Applications of ADVANCED TECHNOLOGY in Transportation

Proceedings of the Ninth International Conference

Edited by

Kelvin C. P. Wang
Brian L. Smith
Donald R. Uzarski
S. C. Wong

ASCE

T&DI
TRANSPORTATION AND DEVELOPMENT INSTITUTE

APPLICATIONS OF ADVANCED TECHNOLOGY IN TRANSPORTATION

PROCEEDINGS OF THE NINTH INTERNATIONAL CONFERENCE

August 13–16, 2006
Chicago, Illinois

SPONSORED BY
Advanced Technology Committee
Infrastructure Systems Committee
The Transportation & Development Institute (T&DI)
of the American Society of Civil Engineers

IN PARTNERSHIP WITH
Transportation Research Board (TRB)
University of Illinois, Chicago

EDITED BY
Kelvin C. P. Wang
Brian L. Smith
Donald R. Uzarski
S. C. Wong

ASCE

T&DI
TRANSPORTATION AND DEVELOPMENT INSTITUTE

Published by the American Society of Civil Engineers

Library of Congress Cataloging-in-Publication Data

International Conference on Applications of Advanced Technology in Transportation (9th : 2006 : Chicago, Ill.)
Advanced technology in transportation : state of the practice : proceedings of the ninth international conference, August 13-16, 2006, Chicago, Illinois / sponsored by Advanced Technology Committee, Infrastructure Systems committee, The Transportation & Development Institute (T&DI) of the American Society of Civil Engineers in partnership with Transportation Research Board (TRB) University of Illinois, Chicago ; edited by Kelvin C. P. Wang ... [et al.].
p. cm.
Includes bibliographic references and index.
ISBN 0-7844-0799-1
1. Transportation engineering—Congresses. 2. Electronics in transportation—Congresses. I. Wang, Kelvin C. P. II. Title.

TA1005.I523 2006
629.04--dc22 2006019292

American Society of Civil Engineers
1801 Alexander Bell Drive
Reston, Virginia, 20191-4400
www.pubs.asce.org

ISBN 0-7844-0799-1
Manufactured in the United States of America.

Preface

The Applications of Advanced Technology in Transportation conference 2006 is the ninth of the AATT conference series. We are fortunate to have two committees of ASCE Transportation and Development Institute (T&DI) to provide technical support, including soliciting abstracts and papers, organizing presentations, sessions, and managing the peer-review process. The two committees are Advanced Technology and Infrastructure Systems. The work from the four editors of the proceedings represents the efforts of the two committees. As in the past, the Transportation Research Board (TRB) also provided sponsorship of the conference. We also appreciate the sponsorship of the conference by the University of Illinois, Chicago.

All papers in the proceedings were peer-reviewed.

We appreciate the diligent work of our two conference co-chairs, Professor Sue McNeil of University of Delaware and Mr. David T. McKay of the US Army Corps of Engineers. The work of the conference co-chairs included organization, communication, marketing, financing, and conference programming. Without their hard work, we would not have a successful conference.

There were a number of volunteers who reviewed the papers included in the proceedings. The conference and the editors appreciate their valuable work. Many of the reviewers have been consistent participants in the AATT conferences.

From the included papers, there is a substantial representation of authors outside of US. The industrialization, modernization, and globalization are taking foot in many of the developing areas, which are certainly reflected in the technical advancement in transportation engineering in these areas.

Lastly, we also thank ASCE and T&DI staff members in marketing, publication, and technical services who made the conference a success.

Editor: Kelvin C.P. Wang, University of Arkansas, Fayetteville,

Co-Editors:
Brian Smith, University of Virginia, Charlottesville
Don Uzarski, Uzarski Engineering, Champaign, Illinois
SC Wong, University of Hong Kong

August, 2006

Acknowledgments

Conference Committee

David T. McKay, P.E., USACE ERDC CERL, Champaign, IL
Sue McNeil, Ph.D., P.E., University of Delaware, Newark, DE

Steering Committee

David T. McKay, P.E., USACE ERDC CERL, Champaign, IL
Sue McNeil, Ph.D., P.E., University of Delaware, Newark, DE
Kelvin C. P. Wang, Ph.D., P.E., University of Arkansas, Fayetteville, AR
Chris Garlick, PBS&J, Orlando, FL

Technical Committee

Chair: Kelvin C. P. Wang, Ph.D., University of Arkansas, Fayetteville, AR
Brian L. Smith, Ph.D., University of Virginia, Charlottesville, VA
Donald Ray Uzarski, Ph.D., P.E., Uzarski Engineering, Champaign, IL
S.C. Wong, Ph.D., The University of Hong Kong, Hong Kong SAR, China

International Advisory Committee

Kumares Sinha, Ph.D., P.E., Purdue University, West Lafayette, IN
Chris Hendrickson, Ph.D., Carnegie Mellon University, Pittsburgh, PA

Contents

Transportation Modeling

GA and ANN Algorithms

Advanced Simulation Models I

Advanced Simulation Models II

Transportation Planning

Demand Modeling I

Demand Modeling II

Urban Transportation Systems

Optimization and Spatial Tools for Planning

Equilibrium Models, Pricing of Land Use, and Public Information System

Indexes

Applying Bayesian Networks to Assess Vulnerability of Critical Transportation Infrastructure

M. K. Jha, Member, ASCE[1]

[1]Center for Advanced Transportation and Infrastructure Engineering Research, Department of Civil Engineering, Morgan State University, 1700 East Cold Spring Lane, Baltimore, MD 21251; PH (443) 885-1446; FAX (443) 885-8218; email: mkjha@eng.morgan.edu

Abstract

The 9/11 terrorist activity followed by Madrid and London mass transit bombings have had far-reaching effects on transportation infrastructure security planning. Transportation infrastructures (such as, airports, metro and subway stations, bridges, and tunnels) are vital to any nation's economy; therefore, in the wake of recent terrorist incidents innovative and robust methods need to be exploited to predict likelihood of terrorist strikes. This research is motivated by the growing vulnerability of transportation infrastructures in the post 9/11 era. Bayesian Networks (BNs) have been proven to be very effective in predicting future scenarios, such as weather forecasts and market predictions. In this paper a BN model is developed for predicting likelihood of a terrorist strike at a critical transportation infrastructure facility, such as an airport or a mass transit station that acts as a hub. An example study is performed to demonstrate the effectiveness of the developed approach. Limitations of the proposed approach are outlined and Dynamic Bayesian Networks are suggested to be exploited in the future for solving real-world problems at the expense of lower computational burden.

Introduction

Terrorism has been the real menace to the U.S. in the post 9/11 era. In recent years billions of dollars have been invested by the federal government in prevention and mitigation of future terrorist activities in the U.S. homeland. The 9/11 terrorist attack followed by Madrid and London mass transit bombings have made it clear that terrorists are constantly targeting locations where a timely attack may cause maximum casualty and economic loss. There also may be long-term psychological impact on people as a result of a terrorist strike. News and media reports suggest that terrorist attacks are a result of well thought-out coordinated set of actions for which, in many instances, years of planning is undertaken by the terrorist networks. From a State's preparedness point of view, which are aggressively engaged in prevention and mitigation, one "loose" action falling though the cracks may result in a

successful terrorist attack. Despite "reasonably good" intelligence available from various sources what often lacks is "coordinated assembly" and "seamless integration" of the available information so that timely and swift action can be taken to avoid an impending terrorist strike. If there was a common modeling tool available to policy makers that would allow predictive capability with "reasonable confidence" it will aid in taking timely and coordinated response. During hurricane Katrina it was observed that lack of "timely" and "coordinated" response by agencies at different levels (such as federal, state, and local) can have disastrous consequences.

The objective of this research is to develop a Bayesian Network (BN) to predict likelihood of terrorist strikes at mass gathering places/events (such as airports, transit hubs, new-year event at time square, independence day parade in Washington, D.C.). Mathematical theories of BNs are briefly reviewed and an innovative model is developed for terrorist strike predictability. An example is performed to predict the likelihood of a terrorist strike at a typical U.S. airport.

Bayesian Networks

A Bayesian Network (BN) (Berger, 1985; Wikipedia, 2005; Zellner, 1971) consists of a directed acyclic graph (DAG). Nodes and arcs in the graph represent dependence relationships. The nodes in the graph are random variables and the arcs show how the nodes connect. There are three connection types: linear, converging and diverging, which make up directed connecting paths. Linear type is just a straight connection from one node to another. Converging type has at least two nodes leading directly into one node. Diverging type has one node that can lead directly into at least two other nodes. The actual connection from node to node carries out the probability of the event to happen along the path that is being followed (Charniak, 1991).

There are many fields that use BNs. For example, economists have used it in market predictions, doctors have used it in medical diagnosis, armed forces have used in countermeasures, and political scientists, and engineers have also used it for various predictions. Microsoft used BN in its office assistance tool and Microsoft head Bill Gates once stated, "Microsoft's competitive advantage lies in its expertise in Bayesian Networks (Helm, 1996)."

An example in the medical field is in predicting protein-protein interaction from genomic data. Integrating evidence from many different sources, researchers decided to use BNs. It allows for probabilistic combinations of the data sets and their illustrations with yeast. The researchers were able to take weakly predictive genomic information and use a BN to produce reliable protein-protein interactions (Jansen, 2003).

Spanbauer (2003), in search for a free Spam filter, came across SpamBayes. It uses BNs to observe and update a database on Spam. It scans the message content and compares it to mail that has been classified as Spam. A spam Probability of about 90 to 100 is moved to the Spam folder. Probability of 0 to 15 is moved to Inbox folder. Those probabilities that were in between the two ranges were put into the Unsure folder to be determined manually if they were spam or not. In one week SpamBayes detected 99 percent of incoming spam. Putting very few messages in the Unsure folder no spam messages were moved to the Inbox folder. In Spanbauer's opinion SpamBayes is the best antispam program that integrates with Microsoft's email software Outlook Express.

Bayesian Networks in Terrorism Related Research

One area of terrorism related research in which BN is being applied is correlated sensor networks. Rob Hill along with the Tactical Systems Section in the Non-proliferation, Arms Control, and International Security Directorate at Lawrence Livermore are using wireless sensors and communications network to detect nuclear terrorist attack. By doing so they want the sensor to share the data they collect. This will allow the user to see a 'bigger picture' (Parker, 2001). Having the BN applied to sensors networks it will lessen the number of false alarms and detect those threats that would have been ignored. BNs are used in deciphering information from a sensor that is communicated to other sensors, to make it more energy-efficient, so that it only communicates new likelihood targets to the other sensors (Parker, 2001).

Information system is another area that is being investigated for a potential BN application to innovate counter-measures to terrorist threats. Kathryn Laskey and others explore the misuse of access privileges. Confidentiality, integrity and availability are three areas that an insider can be compromising in a system. A person is able to get sensitive information making it accessible to unofficial persons or manipulate it to misguide those that have legal access. The BN model extracts normal user behavior by analyzing threatening user behavior over time (Laskey, 2004).

In Laskey's (2004) work BNs are able to combine expert knowledge to collected data. From the information it collects it is able to change and update it to incorporate it with the knowledge it started out with. One variable that it uses in the network is motive and assigned tasks are parent nodes. Motive feeds into global intention which feeds into query task along with assigned task. The end result determines the task relevance. To include more variables researchers used Multi-Entity Bayesian networks (MEBNs) using structures called MEBN Fragments (MFrags). The new model contains seven MFrags. The MFrags are used to model queries and document users accessing information (Laskey, 2004).

Another area using BNs is antiterrorism risk management (Sun 2005; Hudson, 2002). Hudson (2002) and others describe the Site Profiler Installation Security Planner. They apply Bayesian networks in such a way that it allows the user to manage hundreds of threats to make predictions more efficient. This technique is highly developed, systematically sound, adaptable and easily amendable for experts and policy makers (Hudson, 2002).

The BN also performs an important task in the decision-support system for assessing terrorist threats against military installations. It is designed to assess risks specific to a given installation and situation. In Hudson's work a sample BN model contains 146-node that makes up the Risk Influence Network (RIN). RIN solves for the risk of an attack against an asset due to certain threats (Hudson, 2002).

The Mathematical Theory of Bayesian Networks

Using a Bayesian Network (BN) offers many advantages over traditional methods of determining causal relationships. Using BN (Niedermayer, 1998) independence among variables is easy to recognize and isolate while conditional relationships are clearly delimited by a directed graph edge: two variables are independent if all the paths between them are blocked (given the edges are directional). Not all the joint probabilities need to be calculated to make a decision; extraneous branches and relationships can be ignored. The BN algorithm can run in linear time (based on the number of edges) instead of exponential time (based on the number of parameters). The theory of BN is available in standard references (Jason, 2001; Gamez, et al., 2004; Niedermayer, 1998); therefore, has been skipped here.

BNs for Predicting Future Terrorist Strikes

As noted earlier, in recent years BNs have been applied for future predictions in many fields using past data/observations. This makes it an excellent candidate for application in the prediction of future terrorist strikes. Our approach is based on lessons learned from the following three terrorist incidents: (1) 9/11 incident in the U.S.; (2) Madrid mass transit bombing; and (3) London mass transit bombing. A BN model is developed to predict the likelihood of a terrorist strike at a typical U.S. airport. The model is shown in Figure 1. Explanations to Figure 1 legends are provided in Table 1. The dotted links and items in the dotted box are not considered in this analysis. While they are shown for illustrative purposes only, their inclusion will allow us to solve a more precise real-world problem at the expense of higher computational burden.

A scenario is run of predicting the likelihood (probability) of a terrorist strike using hypothetical weekly passenger demand observations. A computer program is developed that computes the marginal and joint probabilities (using exogenous data) of the tree structure shown in Figure 1. Conditional dependencies are carefully examined and eliminated to avoid a problem domain that may be NP-hard as the number of nodes, arcs, and relationship grow. For an average weekly passenger demand of 10,000 the probability of a terrorist attack is found to be 0.39. Please note that this calculation is dependent upon the precision and accuracy of the exogenous data.

Table 1. Figure 1 legend.

Legend	Explanation
Airport	Average daily passenger demand at a typical U.S. airport
US	An airport passenger who is a U.S. citizen
US1	An airport passenger who is a U.S. citizen by birth
US2	A naturalized U.S. citizen
Non-US	An airport passenger who is not a U.S. citizen
Ig	A non-US citizen who is a permanent resident alien
Non-Ig	A non-US citizen who is not a permanent resident alien
S	A non-US citizen who is a Student
T	A non-US citizen who is a temporary Worker
V	A non-US citizen who is a visitor
O	Other non-US citizen
FV-Yes	Foreign Visit in the last five years
FV-No	No foreign visit in the last five years
R-1, R-2, R-3	Visit to a foreign country falling in regions 1-3; regions describe susceptibility to affiliation with a militant/terrorist group for terrorist training and logistical support
TS	Likelihood of a terrorist strike

Limitations and Future Works

The contribution of this paper is in demonstrating the power of BNs in predicting future terrorist strikes if appropriate data/observations were available that may, at times, be classified. It will be desirable to develop a BN-based simulation model for improved

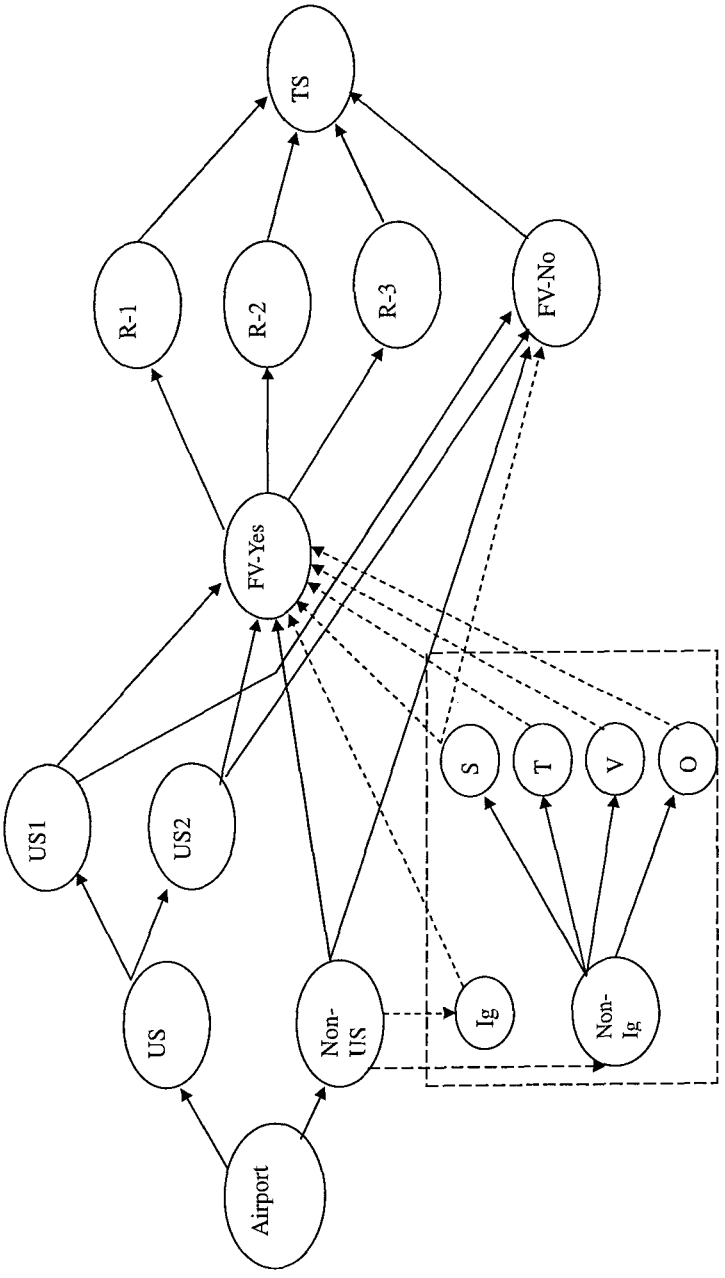


Figure 1. A BN terrorist attack prediction model for a typical U.S. airport

precision and robustness. While adding complexities to the model may yield an enhanced and reliable model (capable of solving "close to real-world scenarios") it will significantly add the computational burden. To combat this issue, a Dynamic Bayesian Network (DBN) is currently being developed, which has the potential to solve complex real-world problems at the expense of reduced computational burden.

Acknowledgements

This work was motivated by the writer's summer 2005 work as a research fellow at the START (Study of Terrorism and Responses to Terrorism) Center of Excellence at the University of Maryland, College Park funded by the Department of Homeland Security. He thanks his student researcher Kimberly Freeman for her assistance with the literature review while working in the research team at the START center. This work, in part, was completed at the Center for Advanced Transportation and Infrastructure Engineering Research (CATIER) at the Morgan State University.

References

Berger, J. (1985). "*Statistical Decision Theory and Bayesian Analysis.*" Springer-Verlag, New York.

Charniak, E. (1991). *"Bayesian Networks Without Tears,"* AI Magazine, 12, Winter 1991, 194.

Gamez, J.A., Moral, S., and Salmeron, A. (eds.) (2004). *Advances in Bayesian Networks,* Springer-Verlag, New York.

Helm, L. (1996). "Improbable Inspiration," *Los Angeles Times.* Oct 28, 1996. p. D1.

Hudson, L., Ware, B. S., Laskey, K.B., and Mahoney, S.B. (2006). *"An Application of Bayesian Networks to Antiterrorism Risk Management for Military Planners,"* http://iie.gmu.edu/~klaskey/papers/Antiterrorism.pdf, (accessed: Dec. 30, 2005).

Jansen, R., et al. (2003). *"A Bayesian Networks Approach for Predicting Protein-Protein Interaction from Genomic Data."* Science, 302, Oct 2003, 449-453.

Jason, F.V. (2001). *Bayesian Networks and Decision Graphs,* Springer-Verlag, New York.

Laskey, Kathryn et al. *"Detecting Threatening Behavior Using Bayesian Networks."* Proceedings of the Conference on Behavioral Representation in Modeling and Simulation, 2004.

Niedermayer, D. (1998). An Introduction to Bayesian Networks and their Contemporary Applications, http:// www.niedermayer.ca/papers/bayesian/ (accessed: Dec. 30, 2005).

Parker, A. (2001). *"Sensing for Danger,"* Science & Technology Review, July/Aug 2001, 11-17.

Spanbauer, S., et al. (2003). *"Purge Unwanted Spam the18th-Century Way,"* PC World, 21, Nov 2003, 194.

Sun, Z., et al. (2005). "Event Driven Document Selection for Terrorism Information Extraction," *Lecture Notes in Computer Science,* 3495, 37-48.

Wikipedia (2005). *Bayesian Network,* http://en.wikipedia.org/wiki/Bayesian_network, (accessed July 13, 2005).

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics,* John Wiley & Sons, Toronto.

Impacts of Condition Assessment Variability on Lifecycle Costs

K.L. Sanford Bernhardt[1] and S. McNeil[2]

[1]Department of Civil and Environmental Engineering, Lafayette College, Easton, PA 18042; PH (610) 330-5584; FAX (610) 330-50509; email: sanfordk@lafayette.edu
[2]Department of Civil and Environmental Engineering, University of Delaware, 360D DuPont Hall, Newark, DE 19716-3120; PH (302) 831-6578; FAX (302) 831-3640; email: mcneil@ce.udel.edu

Introduction

Data describing the condition of physical facilities, such as bridges and pavements, provide the foundation for decisions about maintenance and rehabilitation. Condition data are collected and reported using a variety of technologies, with varying levels of error, uncertainty, and cost. Agencies practicing effective asset management would like to minimize lifecycle costs for their facilities; condition assessment affects lifecycle costs both directly (through the cost of collecting and analyzing the data) and indirectly (through the actions taken in response to the condition assessment). In addition to agency costs, user costs are affected as well.

Uncertainty in pavement condition assessment can have significant effects on average network condition and lifecycle costs. Specifically, the work described uses simulation of pavement deterioration and rehabilitation to track pavement condition and associated maintenance and rehabilitation costs over time. Scenarios evaluated include uncertainty due to systematic errors as well as random variation. Finally, the paper reports the results of a statistical evaluation of the simulation scenarios.

Simulation Model

We developed a simulation model to explore the impacts of errors and uncertainties in the data collection process. The model was originally designed for a bottom-up experiment that demonstrated the behavior of a pavement network as a complex system (Sanford Bernhardt and McNeil 2004). The model simulates the condition of a network of 1000 pavement segments over time in response to varying environmental conditions and maintenance and rehabilitation strategies. The simulation uses standard spreadsheet software (Microsoft Excel®).

Pavement condition is represented by PCI, ranging from 1-100 (worst – best). Initially, all segments are assumed to have the same design, traffic, environment, deterioration rate, useful life, and maintenance actions performed. Initial segment

conditions are randomly and uniformly distributed between 30 and 95. This means that a new pavement has a PCI of 95, and a condition of 30 triggers a need for rehabilitation. After rehabilitation, the PCI returns to 95. We also assume that the rehabilitation cost is the same for all segments and that the deterioration function is linear, with a life of 25 years. We use cost data from Madanat and Ben-Akiva (1994).

As described in Sanford Bernhardt and McNeil (2004), under these conditions, with an unlimited budget, in the steady state, the average pavement will have a PCI of 62.5, and the expenditures on maintenance and rehabilitation are constant. On average, 40 pavement segments will be rehabilitated each year. In reality, a network of pavements is never that well behaved. In our example, with 1000 segments over 25 years, the theoretical equilibrium described above is not reached. The maximum average PCI in any year is approximately 64, while the minimum average PCI in any year is approximately 60. The average PCI over the 25 years is approximately 63. The small group and small time frames do not represent the infinite population and time horizon. The average number of segments rehabilitated each year is 40, as expected.

The Costs of Uncertainty

McNeil (2004) used the simulation model described to review the impacts of uncertainty on user and agency lifecycle costs. User costs are based on the actual pavement condition, whereas agency costs are based on the reported pavement condition. We use the term "reported condition" to refer to the condition as assessed by the inspector, which may be better or worse than the actual condition. Costs are reported as net present values (NPV) with a discount rate of 4%.

Two of the scenarios evaluated reflect systematic uncertainty, and two reflect random uncertainty:

1. The technology consistently reports the pavement condition to be better than actual. In the simulation, the reported condition is calculated as 5 PCI units higher than the actual condition.
2. The technology consistently reports the pavement condition to be worse than actual. In the simulation, the reported condition is calculated as 5 PCI units lower than the actual condition.
3. The technology introduces a random error into the condition with a high level of variability. In the simulation, some segments experience a random change in the condition based on a uniform distribution between +5 and –5 PCI units.
4. The technology introduces a random error into the condition with a modest level of variability. In the simulation, some segments experience a random change in the condition based on a uniform distribution between +2.5 and –2.5 PCI units.

Table 1 summarizes the results of the simulation. Tables 2 and 3 document the analysis of whether the observed changes in average condition and lifecycle cost are

statistically significant. Standard paired t-tests show statistically significant differences in average condition for three of the four scenarios and statistically significant differences in agency and user lifecycle costs for all scenarios.

1. In the case of systematic over-reporting of condition, average network condition declines because necessary maintenance and rehabilitation are not undertaken. Agency costs decline, because fewer/less expensive actions are performed, but user costs rise, because condition is worse. All differences are statistically significant.
2. In the case of systematic under-reporting of condition, the average condition of the network improves because maintenance and rehabilitation actions are more aggressive. Consequently, agency costs also rise, but user costs decline. All differences are statistically significant.
3. Large random variations lead to lower average condition, higher agency costs, and higher user costs. All differences are statistically significant.
4. For small random variations, condition appears slightly better, agency costs lower, and user costs higher. The difference in condition is not statistically significant, but the differences in costs are significant.

Table 1 Summary of Simulation Results

	Average Condition	NPV Agency $/sq m	NPV User $/sq m
Base Case	63.1	$21,546	$70,600
Systematically better condition	60.7	$20,654	$91,613
Systematically worse condition	68.2	$22,412	$61,221
Random errors - large variation	59.1	$23,084	$94,004
Random errors - small variation	63.3	$20,631	$71,961

Table 2 Comparison of Average Condition by Year (n=25)

	Average Condition	Mean Difference to base case	Standard Deviation	t-value
Base Case	63.1			
Systematically better condition	60.7	-2.52	1.80	-7.68**
Systematically worse condition	68.2	4.96	2.73	9.94**
Random errors - large variation	59.1	-4.12	1.21	-17.00**
Random errors - small variation	63.3	0.13	0.93	0.69

** Significant at $p<0.01$

Table 3 Comparison of NPV of Costs by Section (n=1000)

	Average NPV	Mean Difference to base case	Standard Deviation	t-value
Agency Costs				
Base Case	$21,546			
Systematically better condition	$20,654	-$0.89	$0.33	-85.76**
Systematically worse condition	$22,412	$0.87	$0.66	41.42**
Random errors - large variation	$23,084	$2.80	$4.17	21.19**
Random errors - small variation	$20,631	$1.54	$4.95	9.82**
User Costs				
Base Case	$70,600			
Systematically better condition	$91,613	$21.01	$7.08	93.79**
Systematically worse condition	$61,221	-$9.38	$7.51	-39.51**
Random errors - large variation	$94,004	$1.36	$10.57	4.07**
Random errors - small variation	$71,961	$23.40	$21.17	34.96**

** Significant at $p<0.01$

Figures 1 and 2 show that the distributions of average pavement condition remain relatively uniform after 25 years for all five scenarios. Figures 3 and 4 show that the NPV of lifecycle agency costs and user costs are normally distributed for all cases, with shifting means, as expected.

Conclusion

Inaccurate condition assessment leads to statistically significant differences in perceived network condition over time. In addition, it results in statistically significant differences in both user and agency costs. Given that both random and systematic errors in condition assessment impact agency and user costs, further attention is warranted. Recognizing that 1000 segments, each a kilometer long with two 3-meter wide lanes and two 2-meter wide shoulders represents 10 million square meters, agency costs for the base case are $215 m over the life of the pavements (an annual equivalent of almost $14 m per year). Although differences of $0.87 per sq meter (agency costs when reported condition is systematically worse) over the 25 year life of the pavement seem small, $8.7 m (4%) is a considerable value over the life.

References

Madanat, S. and and Ben-Akiva, M. (1994) "Optimal Inspection and Repair Policies for Infrastructure Facilities." *Transportation Science,* 28(10), 55-61.

McNeil, S. (2004) "Selecting Condition Data Collection Strategies based on Life Cycle Costs." *Proceedings of the International Workshop on Integrated Life-Cycle Management of Infrastructures*, December, 2004, Hong Kong.

Sanford Bernhardt, K.L. and McNeil, S. (2004) "Capturing Interdependencies in Pavement Management Decision-making Using Complex Systems Modeling." Proceedings, *6th International Conference on Managing Pavements*, Brisbane, Australia.

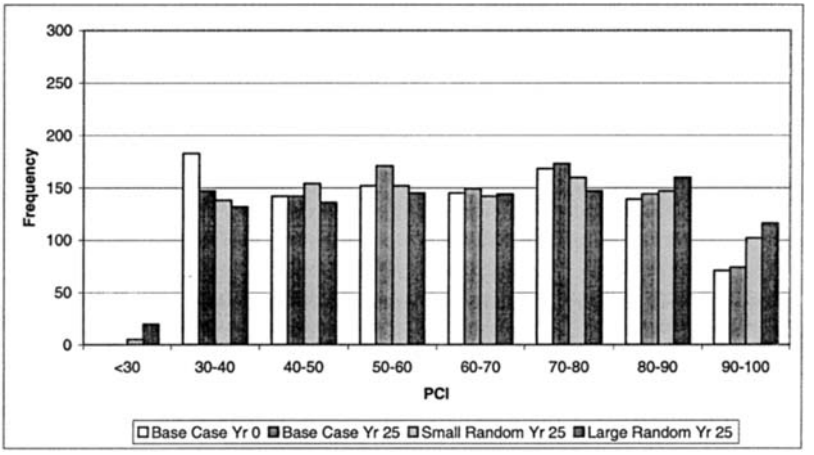


Figure 1 Condition (PCI) Frequency (Random Variation)

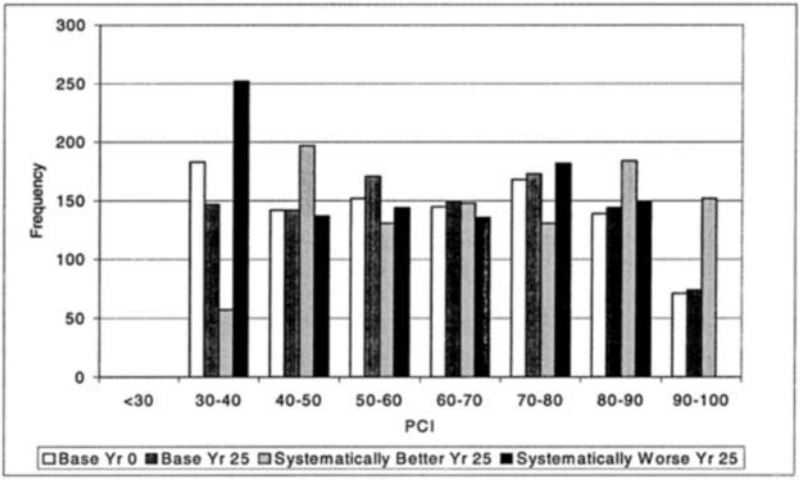


Figure 2 Condition (PCI) Frequency (Systematic Error)

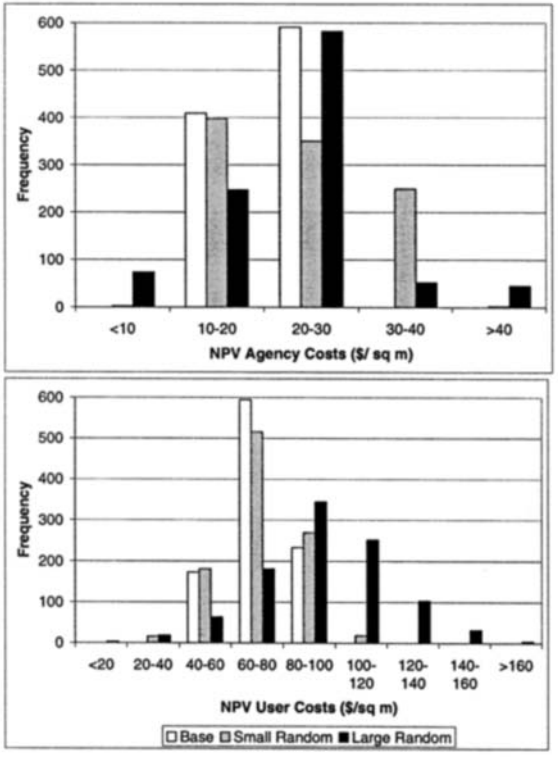


Figure 3 Frequencies of Agency and User Costs (Random Variation)

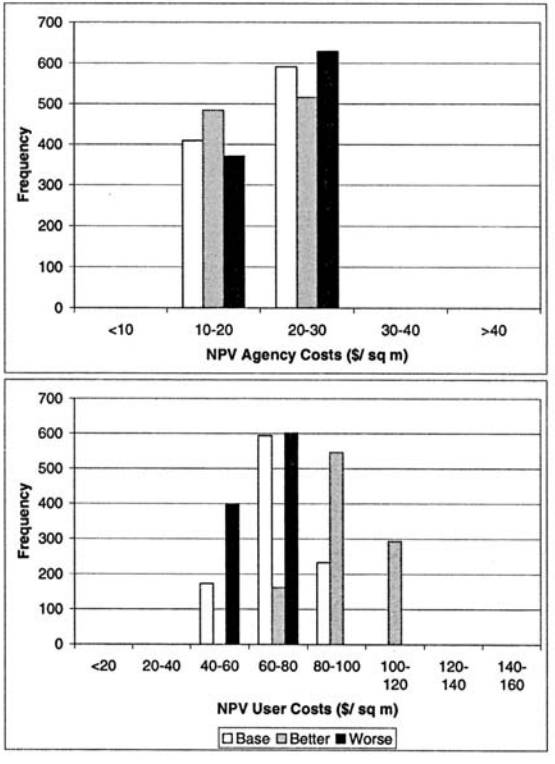


Figure 4 Frequencies of Agency and User Costs (Systematic Error)

Dynamic Programming based Maintenance and Replacement Optimization for Bridge Decks using History-Dependent Deterioration Models

C.-A. Robelin[1] and S. M. Madanat[2]

[1]Graduate Student Researcher, University of California, Berkeley, Dept. of Civil and Environmental Engineering, 116 McLaughlin #1720, Berkeley, CA 94720-1720;
PH (510) 642-7390; FAX (510) 642-1246; email: robelin@berkeley.edu
[2]Professor, University of California, Berkeley, Dept. of Civil and Environmental Engineering, 110 McLaughlin #1720, Berkeley, CA 94720-1720;
PH (510) 643-1084; FAX (510) 642-1246; email: madanat@ce.berkeley.edu

Abstract

In this research, a reliability-based optimization model of bridge maintenance and replacement decisions is developed. Bridge maintenance optimization models use deterioration models to predict the future condition of bridges. Some current optimization models use physically-based deterioration models taking into account the history of deterioration. However, due to the complexity of the deterioration models, the number of decision variables in these optimization models is limited. Some other optimization models consist of a full set of decision variables; however, they use simpler deterioration models. Namely, these deterioration models are Markovian, and the state of the Markov chain is limited to the condition of the facility.

In this research, a facility level optimization model of bridge maintenance and decisions is developed, using a Markov chain whose state includes part of the history of deterioration and maintenance. The main advantage of this formulation is that it allows the use of standard optimization techniques (dynamic programming), while using realistic, history-dependent deterioration models.

This research presents a method to formulate a realistic history-dependent model of bridge deck deterioration as a Markov chain, while retaining relevant parts of the history of deterioration, using state augmentation. This deterioration model is then used to formulate and solve a reliability-based bridge maintenance optimization problem as a Markov decision process. In a numerical example, the policies derived using the augmented Markov chain are applied to a realistic bridge deck, and compared to the policies derived using a simpler Markov chain.

Introduction

Infrastructure management is the process by which agencies monitor, maintain, and repair deteriorating systems of facilities, within the constraints of available resources. More specifically, the management process refers to the set of decisions made by an infrastructure agency over time to maximize the system performance. The basic maintenance and rehabilitation (M&R) decisions an agency has to make are: "in every time period, what M&R activity should be performed on each facility in the system?" The deteriorating bridge population, as well as the limited amount of funds available for maintenance and inspection, led to the development of bridge management systems to assist agencies to make maintenance and rehabilitation decisions by optimizing the use of available funds.

The objective of this paper is to develop a bridge component M&R optimization approach that uses a Markovian deterioration model, while accounting for aspects of the history of deterioration and maintenance. Such a model represents a compromise between simple deterioration models allowing the use of standard optimization techniques, and realistic deterioration models whose complexity prevents efficient optimization of maintenance decisions.

Review of bridge management optimization models

The optimization can be formulated as a Markov decision process (Madanat, 1993; Hawk, 1994; Golabi and Shepard, 1997; Jiang et al., 2000). In these methods, the deterioration is described by a Markov chain, with the state representing the condition of the facility. Optimal solutions are determined using dynamic programming for a single facility. The main advantage of these models is that they enable the use of standard and efficient optimization techniques. As a consequence, these models are implemented in actual Bridge Management Systems such as Bridgit and Pontis (Hawk, 1994; Golabi and Shepard, 1997). The limitation of these Markovian models is the memoryless assumption, according to which the probability for the condition of a facility to transition from an initial state A to a lower state B does not depend on the time spent in state A or on the history of deterioration and maintenance. Although parts of this assumption may be valid for certain bridge states, namely those where the deterioration is primarily governed by mechanical processes, Mishalani and Madanat (2002) have shown empirically that it is unrealistic for bridge states where the deterioration is primarily governed by chemical processes.

Deterioration models in which the history of deterioration is taken into account exist and have been used in bridge maintenance optimization (Mori and Ellingwood, 1994; Kong and Frangopol, 2003; Robelin and Madanat, 2006). However, due to the complexity of their underlying deterioration models, these optimization methods use a very limited number of decision variables in order to remain tractable. To the knowledge of the authors, there does not exist a bridge maintenance optimization method that has more than a few decision variables and that is based on a deterioration model that takes into account the history of deterioration and maintenance. The purpose of the present

article is to develop a bridge maintenance optimization method with a more complete set of decision variables, while using a deterioration model that takes into account important aspects of the history of deterioration and maintenance.

Formulation of a history-dependent deterioration model as a Markov model

Objective. The objective of the present section is to develop a model of the deterioration of a bridge deck with the two following characteristics: the model is Markovian, and it takes into account aspects of the history of deterioration and maintenance.

Definitions and assumptions. The system considered is a single bridge deck managed by an agency such as a state Department of Transportation. Maintenance decisions are made by the agency at discrete points in time, every year. The condition of the deck is represented by its condition index β. By definition of the condition index, the instantaneous probability of failure of the deck (given it has not failed yet) is $\Phi(-\beta)$, where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution.

State of the Markov chain. In earlier Markovian deterioration models, the state is an integer representing the condition of the deck. In the present model, the condition index β of the deck is also part of the state of the Markov chain, as well as three additional variables:

- β^0: the condition index of the deck after the latest maintenance action was performed, or when the deck was new if no maintenance action has been performed yet,
- m: an integer indicating the type of the latest maintenance action performed on the deck (or 0 if no maintenance action has been performed since the deck was new), and
- τ: the time since the latest maintenance action (or the time since the deck was new, if no maintenance action has been performed yet).

The state $x = (\beta, \beta^0, m, \tau)$ consists of two real numbers in the general case (β and β^0) and two integers (m and τ), since there is a finite number of different types of maintenance actions and the unit of time is the year. In practice, the set of possible values for each variable can actually be restricted to small intervals while maintaining the full functionality of the model.

Estimation of the transition probabilities. Transition probabilities represent the probability for a facility that is in state $x_t = (\beta_t, \beta_t^0, m_t, \tau_t)$ at time step t to be in state $x_{t+1} = (\beta_{t+1}, \beta_{t+1}^0, m_{t+1}, \tau_{t+1})$ at the following time step. This transition probability is denoted as $P(x_{t+1} | x_t)$. Note that x_t and x_{t+1} can be any elements of the state space, and may or may not be equal. The original deterioration model of the facility, which is stochastic, is used to estimate the transition probabilities for the resulting Markovian model. In order to accommodate any original deterioration model, Monte Carlo simulation is used to estimate the transition probabilities.

Formulation of the optimization as a Markov Decision Process

Definitions and assumptions. As described earlier, the system considered is a bridge deck managed by an agency. The agency incurs costs when maintenance actions are

performed or when the deck is replaced. Moreover, maintenance actions on a deck or its replacement imply the closure of some or all of its lanes. This leads to delays to the users and/or costs associated with detours.

The agency is responsible for the maintenance and replacement of the bridge for the duration of the planning horizon (T years), after which the bridge is assumed to have a salvage value of V^S.

Problem formulation. Since the deterioration model developed earlier is Markovian, the optimization problem can be formulated as a Markov decision process (Bertsekas, 2001). The following notation is used:
- X: state space of the Markov chain representing the deterioration of the deck. X is the set of all possible values for $(\beta, \beta^0, m, \tau)$, as defined in the previous section.
- U: set of all possible M&R actions, i.e. all types of maintenance actions, replacement, or do-nothing.
- c_u: cost of action u.
- α: discount factor; $\alpha = 1/(1+r)$, where r is the discount rate.
- $V_t(x)$: minimum cost-to-go for the agency to manage the bridge deck from year t to the end of the planning horizon, starting from state x in year t.
- μ: set of optimal decisions. $\mu_t(x)$ is the optimal decision when the bridge deck is in state x in year t. It is the result of the optimization.

The problem formulation is as follows.

$$\begin{array}{lrll} \forall x \in X, & & & \\ V_t(x) & = & \min_{u \in U} \left\{ c_u + \alpha \sum_{y \in X} P(y|x) V_{t+1}(y) \right\} & \text{if } t \in \{0, \cdots, T-2\} \\ & = & V^s & \text{if } t = T-1 \\ \text{subject to} & & & \\ \Phi(-\beta_t) & \leq & P^{\text{acceptable}}, & t \in \{0, \cdots, T\} \end{array} \quad (1)$$

where y is a summation variable representing each state of the Markov chain, and $P^{\text{acceptable}}$ is a user-defined upper bound on the probability of failure.

Solution. The problem formulated above can be solved using backward recursion (Bertsekas, 2001). The minimum discounted cost to manage the bridge over the whole planning horizon is $V_0(x_0)$, where x_0 is the initial state of the bridge.

Case study

This section compares the policies derived using the augmented state Markovian model proposed in this article and the policies derived using a simpler Markovian model, in which the state is the condition of the facility in the current year.

For each Markovian model:
- the coefficients of the transition probability matrices are estimated by Monte-Carlo simulation, as explained earlier, using deterioration parameters adapted from Frangopol et al. (2001),

• the set of policies is determined, using dynamic programming as explained earlier. For this, the costs of maintenance and replacement are adapted from Kong and Frangopol (2003).

Let us call "policies of the simple model" the policies obtained using the simple Markovian model, and "policies of the augmented state model" the policies obtained using the Markovian model with state augmentation. The application of these two sets of policies is finally simulated on two bridge decks having the same deterioration parameters, over a time horizon of 75 years. In the example described in this article, the total cost using the policies of the simple model is 63 percent more than the total cost using the policies of the augmented state model (4,215 dollars per square meter of deck area when using the simple model, and 2,580 dollars per square meter of deck area when using the augmented state model). As shown in Figure 1, the condition of the deck is worse on average when using the policies of the simple model. Moreover, several different values for the deterioration and cost parameters were tried, and the savings obtained through the use of the augmented state model were always significant. The sequence of M&R actions obtained by application of the policies of the simple model is very different from the sequence obtained by application of the policies of the augmented state model, as shown in Figure 1.

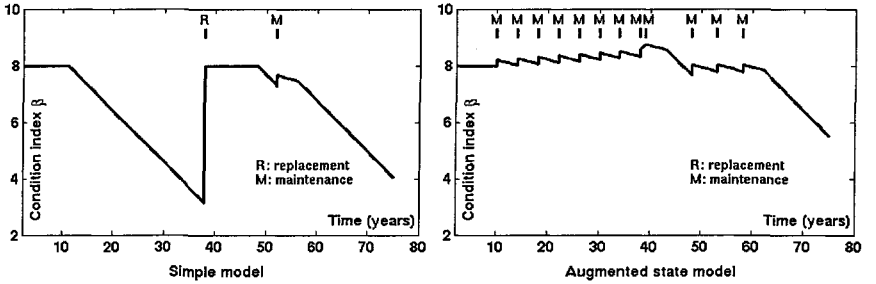


Figure 1: Evolution of the bridge deck condition over the planning horizon, for two scenarios: application of the policies of the simple model (left), and application of the policies of the augmented state model (right).

Using the policies of the augmented state model, the performance of maintenance actions at almost regular intervals is a result of the optimization, and was not provided as an input to the model. A possible intuitive explanation for this fact is as follows. By construction, the state space of the augmented state model captures more detail than the state space of the simple model. In particular, the combination of values for the condition of the facility and for the time since the previous maintenance action is possible in the augmented state model, and not in the simple model. This combination allows for more selective recommendations using the augmented state. For example, if the current condition is 5, the recommendation using the augmented state model may be to perform maintenance if the previous maintenance action was performed seven years before or earlier, and to do nothing if the previous maintenance action was performed less than seven years before. In the same situation, if the current condition is

5, the simple model provides only one recommendation, regardless of the time since the previous action. Thus, the performance of maintenance at regular intervals cannot be recommended by the simple model.

Conclusion

This research presents a method to formulate a realistic history-dependent model of bridge deck deterioration as a Markov chain, while retaining relevant parts of the history of deterioration, using state augmentation. This deterioration model is then used to formulate and solve a reliability-based bridge maintenance optimization problem as a Markov decision process. The numerical example demonstrates the savings brought by the application of the policies of the augmented state model compared to a simpler Markovian model, therefore showing that the additional information included in the model is indeed beneficial.

Acknowledgments

Partial funding for this research was provided by the University of California Transportation Center (UCTC) to the first author through a dissertation-year fellowship.

References

Bertsekas, D. P. (2001). *Dynamic Programming and Optimal Control.* Athena Scientific, Belmont, MA, 2nd edition.

Frangopol, D. M., Kong, J. S., and Gharaibeh, E. S. (2001). Reliability-based management of highway bridges. *Journal of Computing in Civil Engineering*, 15(1):27–34.

Golabi, K. and Shepard, R. (1997). Pontis: a system for maintenance optimization and improvement for U.S. bridge networks. *Interfaces*, 27(1):71–88.

Hawk, H. (1994). *Bridgit: Technical Manual.* Toronto, Canada.

Jiang, M., Corotis, R. B., and Ellis, J. H. (2000). Optimal life-cycle costing with partial observability. *Journal of Infastructure Systems*, 6(2):56–66.

Kong, J. S. and Frangopol, D. M. (2003). Life-cycle reliability-based maintenance cost optimization of deteriorating structures with emphasis on bridges. *Journal of Structural Engineering*, 129(6):818–828.

Madanat, S. M. (1993). Optimal infrastructure management decision under uncertainty. *Transportation Research, Part C*, 1C(1):77–88.

Mishalani, R. G. and Madanat, S. M. (2002). Computation of infrastructure transition probabilities using stochastic duration models. *Journal of Infrastructure Systems*, 8(4):139–148.

Mori, Y. and Ellingwood, B. (1994). Maintaining reliability of concrete structures. II: optimum inspection/repair. *Journal of Structural Engineering*, 120(3):846–862.

Robelin, C. A. and Madanat, S. M. (2006). A bottom-up, reliability-based bridge inspection, maintenance and replacement optimization model. In *Proc. Transportation Research Board (TRB) Meeeting 2006 (CD-ROM)*, TRB, Washington, DC. Paper #06-0381.

Condition and Reliability Prediction Models using the Weibull Probability Distribution

M.N. Grussing[1], D.R. Uzarski[2] and L.R. Marrano[3]

[1]Engineer Research and Development Center - Construction Engineering Research Laboratory (ERDC-CERL), P.O. Box 9005, Champaign, IL 61826-9005; Phone: 217-398-5307; Fax: 217-373-3490; email: michael.n.grussing@erdc.usace.army.mil
[2]Unity Consultants, Inc., 2011 Barberry Circle, Champaign, IL 61821; Phone: 217-398-3984; email: d.uzarski@insightbb.com
[3]Engineer Research and Development Center - Construction Engineering Research Laboratory (ERDC-CERL), P.O. Box 9005, Champaign, IL 61826-9005; Phone: 217-373-4465; Fax: 217-373-3490; email: lance.r.marrano@erdc.usace.army.mil

Abstract

Key to a successful building asset management plan is the ability to measure current condition and predict future condition and degradation trends over a specified planning horizon for each and every individual component-section present in a building. However, this ability poses a difficult challenge because of the vast array of different components, their material make-up, and type found in a building and because each has a different expected life and degradation curve. In addition, condition trends and service lives depend on the amount of preventative and corrective maintenance, including repair, invested during a component-section's lifecycle. Because of these variables, it is difficult to accurately project a condition-lifecycle trend for each individual component-section without a periodic inspection and a meaningful condition metric for measuring condition. This condition metric provides data for the lifecycle prediction process. However, condition data are most often very limited for any given component-section. This paper addresses the use of the Weibull probability distribution function with the data collected during component-section inspections to predict lifecycle condition and reliability over time. The prediction model is self-correcting using attribute information collected during both current and historical inspections to accurately project the unique lifecycle degradation trend for an individual component-section in a building.

Introduction

Recent requirements in civil infrastructure asset management have highlighted the need for improved methods, metrics, and tools to support maintenance, repair, and recapitalization decisions for both public and privately owned facilities (GAO 2003). The objective of these efforts is to minimize total lifecycle costs while maintaining facility condition and performance above specified levels. Lifecycle costs can be optimized by identifying, analyzing, and planning facility repair work in a timely fashion, before the penalty costs due to accelerated facility condition degradation is compounded. This requires knowledge of the relative condition and how that condition degrades over time. For decades, pavement management systems have existed to measure and predict condition for that specific infrastructure domain, but extending that science to building infrastructure management presents a completely new challenge. Currently, existing software applications allow the quantitative calculation of a condition index (CI) value for each component of a building based on an objective condition survey process (Uzarski and Burley 1997). This paper explores the use of this metric to track condition trends and project condition and reliability for the vast and diverse array of building components.

Building Component-Section Lifecycle

Buildings are complex assets comprised of several systems and components, and crossing several specialized civil construction disciplines. Because of this, a rigid hierarchical structure, such as the ASTM Uniformat II (U2) standard for building elements classification (ASTM 2002), is required. The U2 standard divides the building first into major assemblies aligned with the construction trade disciplines, then by building systems, and finally by the individual components that make up those systems. Each component can then be further divided into a component-section, which establishes component attributes based on material, type, age, and location. For example, a wall (component) may be constructed of masonry or wood. The different materials have different responses to their environment over time, have different service lives, and require different work actions at various stages in their lifecycle. As such, the basic management unit for building lifecycle asset management and condition tracking is termed the component-section.

Each component-section works interdependently with other component-sections to support the functions of an efficiently operating building. Each component-section ages and deteriorates over time, adversely affecting its performance and reliability. If left in service for long enough, its condition reaches some limit or failure state at which the component-section can no longer serve its intended function sufficiently (Moubry 2002). It may also adversely affect the function or condition of other component-sections. This limit state occurs at a typical condition index value (CI equals approximately 40) as defined by the building component condition index scale. Due to the nature of their function, certain component-sections, such as structural columns, have a service life designed to correspond to the life of the facility. Other component-sections, such as a roof surface, can have a projected lifespan much shorter than the life of the facility. Periodic repair or replacement of the various component-sections is needed to restore condition and performance capabilities, as well as that of the building as a whole. Depending on the criticality of the component-section and the consequence of a failure, this corrective action may best be performed at or before reaching the failure state.

Predicting this failure state for a unique component-section in a building is difficult, because the true lifespan of a component-section is rarely known, when new. While a designer or manufacturer can provide a generalized idea of design life for a component-section, actual service life depends greatly on local environmental factors, use and abuse, and levels of routine maintenance accomplished. In addition, for many component-sections, simply defining what constitutes a failure state can sometimes be ambiguous. For instance, does a window component-section fail when the vapor barrier is breached, it is no longer operable, a window pane breaks, or some other criteria? This failure state could have a different meaning for different component-sections and to different people. Instead, defining a quantitative failure state based on an objective condition index (CI) metric (Uzarski and Burley, 1997) provides a more consistent definition of component failure.

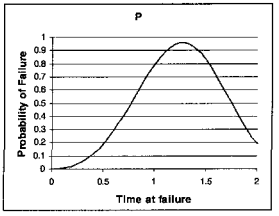
The failure state is rarely the most efficient point when corrective action should be performed. For many component-sections, repair early in the lifecycle can extend life and avert expensive damage caused by accelerated degradation later. The point at which minor corrective action is most efficient is termed the "sweet spot." Experience with the building component CI metric has shown that for a wide variety of components, the repair sweet spot falls in the CI equals 70-80 range. Performing repairs at the sweet spot can result in penalty cost savings from major repair or replacement due to costly critical failure consequences later in the lifecycle.

Reliability and Condition

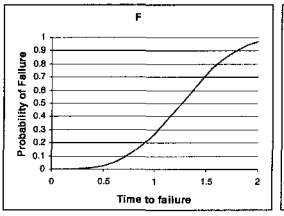
Each building constituent component-section has a finite service life. This service life defines the lifespan for a component-section. Although a building component-section cannot last forever, its service life can be extended with proper operation, maintenance, and even repair. Likewise, a component-section's service life can be significantly decreased by environmental factors, abusive operations, or lack of maintenance. Because of these factors, a wide service life range can exist for a given component-section and there is a limit to the certainty that its service life can be known at the time of construction or installation. Figure 1a shows the probability distribution for the time to failure for a hypothetical component-section. The design service life is the time in service at which the component-section has the greatest probability of failing, but in actuality the true service will be unique for each. Depending on the service life variance within a unique component-section type, there is probability that the actual life could be more or less than the design service life. Design service lives for

a wide range of component-sections are published from different sources based on industry estimates. Unfortunately, unlike pavements, little data exist to describe the variances associated with each service life. In addition, because of the vast and diverse array of individual building components, a family analysis approach, common to pavement infrastructure, to predict condition and service life data is not applicable.

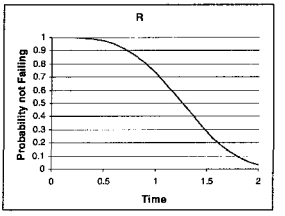
If the variance for the time from component-section construction or installation to failure were known, the statistical probability of that component-section failing at a given year in its lifecycle could be defined. The cumulative failure distribution, Figure 1b, then relates the probability that the component-section will fail at or before a given year. The inverse, Figure 1c, represents the reliability, measured by the probability that the component-section will meet or exceed performance standards at a given year in its lifecycle.



Figures 1a. Failure in year t



1b. Failure before year t



1c. Performs past year t

The model assumes that the condition state measured by the CI and the reliability state are proportionally similar. Both are defined below.

Condition

The physical condition state relates the general health of a building component-section. Physical degradation of the component-section due to normal aging, excessive or abusive use, or poor maintenance can cause a reduction in the component-section's ability to perform as required. In BUILDER® EMS (BUILDER 2005) for example, condition is measured in absolute terms by the use of a condition index metric. The condition index uses a scale of 0-100 with 100 defining "Defect Free" or pristine condition. The lowered condition state is caused by distresses observed during a structured, objective, and repeatable inspection. These distresses have an adverse affect on the component-section's ability to perform. Through a "deduct value" process based on the distress types, severities, and densities present, a condition index is computed. It can be assumed that, when new, a component-section condition index is 100 (i.e. Defect Free).

Reliability

Reliability is the statistical probability that a component-section will meet or exceed performance requirements for a given length of service life. For most building component sections, it is a function of the amount of time a component has been in service. In general, condition and reliability are related as follows:

- Condition and reliability are maximum (100) at or near the start of the service life,
- Condition and reliability approach the minimum state (0) asymptotically,
- Condition and reliability deteriorate unless corrective action is performed, and
- As condition deteriorates, reliability likewise decreases.

Condition Prediction Model Requirements

In order to have a robust condition prediction model, the following methodologies are proposed:

- Model is seeded with reasonable initial assumptions, and "self-corrects" based on collected information,
- Model automatically adjusts the expected service life based on an inspection generated condition index,
- Model takes into account the inspection date and type when calibrating the prediction trend,

- When repair work is completed, model adjusts the prediction trend, and
- Model takes the type of repair work into account when projecting the predicted condition trend.

Weibull Probability Distribution

The Weibull cumulative probability distribution function is used to model the condition lifecycle curve. The Weibull statistical distribution represents the probability of time to failure of a component-section in service. It has natural boundary conditions which abide by the assumptions discussed above, and takes the shape of a classical condition deterioration curve. The resulting mathematical condition prediction model is:

$$C(t) = a \times e^{-(t/\beta)^{\alpha}} \quad (1)$$

Where:

C(t) = component-section condition index as a function of time
t = time, in years, since component-section was installed or constructed
e = exponential
a = parameter, initial steady state component-section condition index
β = parameter, service life adjustment factor
α = parameter, accelerated deterioration factor

Initial Model Seeding

The first step of the self-correcting prediction model process uses initial general assumptions to compute the parameters a, α, β which describe the shape of the condition lifecycle trajectory. When the component-section is new or there are no inspection data the only lifecycle information available is the installation or construction date for the component-section (at which the assumed CI is 100) and the expected service life (at which assumed CI equals some terminal value at end of service life). The model must also be seeded with a degradation factor. An example is illustrated in Figure 2 with an example 30 year service life. The actual degradation factors can be set individually for each component-section based on historical trends, if known. If unknown, any reasonable values may be chosen because the model will self-correct once inspection data are collected.

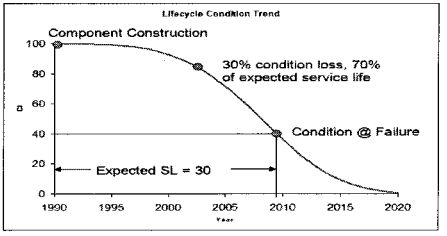


Figure 2. Example Initial Lifecycle Condition Trend

These assumptions initialize the prediction model when no other information exists. Each assumption results in a data point which is a function of component-section time in service and condition. The three model parameters are then solved using those data points to describe the shape and trajectory of the component-section lifecycle curve.

Calibration with Inspection Information

The initial model seeding described above may be reflective of a typical component-section, but they do not account for the unique reliability or behavior of a specific individual component-section present in a specific building. Therefore, as time progresses and the component-section ages and degrades, it becomes important to assess the component-section condition at various points during the lifecycle to compare and calibrate the expected condition with the actual observed condition.

As the component-section progresses farther into its lifecycle and more inspections are performed, these historical inspections form the shape of the observed and projected lifecycle curve. Actual historical condition degradations trends are then used to accurately model the behavior of the component condition and reliability profile. The initial industry average estimate of expected service life is re-adjusted based on information about how the component is actually degrading. Figure 3 illustrates how collected inspection data are used to readjust the expected service life and lifecycle curve. As more data are added, the better the model becomes tailored and unique for any given component-section. Figure 3 also illustrates how the expected service life shifts.

Weighting of Historical Inspection Data

After several inspection points are collected, the model begins to accumulate information about the behavior of the component-section's condition over time. However, some inspections and data points are more accurate than others depending on the inspection type done, the level of inspection detail, and the inspection date. The condition prediction process takes these factors into account when adjusting the model.

The mathematical Weibull model has only three parameters or degrees of freedom to define the condition lifecycle curve. However, with the initial installation or construction date, the expected service life date, and several inspections collected, more than three data points exist. The model uses regression analysis to fit the prediction curve through data points by minimizing the sum of squares residual error. Each point is also associated with a weighting factor that modifies its residual error. The more accurate the data point, the higher the weighting factor and the more affect it has on the adjusted model. Also, the adjusted prediction curve always passes through the last known condition index point.

Factors that affect the weighting values for a data calibration point include the component-section installation or construction date certainty, the time (in years) since the previous inspections, the inspection type of inspection (distress survey or direct rating) (Uzarski 2004), and the change in CI value between inspections.

Scheduling Inspections

The proposed model above illustrates how important inspection scheduling is in the condition prediction process. The quantitative BUILDER® inspections provides the information for self-correcting the model based on actual conditions. The prediction model is always most accurate in the time near a well trusted data point, such as in the case of a recent inspection. As time passes and that inspection becomes dated, the predicted condition becomes less certain. Depending on the predicted condition, the certainty of that predicted value, the consequence of prediction error, or the amount of service calls or trouble tickets, it may be time to schedule another inspection. This new inspection once again verifies and self-corrects the predicted condition model.

The proposed prediction model process provides a way for efficiently scheduling inspections and justifying the benefits. Since inspections cost money and require personnel resources, performing inspections at selected times in the lifecycle, based on need and not the calendar, efficiently allocates these resources.

Calibration when Work is Performed

In addition to adjusting the condition model based on inspection information, corrective repair work has an effect on the future condition trend. If the projected condition drops below a designated acceptable level or standard, component-section repair or replacement may be necessary. If a repair is performed, then the deterioration model experiences a step function increase at the time of repair which raises the CI up to an assumed value of 95. Since it cannot be assumed that the component-section has been repaired to a "Defect Free" condition (CI=100), a reduced CI is used. If a quality control inspection is performed on the work, the actual CI value, including 100, is used in the model. To predict the component-section condition trend response after repair, the established pre-repair degradation trend is extended (see Figure 4). The model also takes into account the age of the component-section at the time of repair when determining the new degradation rate. Generally, the older a component-section is at the time of repair, the faster the post-repair degradation will be

when compared to the pre-repair rate, but there are exceptions. Thus, multiple repairs cannot extend the life of a component-section indefinitely. Eventually, component-section replacement will be required. Replacement will reset the service life clock. Of course, as discussed above, inspections will be periodically conducted on the repaired component-section which will adjust the lifecycle curve.

When the component is replaced with an identical component-section (replacement "in kind"), the prediction model takes the characteristics of the replaced component-section's degradation trend to initialize the new component-section condition prediction model.

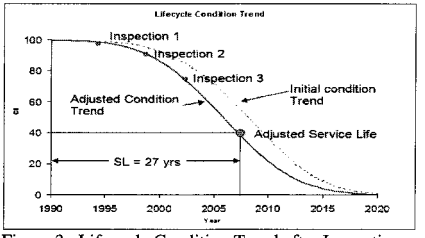


Figure 3. Lifecycle Condition Trend after Inspection Data Calibration

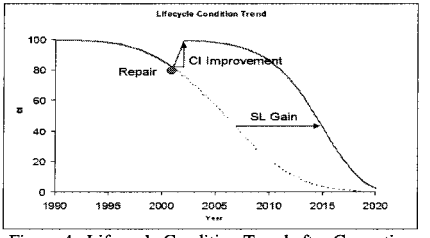


Figure 4. Lifecycle Condition Trend after Corrective Repair

Conclusions

The measurement and prediction of future facility component condition trends is essential to a reliability-centered building lifecycle management program. However, with the large number of dissimilar building components in a building portfolio, and the lack of detailed models to describe the response in service to each of these individual components, predicting condition performance for a building is quite a challenge. Current models and industry estimates of component service lives rarely account for the local conditions of use, maintenance, and environmental factors, making the condition prediction for an individual asset less accurate. The proposed model described above provides a feedback loop to self-correct individual lifecycle trends for a wide variety of building components based on periodic objective assessment observations. In addition, corrective repair work performed during the lifecycle of a component-section is factored into the lifecycle trend analysis. This reliability/risk-based condition prediction model provides the basis for optimized planning of both inspection resources to refine the condition trend as required, and repair and recapitalization resources to plan corrective work at the appropriate time in the component-section lifecycle.

References

ASTM International (2002). Standard Classification for Building Elements – Uniformat II, ASTM E 1557-02.

BUILDER (2005). BUILDER Engineered Management System, http://dwww.cecer.army.mil/td/tips/product/details.cfm?ID=64&TOP=1

GAO (2003). Changes in Funding Priorities and Strategic Planning Needed to Improve Condition of Military Facilities, www.gao.gov.

Moubray, John. (2002). Reliability-Centered Maintenance, Industrial Press Inc., New York.

Uzarski, D.R. and L.A. Burley. (1997). "Assessing Building Condition by the Use of Condition Indexes," proceedings of the ASCE Specialty Conference Infrastructure Condition Assessment: Art, Science, Practice, pp.365-374.

Uzarski, D.R. (2004), "Knowledge-based Condition Assessment Manual for Building Component-Sections" U.S. Army Construction Engineering Research Laboratory. (Distributed with BUILDER® software)

Estimating Pavement Performance Models using Advanced Technologies and Time Series Analysis

Chih-Yuan Chu[1] and Pablo L. Durango-Cohen[2]

[1]Graduate Student Researcher, Department of Civil and Environmental Engineering & Transportation Center, Northwestern University, 2145 Sheridan Road, A321, Evanston, IL 60208-3109, email: jameschu@northwestern.edu
[2]Assistant Professor, Department of Civil and Environmental Engineering & Transportation Center, Northwestern University, 2145 Sheridan Road, A335, Evanston, IL 60208-3109, tel: +847 491 4008, fax: +847 491 4011, email: pdc@northwestern.edu

Abstract

We propose state-space specifications of autoregressive moving average models and structural time series models as a framework to develop and estimate performance models for transportation infrastructure facilities. Time series models in state-space form fit the maintenance optimization model of Durango-Cohen (2005) and are consistent with the latent performance modeling approach of Ben-Akiva and Ramaswamy (1993). To illustrate the proposed framework we develop and estimate performance models for an asphalt pavement using pressure and deflection measurements generated by sensors and falling weight deflectometers, respectively. Analysis of the results shows that the ensuing models are consistent with physical properties of flexible pavements. The results also indicate that state-dependence may be statistically significant and further reinforces the computational and statistical advantages of the proposed framework over Markovian transition probabilities.

Introduction

The ability to generate accurate condition forecasts with a pavement performance model is an essential part of developing efficient Maintenance and Rehabilitation (M&R) policies. We describe and compare AutoRegressive Moving Average (ARMA) models and structural time series models to represent the evolution of pavement performance. We consider state-space specifications of the models because they are consistent with the latent performance modeling approach proposed by Ben-Akiva and Ramaswamy (1993), which means that they rigorously account for uncertainty in forecasting infrastructure performance when condition data are gathered simultaneously using multiple technologies. In addition, the models can be readily used for optimization of M&R decisions by Durango-Cohen (2005), which means that they constitute an alternative to the use of Markovian transition probabilities to

forecast infrastructure condition. Through an empirical study, we illustrate how the proposed methodology is valid to forecast pavement performance when condition data are gathered simultaneously using multiple advanced data collection technologies.

Literature Review

State-Space Specifications for Time Series Models

We adopt state-space specifications to estimate time series models following Harvey (1990). They are consistent with the latent performance modeling approach of Ben-Akiva and Ramaswamy (1993) which has been shown empirically to be adequate to represent the evolution of transportation infrastructure. State-space models, originally postulated for the optimal control of engineering systems, have been adapted to analyze time series. However, time series analysis using state-space specifications is more attractive than Box-Jenkins models because it does not rely on stationary data. The general form of a state-space model is shown in Equation (1) and (2).

$$X_{t+1} = g_t X_t + h_t A_t + \varepsilon_{t+1} \quad (1)$$

$$Z_t = H_t X_t + \xi_t \quad (2)$$

where X_t is a vector used to represent a system's state, e.g., a facility's condition. A_t is a vector of exogenous inputs that may include a maintenance/investment rate, environmental factors, or traffic loading. Z_t is a vector used to represent the condition measurements. g_t is the transition matrix of state vectors. h_t contains the parameters for explanatory variables and H_t is a matrix describing the relationship between measurements and latent condition. Finally, ε_t and ξ_t represent random error terms that are assumed to follow Gaussian Distributions with finite second moments. We also assume, without loss of generality, that they have zero means. Σ_ε and Σ_ξ are the covariance matrices of ε_t and ξ_t. Equation (1) is the system equation (or state equation) and Equation (2) is the measurement error model (or observation equation). Equation (1) governs the dynamic transition of the system and the effect of explanatory variables. The components of the equation are formulated according to the nature of the systems of interest.

Methodology

Formulation of ARMA(p, q) Models

An ARMA(p,q) model with a facility's state and k measurements has the following form. The true condition is generally set to be equal to one of the measurements, the reference measurement. This involves setting $\lambda_1 = 1$ in Equation (4).

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}, t = 1,2,\ldots,T \quad (3)$$

$$z_t^{(i)} = \lambda_i x_t + \xi_t^{(i)}, i = 1,2,\ldots,k, t = 1,2,\ldots,T \quad (4)$$

Formulation of Structural Time Series Models

Structural time series models consist of components that capture trend, seasonality, and deterministic or stochastic irregularity. Lagged dependent variables can also be included to capture the dependence on the history of the process. A general representation of a structural time series model is shown in Equation (5) below and Equation (4). In Equations (5), μ_t is the level of condition of the system at time t, β_t is the trend at time t, and γ_t is the seasonal component at time t. η_t, ζ_t, and ω_t are random errors of level, trend, and seasonality, respectively. Their formulations are shown in Equation (13) through (15). This representation also has one facility's state and k measurements. The special case with no lagged dependent variables in the model is obtained by setting $\phi_1 = \phi_2 = \ldots = \phi_p = 0$ and $t = 1,\ldots,T$.

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + \mu_t + \gamma_t + \varepsilon_t, t = p+1,\ldots,T \qquad (5)$$

Empirical Study

The purpose of this section is to describe the process of preparing the required data and illustrate the estimation of the infrastructure performance models described in the previous section.

Data

MnRoad is at the forefront of using advanced monitoring technologies for condition assessment of pavements. The data for the study consist of deflection and pressure measurements, both representing pavement load-bearing capacity. Both of them are collected in the four year period from December 1999 through November of 2003. The measurements are collected from cell 33, a Superpave test cell located on a closed-loop test track. Superpave (SUperior PERforming Asphalt PAVEments) refers to strict criteria for properly designing and building hot-mix asphalt (HMA) pavements to perform better under the extremes of temperature and heavy truck loadings (Palmquist et al. 2002; Zerfas 2003).

The tool used to collect deflection in MnRoad is a **Falling Weight Deflectometer** (FWD). The features of this automated equipment are accuracy and non-destructive test. However, the problem of this equipment is that it is not operating at driving speed and still requires agencies to block the roads in order to collect data. The data used in our analysis consist of the monthly average deflection measurements (in thousandths of an inch – *mils*) induced by a FWD and the data are presented in Figure 1.

Pressure measurements in MnRoad are collected by inducing a load generated by a five-axle truck. The pressure measurements (in *millivolts*) are collected with a **dynamic soil pressure cell sensor**, which is set on top of subgrade. One of the reasons for choosing the measurement is that the sensor is still fully functional after four years of operation while many other sensors installed in the pavement have failed. Although the data collected by sensors are automatic and frequent, the nature of pavement facility makes the replacement of failed sensors very difficult. The data are presented in Figure 1.

Estimation Results

Generally, the prediction error decreases as the number of estimated parameters increases. To avoid "overfitting" in model selection, we use Akaike's Information Criterion (AIC), which considers both goodness-of-fit and number of parameters, to select the desired model specifications in the empirical study.

ARMA time series models

The data in Figures 1 show clear seasonality. Therefore, it is necessary to consider ARMA models of order 13 or higher for monthly data. Several models were considered and the best results were obtained from an AutoRegressive (AR) model of order 14, which is shown in Equations (6) and (7). The model results in the best AIC; in other words, it has the best trade-off of the number of parameters and model fit. Further, from Figure 1 the model fits both measurements adequately. The residuals of both measurements from AR(14) model are further tested using sample autocorrelation function of residuals. Since over 95% of the function values are within $\pm 1.96\sqrt{T}$, the residuals are i.i.d. (independent and identically distributed) with $N(0,1/T)$ and the model is statistically satisfactory.

From Equation (7), the measurement error of deflection is much greater than pressure, which also can be seen from Figure 1. Since the two measurements have different magnitudes, variance of measurement is not sufficient to compare technologies and we further use coefficient of determination to choose the better technology. The result shows that R^2 of deflection is 0.80 while R^2 of pressure measurements is 0.89. It means that pressure sensor not only has smaller measurement error but provides better fit than FWD. This information might be useful for technology selection as suggested in Ben-Akiva and Ramaswamy (1993).

$$x_t = 0.73x_{t-1} - 0.30x_{t-2} + 0.09x_{t-3} - 0.07x_{t-4} - 0.06x_{t-6} + 0.06x_{t-7} + 0.03x_{t-8} + 0.01x_{t-9} - 0.08x_{t-10} + 0.47x_{t-11} + 0.09x_{t-13} + 0.03x_{t-14} + \varepsilon_t, \varepsilon_t - N(0,0.03) \quad (6)$$

$$\begin{bmatrix} z_t^{(1)} \\ z_t^{(2)} \end{bmatrix} = \begin{bmatrix} 1.00 \\ 0.26 \end{bmatrix} x_t + \begin{bmatrix} \xi_t^{(1)} \\ \xi_t^{(2)} \end{bmatrix}, \Sigma_\xi = \begin{bmatrix} 25.99 & -1.89 \\ -1.89 & 1.08 \end{bmatrix} \quad (7)$$

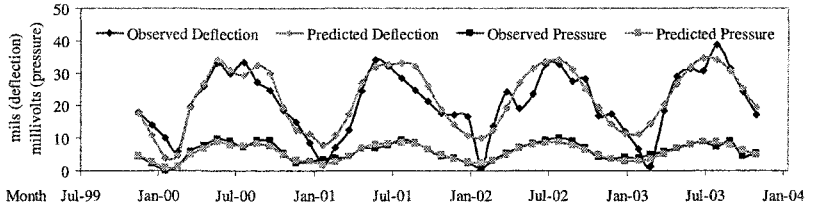


Figure 1. Model vs. Data for AR(14)

Structural time series models

Comparing to ARMA specification, structural time series models are more effective when clear structures such as trend and seasonality exist while ARMA models can

work without any prior knowledge. Moreover, modeling seasonality is parsimonious here. The only parameter required for seasonality in structural time series models is the random error ω_t in Equation (11) and is not decided by the frequency of data. To identify the components of the data set, three Basic Structural Models (BSM) are estimated. The best results in terms of AIC are listed in Equations (8) through (12). The model incorporates both the one and two-step lagged dependent variables. Comparing it and the other two models, virtually all the variances of the random error terms are lowered after adding the lagged dependent variables, i.e., the model captures the transition of the system better and the model itself becomes "less stochastic" due to the lagged dependent variables in the model. Moreover, the fact that the two-step lagged dependent variable is highly significant indicates that the Markovian assumption might not hold. The fact that the one-step lagged dependent variable is not significant can be explained by the inclusion of the "trend" variable in the model.

Figure 2 also shows that the model fits the data adequately. The beauty of the structural time series models is the meaningful components of the series, which is shown clearly in Figure 3. In the figure, the level, trend, and seasonality are clearly decomposed. The seasonality of the state matches the seasonal change of the deflection measurements reasonably. However, another important component of a structural time series model, the trend, is harder to determine from this data set. The negligible trend matches the fact the section had not deteriorated in the period of experiment. This is explained by the facts that the pavement section is new and the higher criteria of design and material are used for Superpave. Note that if structural deterioration of pavement were to exist, this specification would be able to detect it. Moreover, the residuals of both measurements from the model are further tested using sample autocorrelation function. The test shows that the residuals are iid, which indicates that the preferred model is statistically satisfactory.

Note that the measurement error of deflection is still higher than that of pressure. However, deflection has higher R^2 than pressure (0.86>0.77). Thus, for structural time series model, the best technology for measuring pavement condition would be FWD, which contradicts the result in ARMA models. Hence, in this empirical study, the best strategy of technique selection depends on model specifications and is inconclusive.

$$x_t = 0.13x_{t-1} - 0.63x_{t-2} + \mu_t + \gamma_t + \varepsilon_t, \varepsilon_t - N(0, 0.09) \tag{8}$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_{t-1}, \eta_t - N(0, 0.26) \tag{9}$$

$$\beta_t = \beta_{t-1} + \zeta_{t-1}, \zeta_t - N(0, 0.01) \tag{10}$$

$$\gamma_t + \gamma_{t-1} + \gamma_{t-2} + \cdots + \gamma_{t-11} = \omega_t, \omega_t - N(0, 0.0005) \tag{11}$$

$$\begin{bmatrix} z_t^{(1)} \\ z_t^{(2)} \end{bmatrix} = \begin{bmatrix} 1.00 \\ 0.33 \end{bmatrix} x_t + \begin{bmatrix} \xi_t^{(1)} \\ \xi_t^{(2)} \end{bmatrix}, \Sigma_\xi = \begin{bmatrix} 15.52 & -2.54 \\ -2.54 & 2.20 \end{bmatrix} \tag{12}$$

Conclusions

In this paper, we propose state-space specifications of ARMA and structural time series models as a framework to develop and estimate performance models for transportation infrastructure facilities. The framework rigorously accounts for uncertainty in the deterioration process and in the data-collection process, when data are gathered by multiple technologies. The performance models also fit the

optimization model by Durango-Cohen (2005) that supports investment decisions in M&R of transportation infrastructure.

We present an empirical study where we validate the proposed methodology to describe pavement performance. Specifically, we estimate and compare ARMA and structural time series models to simultaneously formulate deflection and pressure measurements. Estimating the measurement errors of inspection equipments could help choose the strategy of technique selection. In addition to verifying that the proposed models provide adequate and reasonable modeling fitting, the results of the empirical study show that the history of the deterioration process is statistically significant.

Acknowledgments

The authors gratefully acknowledge the personnel at MnRoad, who have provided extensive data and technical assistance for the study.

References

Ben-Akiva, M. and Ramaswamy, R. (1993). "An approach for predicting latent infrastructure facility deterioration." *Transportation Science*, 27(2), 174–193.

Durango-Cohen, P. L. (2005). "Transportation infrastructure management under uncertainty: A time series framework." *under review Transportation Research Part B.*

Harvey, A. C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press, New York, NY.

Palmquist, D., Worel, B., and Zerfas, W. (2002). *Mn/ROAD Hot-mix Asphalt Mainline Test Cell Condition Report.* Office of Materials and Road Research, MnDOT, MN.

Zerfas, W. J. (2003). *Superpave Report.* Office of Materials and Road Research, MnDOT, MN.

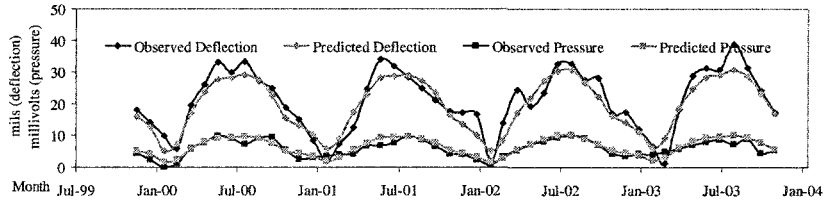


Figure 2. Model vs. Data for BSM

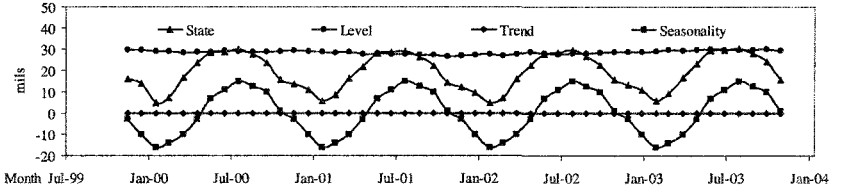


Figure 3. Decomposition of State for BSM

Incorporating Unobserved Heterogeneity in Pavement Deterioration Modeling

Feng Hong[1] and Jorge A. Prozzi[2]

[1]Research Assistant, Department of Civil Engineering, The University of Texas, ECJ 6.510, Austin, TX 78712; Phone: (512) 232-6598; email: fenghong@mail.utexas.edu
[2]Assistant Professor, Department of Civil Engineering, The University of Texas, ECJ 6.112, Austin, TX 78712; Phone: (512) 232-3488; Fax: (512) 475-8744; email: prozzi@mail.utexas.edu

Abstract

Performance model is one of the key elements in transportation infrastructure management systems. Both observed and unobserved difference among the sections (heterogeneities) should be accounted for in modeling the performance. The observed heterogeneity is addressed through explanatory variables; while the unobserved heterogeneity results from other factors not included in the explanatory variables. Most of the existing work only focuses on the observed heterogeneity. Thus, the uncertainty on performance modeling is not well captured. In this paper, the issue on performance heterogeneity is investigated. A structured econometric model with hierarchical parameter structure is established. The Maximum Simulated Likelihood approach is applied to estimate the model. It is found through statistical analysis and test that unobserved heterogeneity is significant. In addition to the population parameter estimates, the individual-specific parameter estimates are also obtained, which are helpful in revealing the deterioration behavior of individual sections.

Introduction

Deterioration models play a critical role in planning and decision making for managing transportation infrastructure facilities. The statistical regression approach is the most widely used in model development. These models express the performance as a function of related explanatory variables such as structure and material properties, traffic load, environmental factors, maintenance or rehabilitation (M&R) treatment, and others (e.g., Garcia-Diaz and Riggins, 1984; Al-Suleiman, et al, 1988). In most of their work, efforts were focused on fitting the performance through linear or nonlinear curves and revealing the significant relationship between the performance and explanatory variables. As an inborn nature, performance heterogeneity exists in infrastructure systems due to the uncontrolled variability of factors such as varying construction quality, material properties, and others. However, this critical issue was usually not fully accounted for or even neglected in the existing work. In such cases, the model estimates are inconsistent.

During performance inspection, many panel data sets are available to assist infrastructure management. These data sets have been proven to lead to a better understanding of infrastructure deterioration behavior, particularly on performance heterogeneity. A random effects (RE) specification was developed to model the bridge-deck deterioration in Madanat et al. (1997). Their models were found to yield improved results over the previous simple models without accounting for heterogeneity. Subsequently, a series of pavement deterioration models were thoroughly investigated in Prozzi (2001), and Prozzi and Madanat (2003). It was found that the unobserved heterogeneity was significant and thus suggested be properly incorporated in performance modeling. These models assumed that the heterogeneity was represented by the model's "intercept" term while the rest of the parameters ("slopes") were fixed. However, some of the "slopes" may vary across the individuals in reality. For instance, the contribution of unit thickness of pavement surface layer to the total structural strength may differ across sections resulting from the varying compaction quality in construction due partly to different thicknesses. Thus, those sections may not share a common surface layer coefficient. In this sense, only considering the random intercept is not sufficient to fully address the heterogeneous effect. In addition to the "intercept", the other parameters deemed to vary across individuals should also be properly addressed accordingly.

Data Set

Generally, two types of data can be used in modeling pavement deterioration, field and experimental data. In this study, the latter is adopted simply because the factors affecting pavement performance can be well controlled and the statistical problems such as endogeneity can be mitigated or avoided. American Association of State Highway Officials (AASHO) Road Test data are used in model development. Six two-lane loop tracks were constructed for the test, with loops 2 through 6 subject to traffic loading. Approximately 1,114,000 axle repetitions were applied on the test sections using trucks with three types of axle configurations. Particularly, the flexible pavement sections are used in this study.

Model Specification

Following the similar deterioration principle of AASHO model (HRB, 1962), a general deterioration model is proposed as,

$$p_t = a - bN_t^{\ c} + \mu \tag{1}$$

Where, p_t = serviceability at time t; N_t = some measure of traffic until time t; a = initial serviceability; b = deterioration rate; c = deterioration curvature; and μ = error term. The two terms to be particularly noted are a and b. What follows is to determine the corresponding specification terms in Equation (1).

It was found that the initial serviceability was positively associated with the surface layer thickness in Prozzi and Madanat (2003). Based on their work, the initial serviceability is expressed as,

$$a = \phi_0 + \phi_1 \exp\{\phi_2 H_1\} \tag{2}$$

Where, $\phi_0 \sim \phi_2$ = parameters to be estimated; and H_1 = surface layer thickness of a pavement section.

The serviceability loss, denoted as $bN_t^{\ c}$ should involve three basic aspects governing pavement performance: pavement structure and material, traffic, and environment. It is known that the serviceability on stronger pavement will decrease more slowly than weaker pavement. The specification for b is designated to account for each pavement layer's contribution to the resistance of serviceability loss:

$$b = \exp\{\gamma_0\}\exp\{\gamma_1 H_1 + \gamma_2 H_2 + \gamma_3 H_3\} \tag{3}$$

Where, γ_1, γ_2, and γ_3 denote the parameters for the individual pavement structures, surface layer, base, and subbase. H_1, H_2, and H_3 are the thicknesses of the three layers respectively. γ_0 represents the parameter for the subgrade. It can be shown that $\exp\{\gamma_0\}$ represents the serviceability loss by the first equivalent traffic load running on the subgrade.

For traffic, due to three typical axle configurations were involved in the Road Test, the following specification for the load equivalency factor for one individual vehicle is adopted, denoted as *LEF*,

$$LEF = \left(\frac{FA}{\alpha_1 18}\right)^{\alpha_0} + A\left(\frac{SA}{18}\right)^{\alpha_0} + B\left(\frac{TA}{\alpha_2 18}\right)^{\alpha_0} \tag{4}$$

Where, $\alpha_0 \sim \alpha_2$ = parameters to be estimated; FA = front axle load magnitude, (kip or kN); SA = single axle (with dual wheels) load magnitude, (kip or kN); TA = tandem axle load magnitude, (kip or kN); A = number of single axles on one vehicle; and B = number of tandem axles on one vehicle.

Thus, the accumulated equivalent traffic can be calculated as:

$$N_t = \sum_{l=1}^{t} \Delta n_l LEF \tag{5}$$

Where, Δn_l is traffic volume during time period l.

Concerning environment, it is found that in the AASHO Road Test pavement serviceability experienced three typical environment-relevant stages: normal, stable, and accelerated deterioration (Prozzi, and Mandanat, 2003). Thus, frost penetration gradient is applied to quantify environmental

changes within each year. As a result, the adjusted accumulated equivalent traffic after incorporating environmental effect is,

$$N_t^{'} = \sum_{l=1}^{t} \Delta n_l (1 + \psi G_l) LEF \tag{6}$$

Where, ψ = parameter to be estimated; and G_l = frost penetration gradient.

The final specification is to take into account of both cross sectional and time series information in the panel data. By integrating Equations (1) to (6), the final specification for the deterioration is,

$$p_{i,t} = \beta_{0i} + \beta_{1i} \exp\{\beta_{2i} H_{1i}\} - \exp\{\beta_{3i} + \beta_{4i} H_{i1} + \beta_{5i} H_{i2} + \beta_{6i} H_{i3}\} N_{i,t}^{\beta_{8i}} + \mu_{i,t} \tag{7}$$

$$N_{i,t} = \sum_{l=1}^{t} \Delta n_{i,l} (1 + \beta_{7i} G_{i,l}) LEF_i \tag{8}$$

$$LEF_i = \left(\frac{FA_i}{\beta_{9i} 18}\right)^{\beta_{11i}} + A_i \left(\frac{SA_i}{18}\right)^{\beta_{11i}} + B_i \left(\frac{TA_i}{\beta_{10i} 18}\right)^{\beta_{11i}} \tag{9}$$

Where, $\beta_{0i} \sim \beta_{11i}$ = parameters to be estimated. The parameters can be divided into two categories, random and fixed. First, those parameters considered evidently varying, and thus treated as random are described. As in RE model, the intercept term is treated as random, i.e. $\beta_{0i} = \beta_0 + \delta_0 v_0$. The implication of the deterministic term β_0 is the same as in the traditional approach: unknown but fixed value. v_0 is a standard normal random variable with its coefficient δ_0 representing the standard deviation of β_{0i}. Concerning pavement structures, except for the observed heterogeneity, the unobserved heterogeneity (such as construction quality) will produce different contribution to resist deterioration. Thus, the parameters for surface, base, and subbase layers are regarded varying among sections. The specifications for β_{4i}, β_{5i}, and β_{6i} are $\beta_{4i} = \beta_4 + \delta_4 v_4$, $\beta_{5i} = \beta_5 + \delta_5 v_5$, and $\beta_{6i} = \beta_6 + \delta_6 v_6$ respectively. Regarding traffic, both the parameters for converting front and tandem axle load into the equivalent 18-kip (80.1 kN) axle load are believed varying across pavement sections. This assumption is supported by the fact that load conversion factor is dependent on pavement structure (AASHTO 1993). Considering the load conversion factor is positive, lognormal distributions are adopted. Thus, by construction, $\beta_{9i} = \exp\{\beta_9 + \delta_9 v_9\}$, where β_9 and δ_9 are the mean and standard deviation of $\log\{\beta_{9i}\}$ respectively. Similarly, the specification for β_{10i} is $\beta_{10i} = \exp\{\beta_{10} + \delta_{10} v_{10}\}$. For the rest of "slope" parameters, no enough evidence is found to support their significant variability across pavement sections. Thus, for simplicity, those parameters are assumed fixed across sections: $\beta_{1i} = \beta_1$, $\beta_{2i} = \beta_2$, $\beta_{8i} = \beta_8$, and $\beta_{11i} = \beta_{11}$. If there is sufficient evidence to support the variation of these parameters, they can also be treated as random parameters. The implementation is straightforward. In addition, a partition form of parameter vector is adopted to facilitate the following discussion, $\underline{\beta}_i = \left\{[\underline{\beta}_i^R]', [\underline{\beta}_i^F]'\right\}'$, where,

$\underline{\beta}_i^R = [\beta_{0i}, \beta_{3i}, \beta_{4i}, \beta_{5i}, \beta_{9i}, \beta_{10i}]'$, including the random parameters, and

$\underline{\beta}_i^F = [\beta_{1i}, \beta_{2i}, \beta_{6i}, \beta_{7i}, \beta_{8i}, \beta_{11i}]' = [\beta_1, \beta_2, \beta_6, \beta_7, \beta_8, \beta_{11}]'$, including the fixed parameters.

Model Estimation Approach

It is shown in Equations (7) to (9) that the underlying model specification is featured highly nonlinear. To estimate such models with random parameters, the Maximum Simulated Likelihood (MSL)

approach is used. The log-likelihood is obtained as (see Greene, 2002, 2004 for details),

$$\ln L = \sum_{i=1}^{n} \ln\left\{ \int_{\underline{\beta}_i^R} \prod_{t=1}^{T_i} g(y_{it}, \underline{X}_{it}, \underline{\beta}_i^R, \underline{\beta}_i^F, \theta) f(\underline{\beta}_i^R) d\underline{\beta}_i^R \right\} \tag{10}$$

Where, y_{it} is performance observation on pavement section i at time t; $\underline{X}_{it}$ contains the explanatory variables; $\underline{\beta}_i$ denotes the parameters; θ is the standard deviation of the disturbance; $g(y_{it}, \underline{X}_{it}, \underline{\beta}_i, \theta)$ takes the form of normal distribution; $f(\underline{\beta}_i^R)$ is the joint density function of $\underline{\beta}_i^R$.

Equation (10) involves high-dimension integral, which is equivalent to and can be approximated by the simulated mean (Greene 2002, 2004). The simulated log-likelihood is obtained,

$$\ln L_S = \sum_{i=1}^{n} \ln\left\{ \frac{1}{M} \sum_{m=1}^{M} \prod_{t=1}^{T_i} g(y_{it}, \underline{X}_{it}, \underline{\beta}_i^{R,m}, \underline{\beta}_i^F) \right\} \tag{11}$$

Where, $\underline{\beta}_i^{R,m}$ is the mth draw from density $f(\underline{\beta}_i^R)$, totally M draws are adopted to simulate the integral.

The simulated log-likelihood is then maximized to arrive at the parameter estimates and their asymptotic standard errors. For the properties of MSL estimator, see Hajivassiliou and Rudd (1994).

Empirical Results and Implications

The estimation results by applying MSL are presented in Table 1, referred to as RP results. For comparison, the results obtained through OLS and RE approaches are also presented. It is shown by the t-statistics that among the three approaches, all of the parameters are significant at a 95% confidence level.

Table 1 also presents the log-likelihood values at convergence for the three alternatives, which can be used to perform hypothesis test based on likelihood ratio. The statistic is $-2(\ln L_R - \ln L_U)$, where $\ln L_R$ and $\ln L_U$ are log-likelihood values for restricted and unrestricted models. First, between OLS and RE models, the homogeneity of the intercept is evidently rejected since the calculated statistic, 3503.36 is much larger than the critical value χ_1^2 at 95% level. The subscript 1 means one degree of freedom. In the same way, the hypothesis of the unobserved heterogeneity only existing through the intercept leads to the rejection of the RE model, in favor of RP model.

β_{4i}, β_{5i}, and β_{6i} are parameters denoting the relative contributions to resist deterioration by surface, base, and subbase layers respectively. The analysis is focused on their mean values that represent their average relative contributions. The ratio $\beta_4 / \beta_5 / \beta_6$ represents the relationship between the contributions by surface, base, and subbase layers. With the subbase layer coefficient as the reference, these ratios are 2.1/1.1/1.0 in OLS, 2.8/1.1/1.0 in RE, and 4.4/1.3/1.0 in RP models. It is suggested that the relative contribution of surface layer is underestimated if the unobserved heterogeneity is not/insufficient addressed. The similar result is found for base layer. This finding leads to a profound implication in optimizing pavement thickness design. Assume the effective thickness of the subbase layer is one. It is shown that the effective thickness of surface layer is 33.3% higher in RE than OLS approaches while the increase is 104.8% in RP over OLS approaches.

As is aforementioned, the conversion factors for front and tandem axles load are assumed to have lognormal distributions. The mean of a variable with lognormal distribution with its location and scale parameters η and ζ in logarithm is $\exp(\eta + \zeta^2/2)$. Based on estimated parameters, the mean load conversion factors for front axle are 0.534, 0.518, and 0.480 in OLS, RE, and RP approaches respectively. Accordingly, the equivalent axle loads are 9.6 kip (42.7 kN), 9.3 kip (41.4 kN), and 8.6 kip (38.3 kN) in the individual models. In the same way, the equivalent loads for tandem axles are 33.1 kip (147.3 kN), 35.7 kip (158.7 kN), and 39.6 kip (176.2 kN) in the three models. Both results suggest that not/insufficient accounting for heterogeneity leads to overestimation of front axle load conversion to equivalent load but underestimation in the case of tandem axle load.

Table 1. Parameters estimation results

Variables	Parameters	OLS		RE		RP	
		Estimates	t-stat.	Estimates	t-stat.	Estimates	t-stat.
Initial Serviceability Related	β_0	4.572	70.7	4.483	88.8	4.398	61.1
	δ_0	–	–	0.352	27.9	0.276	28.9
	β_1	-1.895	-31.2	-2.336	-7.3	-1.399	-14.9
	β_2	-0.403	-10.4	-0.722	-6.5	-0.495	-5.7
Subgrade	β_3	-3.844	-28.7	-2.886	-24.7	-3.851	-35.1
Surface Layer	β_4	-0.189	-15.4	-0.283	-28.2	-0.625	-34.5
	δ_4	–	–	–	–	0.167	42.9
Base Layer	β_5	-0.101	-28.6	-0.113	-31.8	-0.189	-36.9
	δ_5	–	–	–	–	0.038	30.0
Subbase Layer	β_6	-0.089	-27.7	-0.103	-32.2	-0.139	-35.4
	δ_6	–	–	–	–	0.003	4.3
Environment	β_7	-3.109	-62.2	-2.860	-66.5	-2.729	-80.4
Curvature	β_8	0.479	35.8	0.465	46.1	0.688	62.1
Front Axle	β_9	-0.628	-15.3	-0.657	-18.5	-0.664	-17.0
	δ_9	–	–	–	–	0.289	9.4
Tandem Axle	β_{10}	0.609	53.4	0.686	45.0	0.676	61.4
	δ_{10}	–	–	–	–	0.050	8.8
Power	β_{11}	3.743	49.4	4.288	47.0	3.907	60.3
Log-likelihood		-5056.32		-3304.64		-1706.17	

Individual-Specific Parameters

The parameters in Table 1 provide a general view on pavement behavior at a population or average level. The model and estimated parameters can serve pavement management at network level. In addition, at the project level, pavement behavior of a specific structure is of more interest. Compared with the population level parameters, the subpopulation (section specific) level parameters are more informative and relevant to describe that specific pavement behavior. The individual-specific parameters can be obtained through simulation in conjunction with Bayesian theorem.

The means and standard deviations are available through MSL in the previous section (See Table 1, RP results). These estimates are used as the prior information for the subpopulation (specific pavement section *i*). The mean of random parameters for a subpopulation can be derived as (Greene, 2004),

$$\hat{E}\left(\underline{\beta}_i^R\right)=\frac{\frac{1}{M}\sum_{m=1}^{M}\prod_{t=1}^{T_i}\underline{\beta}^{R,m}\,g(y_{it},\underline{X}_{it},\underline{\beta}^{R,m},\underline{\beta}^{F})}{\frac{1}{M}\sum_{m=1}^{M}\prod_{t=1}^{T_i}g(y_{it},\underline{X}_{it},\underline{\beta}^{R,m},\underline{\beta}^{F})} \tag{12}$$

Where, M is the total number of draws in the simulation; m represents the mth draw.

Figure 1 presents an example of performance deterioration at a selected test section (Section 594). The dotted line, denoted as Pr_G, represents the mean prediction line based on the population parameter estimates, and the solid line, Pr_I, is obtained based on that site-specific parameters. It is displayed that

both prediction lines capture the deterioration fairly well. In addition, it is shown the solid line performs the prediction more precisely in that it's obtained based on its site-specific parameters.

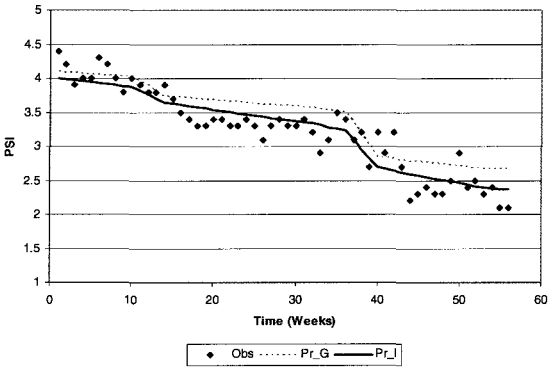


Figure 1. Observed and predicted (using population and individual estimates) performance

Conclusions

This paper investigates performance heterogeneity of transportation infrastructure based on experimental data. Simulated maximum likelihood is applied to estimate the nonlinear deterioration model. It is shown that the deterioration characteristics are well captured by the estimated model. Unobserved performance heterogeneity is proven to exist through hypothesis test. In addition, section-specific parameters are obtained through simulation. The implication for transportation infrastructure management policy is: pavement management at both network and project levels can be accommodated by the population-based and individual-specific parameters. The highlight of this paper is concerned with the flexibility of random coefficient model in transportation infrastructure management rather than the model specification per se. As a further step, the field data will be used in application.

References

AASHTO. (1993). AASHTO Guide for Design of Pavement Structures, Washington, D.C.

Al-Suleiman, T. I., Sinha, K. C., and Anderson, V. L. (1988). "Effect of routine maintenance on pavement roughness." *Transp. Res. Rec., 1205*, Transportation Research Board, Washington, D. C., 20-28.

Garcia, A. and Riggins, M. (1984). "Serviceability and distress methodology for predicting pavement performance." *Transp. Res. Rec., 997*, Transportation Research Board, Washington, D.C., 56-61.

Greene, W. (2002). Econometric Analysis. 5th Ed. Prentice Hall.

Greene, W. (2004). "Interpreting Estimated Parameters and Measuring Individual Heterogeneity in Random Coefficient Models." Department of Economics, New York University, working paper.

Highway Research Board (HRB). (1962). The AASHO Road Test. Special Report Nos. 61A through 61G and 73, Washington, D.C.

Madanat, S. M, Karlaftis, M. G, and McCarthy, P. S. (1997). "Probabilistic Infrastructure Deterioration Models with Panel Data." *J. Infrastructure Sys.*, ASCE, 3(1), 4-9.

Hajivassiliou, V. A., and Rudd, P. A. (1994). Classical Estimation Methods for LDV models Using Simulation, in Engle, R. F. and D. L. McFadden Handbook of Econometrics, Chapter 40, (4).

Prozzi, J. A. (2001). *Modeling Pavement Performance by Combining Field and Experimental Data.* PhD dissertation, University of California, Berkeley, CA.

Prozzi, J. A., and Madanat, S. M. (2003). "Incremental Nonlinear Model for Predicting Pavement Serviceability." *J. of Transp. Eng.*, ASCE, 129(6), 635–641.

Feasibility Study for Gray Theory Based Pavement Smoothness Prediction Models

Qiang Li [1], Kelvin C.P. Wang [1], Robert P. Elliott [1], Kevin D. Hall [1], and Yanjun Qiu [2]

[1]Civil Engineering, University of Arkansas, AR 72701, 479-575-8425; kcw@uark.edu

[2]Southwest Jjiaotong Uuniversity, Chengdu, China

Abstract

In the proposed Mechanistic-Empirical Pavement Design Guide (MEPDG) (NCHRP 1-37A, 2001), the functional performance indicator is pavement smoothness as measured by the International Roughness Index (IRI). The MEPDG IRI prediction models were developed based on the general hypothesis that changes in smoothness result from various distress types that can be predicted by the MEPDG program. Using pavement distress data from the Long Term Pavement Performance (LTPP) database, traditional regression analysis was used to statistically establish the MEPDG prediction equations. This paper attempts to use a new technique for pavement smoothness prediction. The gray system theory was devised in the 1980s for modeling uncertain systems with the characteristics of partially known information. A pavement performance prediction system can fit the domain of the gray system. The gray theory based prediction method is used in this paper to develop IRI prediction equations. With the data exported from the LTPP database, it is found that certain specific types of distresses significantly affect the accuracy of the predictions. After trial and error calculations, Gray Model based smoothness predictions are established using influencing factors similar to the ones used in MEPDG. Based on the comparisons of results from the two prediction methods with LTPP field data, it is shown that the Gray Model based method provides promising results and may be useful for modeling pavement performance.

INTRODUCTION

The International Roughness Index (IRI) is one of the most common methods of measuring smoothness of pavements. In the proposed Mechanistic Empirical Pavement Design Guide (MEPDG), the approach to IRI prediction is to begin with the initial, as-constructed IRI and then predict changes in IRI over time as a function of pavement distress development. The general hypothesis used in the smoothness model is that the various distresses resulting in differing changes in smoothness are represented by separate components within the model (NCHRP 1-37A, 2001). However, as the prediction models are based on linear regression statistics, this hypothesis may have limitations to predict the complex and unclear relationships in actual pavement

performance.

This paper introduces Gray Theory into pavement smoothness prediction. The gray system theory focuses on uncertain systems with the characteristics of partially known information (Deng, 1987). If a system's structure and operations are known and if the relationships among its variables are clearly defined, the system is called a white system. On the other hand, if nothing is known about the system, the system is a black system. A gray system is a system in-between with partially known information. Pavement smoothness prediction is a gray system where many parameters do not have analytical solutions, and limitations of using traditional statistical analysis are present. Therefore, the gray model is a suitable candidate in search for new prediction methods. Using the LTPP data of General Pavement Study (GPS) sections located in various traffic and environmental conditions, the gray model GM (1,N) (Liu,1991) of the gray system theory is employed in this paper to examine the potential for developing gray smoothness prediction models. Comparisons between GM models and MEPDG models are conducted.

DATA COLLECTION

In this study, the data used for smoothness model development are from the LTPP GPS database. Since there is not enough data for GPS-7 (AC overlay on PCC), data from this type of pavement are not included. Therefore, only the remaining six pavement types are used in the study. To be consistent with the MEPDG approach, the same distress parameters were adopted as the influencing factors for development of the GM models.

Two problems were encountered with the LTPP data. First, initial IRI values are not available for any of the GPS test sections. To compensate for this, initial values were estimated for each section using a linear extrapolation. Secondly, the distress and profile survey dates do coincide. This problem was addressed by merging data collected within 3 months as a single data set and not using data separated by more than 3 months.

VERIFICATION OF THE NEW DESIGN GUIDE MODELS

Using LTPP data with MEPDG prediction models, this study shows the presence of certain distresses sometimes have negative effect in IRI prediction accuracy. It is likely that better correlations among distresses and IRI may be needed in the model to improve the predictions.

Gray theory based relational analysis (Liu, 1991) can be used to inspect the correlations between IRI and the influencing factors. From this study's analysis using Gray Theory, the relational coefficients of initial IRI (IRI_0) rank first among all the parameters for all pavement types. This means that future smoothness is significantly related to the initial IRI for all six pavement types and that pavements constructed smoother will generally remain smoother over time. On the other hand, it is also found

that some of the parameters used in MEPDG smoothness prediction models do not have robust correlations with IRIs.

GRAY THEORY BASED PAVEMENT SMOOTHNESS PREDICTION

The GM(1,N) model, which describes the relationship between one main factor, and all other (N-1) influencing factors in a system, is used in smoothness model development. Assume the original data series of main factor, pavement IRI in this paper, is designated as $x_1^{(0)}(k) = (x_1^{(0)}(1) \quad x_1^{(0)}(2) \quad \cdots \quad x_1^{(0)}(k))$. The original data series of (N-1) influencing factors, distress types in this study, are designated as $x_i^{(0)}(k) = (x_i^{(0)}(1) \quad x_i^{(0)}(2) \quad \cdots \quad x_i^{(0)}(k)) \quad (i = 2,3,\cdots,N)$, where k presents the sample number. The analysis procedure of GM (1, N) model is as follows:

(1) Build up the original data series

$$x^{(0)} = \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \\ x_3^{(0)} \\ \vdots \\ x_n^{(0)} \end{bmatrix} = \begin{bmatrix} x_1^{(0)}(1) & x_1^{(0)}(2) & \cdots & x_1^{(0)}(k) \\ x_2^{(0)}(1) & x_2^{(0)}(2) & \cdots & x_2^{(0)}(k) \\ x_3^{(0)}(1) & x_3^{(0)}(2) & \cdots & x_3^{(0)}(k) \\ \vdots & \vdots & \cdots & \vdots \\ x_n^{(0)}(1) & x_n^{(0)}(2) & \cdots & x_n^{(0)}(k) \end{bmatrix} \tag{1}$$

(2) Build up Accumulate Generation Operation (AGO) sequences

$$x_i^{(1)} = AGO x_i^{(0)} = \left[\sum_{k=1}^{1} x_i^{(0)}(k) \quad \sum_{k=1}^{2} x_i^{(0)}(k) \quad \cdots \quad \sum_{k=1}^{n} x_i^{(0)}(k) \right] \tag{2}$$

(3) Develop the GM (1, N) model

Deng (1987) gave the solution to the gray model equation for estimation of the accumulated data series:

$$\hat{x}_1^{(1)}(k+1) = \left[x_1^{(0)}(1) - \frac{1}{a} \sum_{i=2}^{n} b_i x_i^{(1)}(k+1) \right] e^{-ak} + \frac{1}{a} \sum_{i=2}^{n} b_i x_i^{(1)}(k+1) \tag{3}$$

Where, coefficients metrics,

$$\hat{a} = (a, b_2, b_3, \cdots, b_n)^T = (B^T B)^{-1} B^T Y \tag{4}$$

$$B = \begin{bmatrix} -0.5(x_1^{(1)}(1) + x_1^{(1)}(2)) & x_2^{(1)}(2) & \cdots & x_n^{(1)}(2) \\ -0.5(x_1^{(1)}(2) + x_1^{(1)}(3)) & x_2^{(1)}(3) & \cdots & x_n^{(1)}(3) \\ \vdots & \vdots & \vdots & \vdots \\ -0.5(x_1^{(1)}(n-1) + x_1^{(1)}(n)) & x_2^{(1)}(n) & \cdots & x_n^{(1)}(n) \end{bmatrix} \quad (5)$$

$$Y = \begin{bmatrix} x_1^{(0)}(2) \\ x_1^{(0)}(3) \\ \vdots \\ x_1^{(0)}(n) \end{bmatrix} \quad (6)$$

(4) Obtain the estimations of the original data series

$$\hat{x}_1^{(0)}(k+1) = \hat{x}_1^{(1)}(k+1) - \hat{x}_1^{(1)}(k) \quad (7)$$

(5) Calculate the estimation error criteria: Average (Ave.), percentage Error (%e) and Root-Mean Square Error (RMS)

$$\%e = \frac{x_1^{(0)} - \hat{x}_1^{(0)}}{x_1^{(0)}} \times 100\% \quad (8)$$

$$RMS = \sqrt{\frac{\sum_{k=1}^{n}\left[\hat{x}_1^{(0)}(k) - x_1^{(0)}(k)\right]^2}{n}} \quad (9)$$

The values of coefficients reflect the relationships among the main factor and the influencing factors (Liu, 1991). The larger the value is, the closer the relationship is. If any developed coefficient is negative or too small, it implies that the relationship is negligible, or its contribution to smoothness has been considered by other parameters. After trial-and-error calculations, based on 40 data sets of each pavement type and following the GM(1, N) model developing procedure, the GM prediction equations (10) to (15) are listed below. The designations in the equations have the same meanings as those in the MEPDG design guide. Due to the length limitation, definitions of parameters are not given in this paper.

GPS1--Conventional flexible pavement with thick granular base

$$\hat{IRI}^{(1)}_{(k+1)} = 1.298e^{-0.2571k} + 3.8895(1 - e^{-0.2571k})[0.05934IRI_0^{(1)}{}_{(k+1)} + 0.0007594(SF \times e^{(age/20-1)})^{(1)}{}_{(k+1)} + 0.00003874(TC_L)^{T(1)}{}_{(k+1)} + 0.004652(FC)_T^{(1)}{}_{(k+1)}] \quad (10)$$

GPS2AT--Deep strength pavements (Asphalt treated base)

$$\hat{IRI}^{(1)}_{(k+1)} = 1.456e^{-0.1710k} + 5.848(1-e^{-0.1710k})[0.1525IRI_0{}^{(1)}_{(k+1)} + 0.00001226(FI)^{(1)}_{(k+1)} + 0.0127(FC)_T{}^{(1)}_{(k+1)} + 2.7479(\frac{1}{(TC_S)_H})^{(1)}_{(k+1)}] \quad (11)$$

GPS2SR--Semi-Rigid Pavements

$$\hat{IRI}^{(1)}_{(k+1)} = 0.789e^{-0.1314k} + 7.6104(1-e^{-0.1314k})[0.1720IRI_0{}^{(1)}_{(k+1)} + 0.03419(SD_{RD})^{(1)}_{(k+1)} + 0.00002642(TC_L)_T{}^{(1)}_{(k+1)}] \quad (12)$$

GPS3--Jointed Plain Concrete Pavement (JPCP)

$$\hat{IRI}^{(1)}_{(k+1)} = 3.38e^{-0.7473k} + 1.3382(1-e^{-0.7473k})[0.8370IRI_0{}^{(1)}_{(k+1)} + 0.0009929(SPALL)^{(1)}_{(k+1)} + 0.001203(TFAULT)^{(1)}_{(k+1)}] \quad (13)$$

GPS5--Continuously Reinforced Concrete Pavement (CRCP)

$$\hat{IRI}^{(1)}_{(k+1)} = 0.93e^{-0.4194k} + 1.3243(1-e^{-0.4194k})IRI_0{}^{(1)}_{(k+1)} \quad (14)$$

GPS6--HMA overlay of flexible pavement

$$\hat{IRI}^{(1)}_{(k+1)} = 0.903e^{-0.1467k} + 6.8166(1-e^{-0.1467k})[0.1177IRI_0{}^{(1)}_{(k+1)} + 0.01187(Age)^{(1)}_{(k+1)} + 0.02069(FC)_T{}^{(1)}_{(k+1)}] \quad (15)$$

RESULTS, COMPARISONS AND ANALYSIS

Table 1. Comparisons of IRI predictions from MEPDG and GM models

GPS1	# Sites	Ave.	% e	RMS	GPS3	# Sites	Ave.	% e	RMS
Actual data	57	1.686			Actual data	140	1.832		
MEPDG	57	1.332	17.0	0.641	MEPDG	140	1.908	-4.70	0.448
GM	57	1.658	0.30	0.582	GM	140	1.876	-3.90	0.481
GPS2AT	# Sites	Ave.	% e	RMS	GPS5	# Sites	Ave.	% e	RMS
Actual data	56	1.637			Actual data	49	1.402		
MEPDG	56	2.049	-25.7	1.167	MEPDG	49	1.511	-7.80	0.205
GM	56	1.571	0.40	0.725	GM	49	1.682	-20.4	0.341
GPS2SR	# Sites	Ave.	% e	RMS	GPS6	# Sites	Ave.	% e	RMS
Actual data	83	1.622			Actual data	54	1.903		
MEPDG	83	1.281	15.9	0.643	MEPDG	54	2.666	-53.0	1.239
GM	83	1.735	-13.0	0.529	GM	54	1.896	9.00	0.693

Except for the 40 data sets for model development, other available data sets are used for validation of the models. The error analysis is presented in Table 1 for actual data, MEPDG predictions and GM predictions.

Although the GM models use fewer distress parameters than do the MEPDG models, the GM models appear to produce better predictions. For GPS1, as an example, the predictions from MEPDG equation estimate the IRI on the average 17.0% smaller than the actual field data. In comparison, GM model predictions on the average differ from the field data by only 0.3%. For the GPS2AT and GPS6, the MEPDG predictions overpredict the field data by averages of 25.7% and 53.0%, while the GM method has average prediction errors of only 0.4% and 9.0%.

The lone exception to this trend is the GPS5 model. The GM model for GPS5 data set did not predict IRI as well as the MEPDG model. The gray relational analysis indicated that the correlations between smoothness and the influencing parameters for these pavements was not good. Therefore, only initial IRI is used in GPS5 GM model simply to illustrate the complexity of smoothness prediction. This is possibly due to variable qualities of the LTPP data (IHRB, 2005).

CONCLUSIONS

This paper presents an initial feasibility study on the use of gray theory based methodology to develop smoothness prediction models. Using the same distress parameters as used in the proposed MEPDG guide, LTPP data of six pavement types are exported and analyzed. From the gray relational analysis results, the influencing factors adopted for GM model developments are selected among the existing factors in the MEPDG; GM models of the six pavement types are then established. From the comparisons of the predictions originated from MEPDG equations and GM models with the field LTPP data, the GM models show promising results and better predictions. It is anticipated that future research will center on further data validation of the GM models with more LTPP data sets and correlation studies among influencing factors.

REFERENCES

Deng Julong (1987). "The basic method for gray system." *Huzhong University of Science & Technology Press*, Wuhan, HUBEI, P. R. China.

IHRB Project TR-509 (2005). "Implementing the mechanistic-empirical pavement design guide: technical report." *Iowa State University*, Ames, IA

Liu Sifeng (1991). "An Introduction to Gray System: Foundation, Methodology and Application." *Henan University Press*, Kaifeng, Henan, P. R. China.

National Cooperative Highway Research Program 1-37A(2001). "Guide for Mechanistic-Empirical design of new and rehabilitated pavement structures." *Transportation Research Board*, Washington D.C.

Automated Tracking System for Inventory of Roadway Signs

Kelvin C.P. Wang, Zhiqiong Hou, and Weiguo Gong

4190 Bell, University of Arkansas, Fayetteville, AR 72701, Email: kcw@uark.edu

Abstract

Roadway signs represent a substantial investment of public money in road and highway infrastructure. However, the current level of automation in sign identification and recognition, size dimensioning, and location identification is not satisfactory. This paper targets the development of an automated road inventory system addressing these issues. A framework of combining the conventional image processing methods with the Kalman filter tracking method is applied to improve the accuracy and efficiency of the ROW image processing. Through the tracking technique, the candidate region of the road sign in the picture can be predicted based on the previous image frame. Detection efficiency and accuracy of the sign can be improved. The methodologies described in the paper fit a dynamic and motion environment, appropriate for a highway survey vehicle.

INTRODUCTION

Recognizing road sign from the ROW images or videos frames have been studied for some time. Common techniques for detecting and recognizing road signs are algorithms based on color information and shape information (Yulie et al., 1998), template matching (Escalera et al., 1997) and techniques of Neural Network (Liu et al., 2001). However, traditional methods based on imaging processing would search road sign in every image frame. This method costs too much computing resources and is therefore too expensive for the real time, or high speed processing. It is even more inefficient when many road signs are in one image frame or a series of image frames. Therefore, in this paper tracking technique is investigated as a solution to the problem of improving reliability and computing efficiency of algorithm implementation. This will help improve the automation level of the road sign inventory system now available in US such as techniques in Lambda Tech, Transmap Corp. and Geo-3D Corp.

So far, only a few publications addressed algorithms of using tracking technique in the road sign detection. Fleischer et al. (2002) and Fang et al. (2003) applied Kalman filter in the road sign recognition. However, the stability of the Kalman model in (Fang et al., 2003) is not verified because the formation of the transition matrix has a trend to be ill conditioned. In addition, both of Fang et al. (2003) and Fleischer et al. (2002) focus on the automated navigation and driving. The model presented in this paper is specifically designed for automated sign inventory. It simplifies the transition matrix and provides more explanation on the modeling and pretest.

SOLUTION METHODOLOGY

This paper proposes a framework that can incrementally detect road signs from a sequence of digital video frames. The general algorithms of the framework are as follows.

1. Detect the candidate road sign region with specific color threshold and geometrical properties of shape.
2. Track the candidate region of road sign and predict the potential region in the following frame using a Kalman filter.
3. Micro-scan the predicted region by image processing when the region is big enough and distinctive.
4. Global Positioning System (GPS) positioning of the road sign.

Candidate region detection: The purpose of candidate region detection is to find all possible road sign regions. These candidate regions will be used as the start point to the tracking. This candidate region is detected by image processing. Specific color threshold is first used to segment the image into meaningful regions. Then the aspect ratio and compactness of the segmented images are evaluated to find the possibility of the region to be a candidate road sign region.

Kalman filter tracking: First, a road sign is determined in the current frame, which is accomplished in the previous step of candidate region detection. By computing the geometric relationship of the road sign between adjacent frames, the framework finds the candidate region for the road sign in the next frame. The image processing step analyzes the candidate region in detail when it comes to the final stage, which means the sign is right before out of view in the image.

The Kalman filter technique includes two phases: time update and measurement update (Kalman, 1960). The time update procedure is based on the dynamic equation which is derived from the spatial constraints from the two successive frames. The measurement update is based on the image processing location in the proximity of the predicted candidate region.

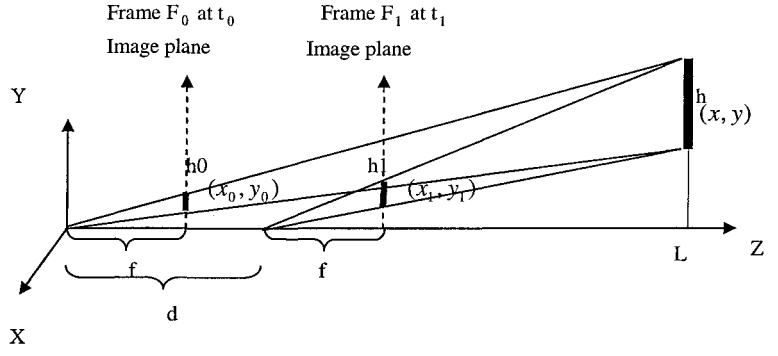


Figure 1. Spatial constraints in two successive frames.

As shown in the Fig. 1, camera coordinate system at time t_0 is taken as the basic coordinate system, Z-axis is the camera optical axis and the X-axis is parallel to the vehicle's horizontal plane. f is the camera focus length. The camera moves the distance d in the traveling direction, which is measured from the on-board Distance Measurement Instrument (DMI).

Road sign size prediction:

Assume h is the actual height of the road sign, h_0, h_1 are the heights of the road sign in two successive frames shown in Fig. 1. Based on triangular principle, the relationship can be found as

$$h_1 = \frac{1}{1-\frac{dh_0}{hf}} \cdot h_0 \qquad w_1 = \frac{1}{1-\frac{dw_0}{wf}} \cdot w_0 \tag{1}$$

Road sign location prediction:

Similarly, the relationship can be formed as:

$$y_1 = \frac{1}{1-\frac{dh_0}{hf}} \cdot y_0 \qquad x_1 = \frac{1}{1-\frac{dw_0}{wf}} x_0 \tag{2}$$

Here, x, y, h, d, and f are constants. Based on the above relationships, the model is constructed as follows.

$$\text{the location state } X = \begin{bmatrix} x_t \\ y_t \end{bmatrix}, \text{ the size state } Y = \begin{bmatrix} h_t \\ w_t \end{bmatrix} \tag{3}$$

$$\text{dynamic equation } x_t = Ax_{t-1} + \omega_t \tag{4}$$

where ω_t represents system perturbation which comes from the drive direction variance, DMI measurement error, etc.

Transition matrix A

$$A = \begin{bmatrix} \frac{1}{1-\frac{d \cdot h_{t-1}}{h \cdot f}} & 0 \\ 0 & \frac{1}{1-\frac{d \cdot w_{t-1}}{w \cdot f}} \end{bmatrix} \tag{5}$$

It is further assumed that the image processing estimates

$$z_t = Hx_t + v_t \tag{6}$$

where v_t represents measurement uncertainty.

Measurement matrix H:

$$H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{7}$$

The test conducted in the research shows that the noise of the location state $[x_t, y_t]$ is much higher than the size state $[h_t, w_t]$. On the other hand, the smaller dimension matrix will have lower possibility to be ill conditioned. These are the two reasons why the location state and size state were modeled separately. Based on our test, this modification is superior to the formation of the transition matrix in Fang et al. (2003) in terms of the model stability and convergence.

Recognition: As the candidate region is tracked when the vehicle moves toward the sign and size of the sign is increased, it will come to a stage that the sign in the picture is obvious and the disturbance from the background is insignificant to influence the outcome. The candidate region is matched by template before it is out of view in the camera, that is, in the last frame in the tracking process. Five series of templates are used in pixel: 120×120, 100×100, 80×80, 60×60, 40×40. For each size template, three levels of contrasts are used. Once the road sign is recognized in the template matching, the tracking is marked as successful.

GPS Positioning: The last step is mapping the detected road sign using GPS and Gyroscope (GYRO) information. Depending on the GPS receiver, the positioning accuracy varies. The data from GPS receiver is adjusted by local GYRO information. The integration of the GPS receiver and GYRO sensor provide more accurate location information for the detected road sign than just using the GPS data alone.

EXPERIMENT

The images were captured by the Right-of-Way (ROW) images system in the Digital Highway Data Vehicle (DHDV) which was developed by the research team at the University of Arkansas since mid-1990's. DHDV is a multi-functional highway data collection and analysis platform with pavement surface imaging system, longitudinal roughness laser sensors and accelerometers, laser based rutting measurement device, GPS receiver, DMI, and Gyro sensor. The software system used in the on-board computers of the DHDV employs real-time relational database engine, inter-computer communication techniques, multi-computer and multi-CPU based parallel computing, real-time control of digital sensors, and the generation of multimedia databases.

For right-of-way imaging, there are two digital color frame cameras mounted on top of the DHDV. Each camera is at the resolution of 1300 by 1024. The imaging system can record and archive one frame color image from the camera at user-determined interval up to one frame per 10 feet. In this experiment, one camera is used for the research.

STATISTICAL TEST AND CALIBRATION

The Q and R term in the Kalman filter model are determined based on the statistical test. The error of the dynamic model Q will come from the instrument (the vehicle velocity and direction, DMI and other instrument). Q is obtained by taking photos for the same road sign from the same distance 20 times. The error covariance of the location state vector is first calculated from the 20 images.

$$e = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \tag{8}$$

Then substitute to equation (9) to obtain the Q term.

$$Q = e^T \cdot e = \begin{bmatrix} e_1 & e_2 \end{bmatrix} \cdot \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} e_1 e_1 & e_2 e_1 \\ e_1 e_2 & e_2 e_2 \end{bmatrix} \tag{9}$$

The Q term for the size state vector is obtained in the same way. Measurement error covariance term R is determined from the success probability in image processing which is based on correct/false detection ratio. The error of the measurement (image processing) comes from the accuracy of the detection algorithms. In this test, it is set to be very small since the measurement (image processing) in the predicted limited region is relatively accurate.

Another important thing is to calibrate the effect focal length. The focus of the camera lens is 11-mm in this experiment. The effective focal length is calibrated by using the known distance L and known road sign size h in the 20 pictures. H is the physical height of the stop sign. The effective focal length of the camera is the average value of the calibration based on the 20 images.

$$f = \frac{hL}{H} \tag{10}$$

EXPERIMENT RESULT

The determined Q, R and f are input parameters in the system. In the experiment, the size of the image is 1300 x 1024, the DMI distance, GPS latitude, longitude, altitude, Gyro information are given for each image. The interval distance between two successive images is obtained from DMI measurement. The GPS information, current state vector and tracking status are displayed instantly while the vehicle is in motion. The ROW images are processed in real time.

Five road tests are conducted. The results in Table 1 show the accuracy based on the Kalman filter tracking algorithm. Table 1 also lists the false and success detection of the road sign in the five road tests. The majority of the road sign can be detected and tracked accurately. The successful detection rate is about 95%. The results also show that the tracking technique is effective in automated inventory of road signs, overcoming difficulties in using traditional imaging techniques for multiple targets.

TABLE 1 Performance of the Stop Sign Detection

Test	#Frames	#Stop Signs	#Correct Tracking	#False tracking
1	8324	35	33	1
2	12549	47	44	0
3	6123	23	22	0
4	2678	15	12	1
5	9082	38	36	0

ACKNOWLEDGEMENT

The authors would like to thank Mr. Alan Meadors, head of Research and Planning of the Arkansas State Highway and Transportation Department (AHTD) for his vision in initiating this project. Mr. Ron Strickland of AHTD provided technical and administrative assistance for this project. Mr. Webb Winston, Mr. Nguyen Son, and Mr. Matt Jones of the University of Arkansas provided various support in data collection of the ROW images.

CONCLUSION

There have been limited published literature on automated inventory of roadway signs, despite the fact that there were several attempts in recent years in the ITS arena to detect sign for the purpose of automated driving and navigation. A methodology for automated inventory of road sign by image analysis and Kalman filter tracking with ROW images has been tested successfully in this research. The algorithms have four main parts, the detection, tracking, recognition and GPS mapping. The detection is based on color and shape information. The Kalman filter tracking algorithm makes the detection more robust and efficiency. The recognition is accomplished by template matching. The GPS position system provides positioning data for road sign. It is also demonstrated that the techniques can successfully track and detect multiple signs in the same sequence of images. Based on our initial experimentation with the algorithms and models presented in the paper, the percentage of falsely identified signs is very low. It is hopeful that comprehensive and fully automated inventory of roadway signs can be realized with more signs built into the database.

Reference

Yulie, A.L., et al., Using color to detect, localize and identify informational signs. *International Conference on Computer Vision (ICCV98)*, Bombay, India. 1998, pp. 628-633.

Escalera, A., Moreno, L., et al., Road Traffic Sign Detection and Classification, in *IEEE Transactions on Industrial Electronics*, Vol. 44, 1997, pp.848-858.

Liu, X. H., Ran, B., Vision-Based Stop Sign Detection and Recognition System for Intelligent Vehicle, *TRB Annual Meeting* , 2001.

Fleischer, K. et al., (2002), 3D-Model-Based-Vision for innercity Driving Scenes, *IEEE Intelligent Vehicles Symposium (IV'2002)*, Versailles, France, 2002, pp. 477-482.

Fang, C.Y., Chen, S. W. and Fuh, C. S., Road Sign Detection and Tracking. *IEEE Trans. On Vehicular Technology*, Vol. 52, No. 5, 2003, pp.1329-1341.

Kalman, R. E. A New Approach to Linear Filtering and Prediction Problems. *Transaction of the ASME—Journal of Basic Engineering*, 82(Series D), 1960, pp. 35-45.

Scheduling Bridge and Highway Inspection/Test Activities with QUALITIME

Mireille Battikha[1]

[1] Consultant, 90 Gerrard West, Toronto, ON, M5G 1J6

Abstract

This paper reports on the assessment of the effectiveness of an add-on software tool named QUALITIME, which is conceived to schedule and represent quality activities in a graphical mode. QUALITIME is expected to integrate with existing scheduling software tools such as Microsoft Project and Primavera, and provide an environment to schedule quality activities based on the frequency and timing of their occurrence, and in connection with schedules of construction activities. Its use will render the management integration between construction and quality more beneficial, and will furnish more effective automated schedules for improved analyses and decision-making processes. The assessment is performed by analyzing outcomes of case studies from the highway and bridge domains, in which QUALITIME capabilities are utilized to schedule inspection/test activities. The major criteria for assessment include: (1) the needs and benefits to schedule quality activities, to computerize their information and tasks, and to represent them graphically and in association with construction schedules; (2) the determination of the relevant information to be captured by the tool; (3) the definition of the tasks to be performed; and (4) the implications of the consideration of the quality activities, as distinct time entities, on the duration of construction schedules. Refinement of the software tool design will be carried out based on the results of the case studies.

Introduction

The main purpose of construction schedules is to determine project completion. Consequently, contractors can then modify crew sizes, shifts, or equipment to speed up or slow progress. Schedules are also used to determine activities start and finish dates, which are needed to arrange for material, tools, and equipment. Other uses include resources control and cash flow studies. In addition, construction schedules are employed to evaluate changes, and perform project monitoring and control, as well as delay analyses. Schedules of construction activities are generally performed using the Critical Path Method and represented graphically with Primavera, Microsoft Project, or other software tools. These tools are designed to handle scheduling techniques and represent construction activities graphically based on their durations and relationships. In the quality management domain, inspection/test activities are performed at intermittent periods during, before, or after a construction activity. These quality activities are assigned frequency and timing requirements to be accomplished by trained personnel (Battikha 2003).

With current increasing demands for quality systems implementation in construction, there is a need for effective and efficient scheduling of quality activities. In an attempt to improve the planning and scheduling of quality activities, an add-on software tool has been conceived to schedule and represent inspection/test activities in a graphical mode (Battikha 2005). The tool, named QUALITIME, is designed to integrate with existing scheduling software tools such as Microsoft Project and Primavera and to schedule inspection/test activities based on their frequency and timing, and in association with schedules of construction activities. Its use will render the management integration between construction and quality more beneficial, and will furnish more effective automated schedules for improved analyses and decision-making processes. This paper reports on the assessment of the effectiveness of QUALITIME. The assessment is performed by analyzing outcomes of case studies from the highway and bridge domains, in which QUALITIME capabilities are utilized to schedule inspection/test activities. The major criteria for assessment include: (1) the needs and benefits to schedule quality activities, to computerize their information and tasks, and to represent them graphically and in association with construction schedules; (2) the determination of the relevant information to be captured by the tool; (3) the definition of the tasks to be performed; and (4) the implications of the consideration of the quality activities, as distinct time entities, on the duration of construction schedules. Refinement of the software tool design will be carried out based on the results of the case studies.

Previous research has been directed at developing a computer-based system (QUALICON) to assist quality management teams, in the construction industry, to deal with information and consequent decision-making processes pertaining to nonconformance of construction projects, for (1) the detection of problems and/or their prediction; (2) the diagnosis of their root causes; and (3) the specification of appropriate remedial, corrective and/or preventive actions. A set of quality management tasks has been identified for automation and the first phase of its realization achieved. The realized phase of the system is designed to integrate with existing computer-aided project management functions, and can assist management in (1) the definition of requirements/criteria for design, construction, and quality management; (2) the development of inspection and test plans; (3) the tracking of actual inspection/test results; (4) the verification of their conformance to defined criteria; (5) the documentation of past experience in the form of standard templates for assisting the tasks involved; and (6) the generation of reports (Battikha 2002b). The second phase of the system realization treats the reasoning tasks pertaining to (1) the problem identification and/or prediction; (2) the diagnosis of their causes; and (3) the specification of appropriate actions (Battikha 2002a).

QUALICON has addressed the flexible representation of the inspection/test requirements to include their frequency and timing. Also, QUALICON tasks are enhanced by the use of templates devised to enable the reuse of predefined packages of information from project to project. These templates allow the user to define a single or a set of reusable requirements/criteria, as well as inspection and test plans, which are stored in the system *Standards*. These requirements/criteria, and inspection and test plans, can then be selected and copied to a project file. This system capability

dramatically reduces the amount of time and effort necessary to input, in a new project, the large amount of information related to quality requirements/criteria and inspection and test plans. *Criteria*, as defined by QUALICON, are requirements to which inspection/test results must conform, and upon which an inspection/test passes or fails. The inspection and test plans and their respective collections of inspections/tests (together or separated), can be added, deleted, or edited by the user. It is suggested, yet not mandatory, that relevant information documented in QUALICON be imported to QUALITIME, at the user's discretion, to facilitate the input of information needed for scheduling inspection/test activities. The availability of this information in a computerized environment to be imported to QUALITIME will be very beneficial to the user.

Outcomes of Case Studies

Case studies from the bridge and highway domains have been collected from documented inspection manuals and quality assurance programs revealing the practical use of quality activities schedules (MoTH 1992; MoTO 1996). Selected activities of the cases have been analyzed to assess QUALITIME capabilities. Bridges and highways have been adopted for the study because their quality is crucial. Also, these domains are well defined and representative of other construction domains. However, all development steps have been considered for their applicability to any type of construction in order to preserve a generic aspect of QUALITIME. Table 1 presents examples of an inspection scheme for concrete control that are representative of precast concrete bridges (Fib 2004). Table 2 provides an example schedule of testing for highways, which is also used in the assessment (Asphalt Institute 1989).

Table 1. Example of Inspection/Test Activities for Precast Concrete Bridges

Inspection/Test	Method	Purpose	Frequency/Timing
Mixture – composition (except water content)	- Visual on weighing equipment - Checking against production documents	Conformity with intended production	- Daily for each composition used - After each change
Water content of fresh concrete	Appropriate method	To provide data for the water/cement ratio	- Daily for each composition used - After each change - In case of doubt
Chloride content	Calculation	To ensure that the maximum chloride content is not exceeded	In case of increase of the chloride content of the constituents
Water/cement ratio of fresh concrete	Calculation according to specific standard	To assess specified water/cement ratio	Daily, if specified
Air content of fresh concrete	Test according to specific standard	To assess conformity with specified content of entrained air	First batch of each production day until values stabilize
Concrete mix	Visual check	Correct mixing	Daily for mixer
Potential concrete strength	Testing according to specific standard	To assess conformity with intended value	Daily for each type of concrete
Density of hardened concrete	Testing according to specific standard	To assess specified density	As frequently as potential strength

Table 2. Example Schedule of Sampling and Testing for Highways

Sample of	Minimum Sample Frequency	Test to be Performed	Test Method Designation
Hot Aggregate, Combined	2 Daily	Sieve Analysis	AASHTO T-11 AASHTO T-27
Mineral Filler	As Necessary	Sieve Analysis	AASHTO T-37
Asphalt	2 for each transport delivered	Send to Central Laboratory	AASHTO T-40
Uncompacted Mix	2 Daily	Density Stability	AASHTO T-209 (ASTM D 2041) Project Specification Requirements

Other expressions than the aforementioned ones listed in Tables 1 and 2 have been used in the case studies to define frequency and timing of inspection/test activities. They include "as soon after original construction as possible", "after impact by a vessel", "after a flood event", and "in exceptional ice conditions". The analyses of the cases based on the four assessment criteria indicated that the vast amount of inspection/test activities assigned on a project justifies the needs and benefits to handle and schedule inspection/test activities in a computerized environment. Representing inspection/test activities graphically and in association with construction schedules enhances the management integration of construction and quality. The determination of the relevant information to be captured by the tool and the definition of the tasks to be performed, have been refined according to the common requirements derived from the cases. In addition, the consideration of the quality activities, as distinct time entities, on the duration of construction schedules can be beneficial by providing the option to the user to select whether to merge it with the construction activity or to separate it, as discussed in the following section.

Assessment of QUALITIME

Based on several iterative design trials, samples of the interfaces, tasks, and information to be captured by the tool have been developed and populated with examples of selected activities of cases from the bridge and highway domains. These samples have been assessed and refined based on the outcomes of the cases. Figure 1 captures the refined version of the user input interface to schedule inspection/test activities with QUALITIME. Users (e.g., management personnel, inspectors) can access the quality chart/schedule via the construction scheduling software tool, after installing QUALITIME. Once the construction activity is defined a series of inspection/test activities can be named, their criteria identified, and their stage determined. Then if applicable, they can be timed with relevance to the construction activity. Consequently, their frequency, unit, and other constraints (date, additional timing descriptions, durations) can be assigned. Inspection/test activities can then be scheduled and represented graphically in connection with the relevant construction schedule. Once displayed, the graphical representation of the inspection/test activity can be used to sort the activities by stage and timing. Conversely, the activities with the same stage and timing can be defined by a graphical input and their definition would then follow. It should be noted that the duration of the inspection/test activity

is usually negligible compared to that of the construction activity. However, if the duration of the inspection/test activity is significant, an optional and additional inclusion of its schedule/representation is provided within the construction activity or separately, to allow a flexible capability, in response to the practitioner mode of selection. For example, if a test requires three days to be completed, it will have a frequency and timing like any other inspection/test activity. In addition, an option is provided to consider the three days either within the construction activity, thus making it longer, or separately, showing the three days duration of the test.

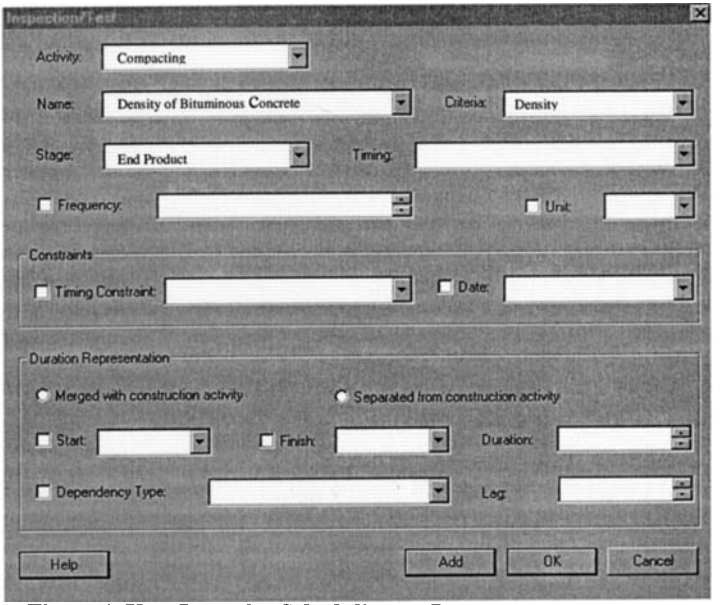


Figure 1. User Input for Scheduling an Inspection/Test Activity

The stage and timing allocate the activity to the chart on the screen. Additional constraints are used for reporting details. The stage classifies the criteria and/or the inspection/test activity into in-process, intermediate product, end product, and performance. The timing constraint is allocated for the same inspection/test activity to further define its point in time. The frequency is also used for the information report. Frequency is not necessarily numerical, for example when the expression "in case of doubt" is used. Furthermore, shape, color and pattern selection is flexible and changes are allowed to facilitate and enhance the graphical representation of the inspection/test activity. The size of the representation can also be modified to suit the screen. A graphical input is initiated and followed by a dialog box to allow the user to perform these manipulations. Graphical interaction or report generation from a menu can be furnished to identify the list of inspection/test activities to be performed on a certain day. Details about the chart representation of the inspection/test activities in association with construction activities can be found in Battikha (2005).

Conclusions

This paper has summarized the assessment results of an add-on software tool to schedule and represent inspection/test activities based on their frequency and timing and in association with construction schedules. Based on the outcomes of the case studies, refinement of the tool has been carried out using several iterative design trials. QUALITIME offers flexibility in integrating inspection/test activities with construction schedules. This allows the quality activities to maintain their originally specified frequency, timing, and association with their corresponding construction activities, in case there are changes in the construction schedules. The representation of an inspection/test activity with a significant duration, similar to the representation of a construction activity, could be provided to the user in a dialog box as an advanced information input, in order to keep the interfaces simple and easy to use. The location of the inspection/test activity and the component to which the activity is related could also be considered in further refinement. Final testing and validation of QUALITIME will be performed in future research, along with studies to instill and integrate its use in practice, by seeking opinions from industry professionals, and contemplating re-engineering the management tasks and using hand-held computers.

References

Asphalt Institute. (1989). "The Asphalt Handbook. Manual Series No. 4 (MS-4)." *The Asphalt Institute*, Lexington, Kentucky.

Battikha, M. (2005). "Scheduling Quality Activities." *Proc., 6th Construction Specialty Conference* (CD-ROM), CSCE, Montreal, Canada.

Battikha, M. (2004). "Bridge Failure Root Cause Analysis." *Proc., 5th Structural Specialty Conference* (CD-ROM), CSCE, Montreal, Canada.

Battikha, M.G. (2003). "Quality Management Practice in Highway Construction." *Int. J. of Quality and Reliability Management*, 20(5), 532-550.

Battikha, M. (2002a). "Problem Patterns for Infrastructure Construction Quality Management." *Proc., 7th Int. Conference on Applications of Advanced Technology in Transportation*, ASCE, Reston, Va., 545-552.

Battikha, M.G. (2002b). "QUALICON: Computer-Based System for Construction Quality Management." *J. of Constr. Engrg. and Mgmt.*, ASCE, 128(2), 164-173.

Fib. (2004). "Fib bulletin 29: Precast concrete bridges." *International Federation for Structural Concrete*, Lausanne, Switzerland.

MoTH. (1992). "Quality Assurance Program Manual." *Province of British Columbia Ministry of Transportation and Highways*, British Columbia, Canada.

MoTO. (1996). "Ontario Structure Inspection Manual." *Ministry of Transportation, Ontario*, Ontario, Canada.

Realtime Image Processing Algorithms for the Detection of Road and Environmental Conditions

James G. Haran,[1] John Dillenburg[2] and Peter Nelson[3]

[1]Department of Computer Science, University of Illinois at Chicago, 851 S. Morgan (M/C 152), Room 1120 SEO, Chicago, IL 60607-7053; email:jharan1@uic.edu
[2] Ph.D., Department of Computer Science, University of Illinois at Chicago, 851 S. Morgan (M/C 152), Room 1120 SEO, Chicago, IL 60607-7053; email:dillenbu@uic.edu
[3] Ph.D., Department of Computer Science, University of Illinois at Chicago, 851 S. Morgan (M/C 152), Room 1120 SEO, Chicago, IL 60607-7053; email:nelson@cs.uic.edu

Abstract

This paper discusses image processing algorithms for the recognition of environmental and road conditions from real-time camera images. This research addresses various implementation techniques for and considerations of the implementation of algorithms to extrapolate various features from images taken by stationary traffic cameras. These algorithms have uses in the areas of traffic pattern analysis, emergency response, real-time traffic situation awareness, and homeland security. This paper covers the design and implementation of an automated camera heading detection system to determine the directional components of a camera's position using the current camera image, various computer vision techniques, and a series of classification training images. Image processing topics addressed as part of this research include edge detection, line detection, and two-dimensional filtering.

Introduction

Thousands of traffic cameras are situated along roads throughout the United States. These cameras serve many purposes such as speed enforcement, accident investigation, and traffic flow analysis. In addition, a wealth of information on environmental and road conditions such as weather, traffic flow, and visibility can be obtained from the visual cues in the images. The additional information contained within these images is often very difficult to extract or determine automatically. Therefore, human intervention is usually required to analyze and interpret the images and collect important data. Furthermore, the images obtained from these cameras often suffer from severe noise distortion due to weather conditions and camera characteristics.

Another difficulty in the analysis of these images is the distortion which occurs as a result of perspective projection of lines within the image to a horizon of vanishing points. This perspective distortion is further compounded by the fact that traffic cameras are generally situated at heights of less than 30 meters.

A further difficulty in the harnessing of visual information from traffic images is the lack of multisensor information to detect camera heading or pitch. When translating from camera and image geometries to real world geometries, this lack of directional heading information means that camera observers cannot tell the direction of the traffic flow.

This paper discusses the analysis of 2D images from stationary expressway traffic camera systems. The result is a camera heading feedback mechanism for the detection of the two dimensional rotational angle of a traffic camera based on a single training image and the

analysis of the perspective projection vanishing point of straight path roadways. The core of this paper discusses a method to determine the perspective vanishing point from 2D camera snapshots. The remaining portion discusses a method for determining the rotational angle based on the image vanishing point and a single training image.

This research focuses on the interpretation of low resolution snapshot images from a real-time traffic monitoring system and uses various computer vision and trigonometric techniques to determine the perspective vanishing point and camera rotation angle.

Related studies have analyzed aerial images for the purpose of unmanned aerial vehicle navigation and transportation system mapping (Koutaki and Uchimura 2005, Laptev et al. 2000, Gorte et al. 2005, Kim 2005) using large area, high resolution photographs. The results of these studies do not, however, have direct translation to studies involving low-altitude, low-resolution, stationary camera situations. Therefore, the development of general purpose camera visual feedback algorithms is needed to allow for collection of real world information from traffic camera.

Related Work

A considerable amount of research has been performed on the analysis of 2D images and video to determine various traffic characteristics (Koutaki and Uchimura 2005, Laptev et al. 2000, Zhang et al. 2001, Gorte et al. 2005) for the purpose of traffic simulation calibration and traffic network mapping. Gorte et al 2005 used helicopter image sequences to extract microscopic traffic flow information during traffic congestion. In this case, linear road segments were extracted using various edge and line detection techniques on images taken from a height of approximately 500 meters. Laptev et al 2000 also analyzed aerial images to extract road patterns for the purpose of mapping and navigation. Other research (Kim 2005) has focused on the realtime analysis of aerial images for the purpose of Unmanned Air Vehicle (UAV) navigation. This work instead entails the analysis of images taken instead from low heights by stationary cameras.

Likewise, there has been a great deal of research into the analysis of images to determine the perspective vanishing point (Rother 2002, Shufelt 1999, Cantoni et al. 2001, Tuytelaars et al. 1997). It is well known (Shapiro and Stockman 2001) that the use of common points from multiple images can be used through the process of rectification to determine the angular distance between these images. Using this knowledge, it is possible to determine the 2D rotational angle of a camera based on the location within two images of a common camera point and the projection line from the camera to this point in each image. Additional camera calibration information such as camera height allows for the extraction of 3D characteristics of image scenes.

Traffic images are further complicated by the dynamic nature of the scene vehicles and the somewhat uniform color of road sections. This makes it difficult to distinguish individual points for comparison between multiple images of a single scene. Therefore, a general purpose algorithm for determining reference points or lines within traffic images is needed to allow use of the well known analysis algorithms in computer vision research.

Image Analysis

The images used in this research (see Figure 1) were acquired using stationary traffic cameras mounted at various locations in the metropolitan Chicago area. These 2D images used JPG format and were represented using a two dimensional array such that each value in the array represented the RGB color triplet for an individual image pixel. This section details the techniques used to enhance the traffic images and extract the road direction and vanishing point for use in heading detection.

Grayscale Conversion

The first step in image analysis was the conversion of the RGB image into grayscale using standard techniques. An example grayscale image is shown in Figure 2.

Figure 1: Image obtained from traffic camera (copyright Gary-Chicago-Milwaukee Corridor).

Figure 2: Grayscale converted image.

Median Filtering

After grayscale conversion, the noise within the image was smoothed using a median filter (see Figure 3). A median filter modifies an image pixel such that the new pixel value is selected as the median of the neighboring values within a range. This filtering has the effect of removing random noise in images while still maintaining the overall integrity of image regions and boundaries. This type of filter is typically categorizes as a blurring filter in many image processing toolkits.

Thresholding

The next step in processing involves the conversion of the grayscale image into a black and white image through a process called thresholding. This process involves the selection of a threshold value in the grayscale range which acts as a cutoff for determining which pixels should be converted to white or black in the resulting image. The resulting black and white image is shown in Figure 4. This image shows a clear perspective flow of image lines from the lower right corner toward the upper left vanishing point.

Figure 3: Image after median filtering.

Figure 4: The thresholded black and white image.

Line Detection

Using the thresholded image, a Hough Transform (HT) is performed. The HT algorithm maps all points in the image space into sinusoids within an alternate polar coordinate

parameter space. The polar coordinates defined within this parameter space define the angle and magnitude or length of vector normals for probable lines in the image space (see Figures 5 and 6). Using these parameters, it is possible to map points in the HT parameter space to lines in the image space.

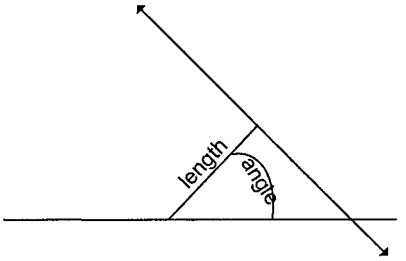


Figure 5: The polar representation of a line.

Figure 6: The Hough Transform parameter space.

Each point in parameter space is associated with a numerical value which relates to the likelihood of the line defined by the polar coordinates actually existing in the real scene from the image. The local maxima of this space define the most prominent lines within the image scene and provide the basis for use of HT parameter space to detect road paths.

Local maxima in the parameter space are reflected as bright spots in the resulting parameter space image (see Figure 6) and reflect the most well defined lines in the image.

Further analysis of this parameter space will reveal that the local maxima for the parallel lines within the image are themselves represented in a single parameter space line. This feature explains the appearance of a ridge in the parameter space image of Figure 6.

Perspective Line Detection

The revelation that collections of parallel image lines represent lines within the parameter space allows for the further reduction of our parameter space using another Hough Transform (see Figure 7). Within this secondary parameter space, a new set of sinusoids are plotted representing the likely ridges within the initial HT space. The maxima within this space reflect the ridges within the initial Hough space and the sets of parallel lines within the original image scene. This method is commonly known as a Cascaded Hough Transform or CHT (Tuytelaars et al. 1997). The appearance of multiple peaks reflects the existence of multiple sets of parallel lines within the initial image which is common in images. Figure 7, however, reflects only a single peak which directly relates to the parallel scene lines representing the edges and flow patterns of the road in Figure 1.

This single peak reflected in this secondary parameter space reflects a line in parameter space which passes through all image space perspective projection lines. Using the line represented by this peak, it is, therefore, possible to collect all parallel image lines into a single set for common analysis and trigonometric determination of the perspective projection vanishing point. One can imagine the ridge in the initial HT parameter space reflecting an arc in the image space which passes through all component lines of a single parallel line set.

Figure 7: The secondary parameter space mapping of Figure 6.

Vanishing Point Determination

Using the secondary parameter space mapping, it is possible to use an inverse Hough Transform or trigonometric techniques to determine the perspective projection lines. Selection of any two parameter space points allows us to form the equations of two parallel image space lines. The intersection of these lines is the perspective vanishing point. In our example, the perspective vanishing point does not lie within the image space, but is easily determined using a planar space expanding beyond the image boundaries on an alternate coordinate system.

Angle of Rotation

Once the vanishing points of the current image and our training image have been determined, it is possible to calculate the distance between the two points and hence the road line relative to the stationary traffic camera to determine the heading direction or angle or rotation of the camera in two dimensional space. The relative rotation of the roadway in this image relative to a training or reference image roadway allows for the determination of the angular rotation of the camera. This angular rotation is easily mapped to real world directional coordinates based on the calibration information for our camera and reference image. This approach assumes a fixed camera location and the knowledge of certain camera calibration settings.

Results

The research presented within this paper reflects the use of computer vision techniques to extract important information from 2D images of roads. This work also detailed the use of a secondary Hough Transform to detect parameter space ridges within Hough Transform parameter space. These secondary HT parameter space maxima allowed the collection of multiple parameter space lines into sets for common analysis.

Using these techniques, it is possible to map the various images from single traffic locations and stationary cameras into true directional heading components. In addition, the methods detailed allow for use in detecting ribbons of curved road sections using methods detailed in other research (Laptev et al. 2000, Kim 2005).

Conclusion

This paper discusses the implementation of a vanishing point based camera heading detection system capable of analyzing the two dimensional heading from traffic camera images. Although this implementation performed well for the problem of detecting two dimensional vanishing points and camera rotational heading in a 2D image plane, it must be further adapted to detect 3D characteristics and account for the image acquisition distortion

common to video monitoring systems. In addition, the Hough Transform implementation discussed suffers from high computational complexity which may hinder use in a realtime video situation. Adaptation to realtime camera video would likely benefit from the implementation of optical flow methods (Shapiro and Stockman 2001) across multiple image frames in addition to the methods described in this paper.

The perspective projection vanishing point determination technique discussed within this paper allows for simple vanishing point detection and is easily adapted to the detection of multiple vanishing points. Additionally, the secondary Hough Transform method allows for clear distinction of multiple sets of parallel lines without the difficulties common to simple edge detection techniques.

Future study on this topic will analyze the 3D camera position and consider the calibration information for the individual traffic camera position. In addition, this research will be adapted to analyze images from curved road positions. It is expected that the multiple instances of parallel line segments present in curved road sections will be easily detectable within the two HT parameter spaces.

References

Cantoni, V., Lombardi, L., Porta, M., Sicard, N., (2001). "Vanishing Point Detection: Representation Analysis and New Approaches." *Proc., 11th Conf. on Image Analysis and Processing.*

Gary-Chicago-Milwaukee Corridor Trans. Information. http://www.gcmtravel.com, (2005).

Gorte, B.G.H., Karimi Nejadasl, F., Hoogendoorn, S.P., (2005). "Outline Extraction of a Motorway from Helicopter Image Sequence." *CMRT05, ISPRS*, Vol. XXXVI, Part 3/W24.

Kim, Z., (2005). "Realtime Road Detection by Learning from One Example." *Proc., IEEE Workshop on Applications of Comp. Vision*, pp. 455-460.

Koutaki, G., Uchimura, K., (2005). "Automatic Road Detection Based on Cross Detection in Suburbs." *J. of Imaging Science and Technology*, vol.49. no.2, pp.163-169.

Laptev, I., Mayer, H., Lindeberg, T., Eckstein, W., Steger, C., Baumgartner, A., (2000). "Automatic Extraction of Roads from Aerial Images Based on Scale Space and Snakes." *Machine Vision and Applications*, Vol. 12, P. 23-31.

Rother, C., Carlsson, S., (2002). "Linear Multi View reconstruction and Camera Recovery using a Reference Plane." *Int'l J. of Comp. Vision*, 49(2/3):117-141.

Rother, C., (2002). "A New Approach to Vanishing Point Detection in Architectural Environments." *IVC*, 20(9-10), 647-656.

Shapiro, L.G., Stockman, G.C., (2001). *Comp. Vision*, Prentice Hall, New Jersey.

Shufelt, J.E., (1999). "Performance Evaluation and Analysis of vanishing Point Detection Techniques", *IEEE Trans. on Pattern Analysis and Machine Intel.*" Vol. 21, No. 3.

Tuytelaars, T., Proesmans, M., Van Gool, L., (1997). "The Cascaded Hough Transform as Support for Grouping and Finding Vanishing Points and Lines." *LNCS*, vol. 1315, pp. 278-289, Sommer G. and Koenderink J.J. eds., Springer-Verlag.

Zhang, C., Baltsavias, E., Gruen, A., (2001). "Knowledge Based Images Analysis for 3D Road Reconstruction." *Asian J. of Geoinformatics*, Vol. 1, No. 4, P. 3-14.

Optimal Spatial Sampling of Infrastructure Condition

Rabi G. Mishalani[1] and Liying Gong[2]

[1]Associate Professor, Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, 470 Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210; PH (614) 292-5949; FAX (614) 292-3780; e-mail: mishalani@osu.edu

[2]Graduate Research Associate, Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, 470 Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210; PH (614) 688-3761; FAX (614) 292-3780; e-mail: gong.24@osu.edu

Abstract

Infrastructure management is the process through which inspection, maintenance, and rehabilitation decisions are made with the aim of minimizing the total life-cycle cost. The inputs to decision-making include current facility condition as assessed from field observations, and future facility condition as forecasted based on the current condition and possible maintenance and rehabilitation (M&R) actions using a deterioration model. The quality of the estimated current condition depends on the accuracy of the inspection technology, the spatial sampling employed, and the nature of correlation between condition variables at different locations. The uncertainty associated with spatial sampling has not been recognized and quantified in the infrastructure management literature. However, its impact on arriving at the optimal M&R decisions can be appreciable. This paper motivates the importance of sampling, focuses on quantifying the spatial sampling uncertainty, and investigates the effect of incorporating spatial sampling as a decision variable on the expected minimum total life-cycle cost.

Introduction and Motivation

Infrastructure management is the process through which inspection, maintenance, and rehabilitation decisions are made with an aim to achieve the minimum expected total life-cycle cost. One primary input to this process is the current condition of the facility, which is used to forecast future condition. The mean of a sample of condition observations over a homogenous infrastructure section is an estimate of its current condition (Mishalani and Koutsopoulos, 1995, 2002). The quality of measurements, the sample size, and the nature of correlation between condition variables at different locations determine the accuracy of condition estimates, and more accurate estimates have the potential to lead to more effective M&R decisions. Consequently, on the one hand the expected combined user costs and M&R costs are reduced over the planning horizon. On the other hand, more accurate information requires more resources such as increased inspection frequency, advanced inspection technologies and data processing methods, and larger sample sizes or less positively correlated observations. Therefore, the optimum combination of inspection decisions and M&R decisions should be determined based on an economic evaluation that captures the long-term costs and benefits.

In this paper three sources of uncertainty are considered: measurement, spatial sampling, and forecasting. Measurement uncertainty is mainly due to technological limitations and environmental effects. The use of forecasting models leads to forecasting uncertainty. The uncertainty associated with spatial sampling is due to the spatial variation and the spatial correlation of condition variables within a homogenous section. Humplick (1992) developed a measurement error model whereby the difference between the true value of an indicator at a given location and its measured value is explained in terms of systematic biases and a random error. Madanat (1991, 1993), and Madanat and Ben-Akiva (1994) extended the traditional infrastructure management framework known as the Markov Decision Process (MDP) to the Latent Markov Decision Process (LMDP) to capture both measurement errors and forecasting uncertainty. Nevertheless, decisions regarding inspection spatial sampling in the context of infrastructure management have thus far not been considered and the associated uncertainty has not been recognized or quantified. Even though the cost associated with inspection activities might be small, its impact on arriving at optimal M&R decisions can be appreciable. Therefore, the primary focus of this paper is to capture inspection sampling uncertainty within the decision-making framework and to develop a methodology for making inspection spatial sampling decisions in the field of infrastructure management.

In the following section an overview of the LMDP framework is provided. The next section presents the spatial sampling formulation for M&R decision-making in the presence of spatial sampling. The resulting new framework is thus referred to as the Latent Markov Decision Process in the presence of Spatial Sampling (LMDPSS). In the fourth section an illustrative example is presented and a numerical comparison between the LMDP and LMDPSS frameworks is discussed for the purpose of assessing the value of capturing the uncertainty associated with spatial sampling and incorporating the sample size as a decision variable. The last section summarizes the paper and discusses possible future research.

Latent Markov Decision Process

LMDP is an extension of the traditional Markov Decision Process (MDP), but differs from it in several aspects including the recognition that condition measurements are not error-free. The LMDP framework assumes that the decision maker's observations are only probabilistically related to the true condition of the facility. This relationship can be mathematically represented as follows (Madanat, 1991, 1993; Madanat and Ben-Akiva, 1994):

$$q(\hat{x}_t^d = k \mid x_t^d = j, r_t) = f(\sigma_r^2); \quad 1 \le k, j \le w; \quad t = 0,1,\cdots,T-1 \tag{1}$$

where x_t^d = discrete true condition state of the facility at time t, $\hat{x}_t^d$ = measured state at time t, r_t = measurement technology index at time t selected at time $t-1$, $q(\cdot|\cdot)$ = conditional probability that the measured state $\hat{x}_t^d$ is k given that the true state x_t^d is j and that the measurement technology r_t is employed, $f(\cdot)$ = a known function, σ_r^2 = variance of the measured condition using inspection technology r after it has been corrected for the systematic biases captured by the measurement error model (Humplick, 1992), w = total number of condition states, and T = planning time horizon.

In the LMDP framework a new augmented state is defined to account for all the information available to the decision maker at a given time and relevant to future decisions (Madanat, 1991, 1993; Madanat and Ben-Akiva, 1994), which is denoted by I_t as follows:

$$I_t = \{I_0, a_0, r_1, \hat{x}_1, \cdots, \hat{x}_{t-1}, a_{t-1}, r_t, \hat{x}_t\}; \quad t = 1,2,\cdots,T \tag{2}$$

where I_0 = the initial information available at the start of the planning horizon, and a_t = M&R actions at time t.

With the introduction of the augmented state, the same solution method of dynamic programming that is used to find optimal M&R policies by minimizing expected costs in the MDP framework can be adopted in the LMDP one. At every time t, the decision maker has the available state of information I_t, on the basis of which the system transitions to one of the augmented states $I_{t+1} = K$ given a selected M&R action a_t and inspection action r_{t+1} with probability $P(I_{t+1} = K|I_t, a_t, r_{t+1})$. The same process is then repeated in time period $t+1$, and so on, thus reflecting the evolution of the state of information. The dynamic program is then solved recursively for every state of information I_t for each year t such that the optimal values of a_t and r_{t+1} are determined whereby the total expected cost over the planning horizon is minimized.

Methodology

A field is defined as a section of infrastructure where condition behaves homogeneously over space and time (Mishalani and Koutsopoulos, 1995, 2002). In this study, a field represents a facility of length L for which decisions are made. The following condition related notation is adopted: x = condition variable represented on a continuous scale, μ_x = true mean of the condition variable for the field in question, σ_x^2 = true variance of the condition variable, and $\rho(s)$ = correlation between condition variables at two points within the field that are separated from one another by a distance s.

As was discussed in the introduction, the facility condition μ_x is estimated by the sample mean, and represents one primary input to M&R decision-making. This sample mean is estimated by the following:

$$\bar{\hat{x}}_{n,r} = \frac{1}{n}\sum_{l=1}^{n} \hat{x}_{l,r} \tag{3}$$

where $\bar{\hat{x}}_{n,r}$ = sample mean across condition observations within the field, $\hat{x}_{l,r}$ = measured condition at location l ($l = 1,2,\cdots,n$) using measurement technology r, and n = sample size.

The uncertainty associated with the sample mean is captured by its variance, which is denoted by $\mathrm{Var}(\bar{\hat{x}}_{n,r})$. The following assumptions are made to arrive at a specific solution to the variance of the sample mean:

(i) The true condition over space $\{x_l, l = 1,\cdots,n\}$ is a stationary process,
(ii) The condition variable and random measurement errors are independent,
(iii) The measurement random errors at different locations are also independent, and
(iv) Distance between locations l and l+1 is constant and equal to h for all l.

Based on the above assumptions, the variance of the sample mean takes the following form:

$$\mathrm{Var}(\bar{\hat{x}}_{n,r}) = \frac{\sigma_r^2}{n} + \frac{\sigma_x^2}{n} + \frac{2\sigma_x^2\sum_{i=1}^{n-1}(n-i)\cdot\rho(i\cdot h)}{n^2} \tag{4}$$

Equation (4) combines both the sampling and measurement uncertainties through the variance of the condition variable σ_x^2, the sample size n, the variance of the corrected

condition measurement, σ_r^2, and the spatial correlation among observations, $\rho(s)$. Clearly, the larger the sample size is, the smaller the contributions of the variances of the condition variable and its measurement are to the overall variance; and the smaller the correlation among observations is, the more accurate the estimate of facility condition is. Since the estimate of mean condition within a field is the basis for M&R decision-making, the nature of the sample mean statistic is further developed by introducing distributional assumptions. The condition variable x is assumed to follow a normal distribution with mean of μ_x and a variance of σ_x^2. The corrected condition measurement is assumed to follow a normal distribution with mean zero and a variance of σ_r^2. It follows that the distribution of the sample mean, $\bar{\hat{x}}_{n,r}$, is normal with mean of μ_x and a variance as given by equation (4) (Casella and Berge, 2002).

As a result, the measurement probability represented by equation (1) is revised to reflect the effects of sampling and is now referred to as the measurement and sampling probability. It is mathematically represented as follows:

$$q(\bar{\hat{x}}_{n,r,t}^d = k \mid \mu_{x,t}^d = j, n_t, r_t) = f(\sigma_x^2, \sigma_r^2, \rho, n_t) \tag{5}$$

where $\bar{\hat{x}}_{n,r,t}^d$ = discrete condition state based on the continuous value of the sample mean across condition observations within the field at time t, and $\mu_{x,t}^d$ = discrete true condition state based on the true mean of the condition variable for the given field at time t.

Comparing equation (5) with equation (1), it is clear that the measurement and sampling probability $q(\cdot|\cdot)$ is not only a function of the factor related to the inspection technology, σ_r^2, as in equation (1), but also a function of factors corresponding to the properties of the observed facility, σ_x^2 and $\rho(s)$, and the sample size n_t where the index t represents the time the sample is taken.

In order to incorporate the spatial sampling in the LMDP framework, the three main revisions to the LMDP framework required are the following:

(i) The measured condition state is replaced by the condition sample mean,
(ii) The uncertainty associated with the sample mean is captured through the measurement and sampling probability, and
(iii) The sample size becomes a decision variable.

Therefore, the decision variables are M&R activity a_t, inspection technology r_{t+1}, and sample size n_{t+1} for each time period t. Then, the same solution method of dynamic programming that is used in the LMDP framework can be adopted in the LMDPSS one to find the minimum total expected cost over the planning horizon and the optimum M&R, inspection, and sampling activities for each possible state of information I_t at time t.

Numerical Comparison of LMDPSS to LMDP

In this analysis an example 1 km long homogeneous pavement section consisting of the 3.66 m wide lane of an interstate highway is assumed. This section is divided into 100 subsections with equal area. Each 10 m long subsection is considered a potential observation unit. The planning horizon T is 10 years. There are 11 combinations of inspection activities: two types of inspection technologies and five levels of sample sizes (20%, 40%, 60%, 80% and 100%), in addition to the "no inspection" option. The condition scale is described by eight states where state 7 represents the best and state 0 the worst. The initial condition state

is set at 7 with a probability of one. The observations are assumed to be independent over space in this example.

In order to assess the value of capturing spatial sampling uncertainty and incorporating the sample size as a decision variable in the LMDPSS framework with respect to the LMDP framework, several total life-cycle expected costs are defined. First, the true LMDP total cost is defined as following. An agency will perform the maintenance and inspection activities according to optimum policies obtained through the application of the LMDP framework. However, the resulting current and future conditions of the facility are simulated through the application of the LMDPSS framework. Thus, the calculated total expected cost based on the LMDPSS framework represents the true total cost associated with the optimum LMDP policies. This cost is denoted by C_1. Second, the total life-cycle cost resulting from the application of the LMDPSS framework with the sample size set at a pre-determined value is also determined and is defined by C_2. Third, the minimum total expected cost is determined by applying the LMDPSS framework when the sample size is selected optimally. This cost is denoted by C_3. The difference between C_1 and C_2 represents the value of capturing the uncertainty in the decision-making process associated with spatial sampling. The difference between C_2 and C_3 represents the effect of forcing the sample size to take a pre-determined value while still recognizing the uncertainty associated with sampling. This difference reflects the value of incorporating the spatial sample size as a decision variable.

The values of C_1, C_2, and C_3 for the example pavement section described above are shown in Figure 1. It is clear that capturing the spatial sampling uncertainty and introducing the sample size as a decision variable are equally important in reducing the total expected cost. Moreover, based on the relative percentage savings with respect to the minimum cost given by C_3, the application of the LMDPSS framework results in a total life-cycle cost saving ranging from 4.24% to 4.78% for this particular example. Since the order of magnitude of the costs is large, such percentage savings still represent marked improvements. For example, in maintaining the 209,107 lane-miles of interstate highways in the United States over a 10-year planning horizon, the savings would be in the order of $9.8 billion based on the inputs assumed in this example.

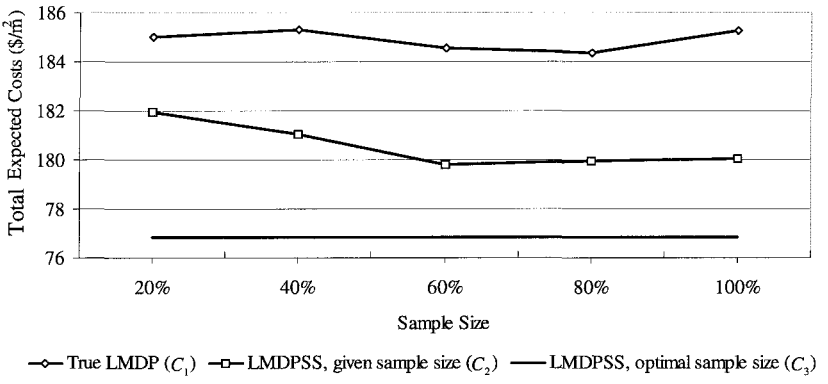


Figure 1. Numerical compositions.

Summary and Future Research

The uncertainty associated with spatial sampling has not been recognized or quantified in the literature of infrastructure M&R life-cycle cost-based decision-making. Moreover, the sample size is not incorporated as a decision variable. This paper presents a methodology where the uncertainty in question is captured and the sample size is incorporated as a decision variable in an optimization framework. The results for the example considered indicate that while the percentage savings of total expected cost from such treatments are not large, the order of magnitude of the savings is.

An extension of this numerical analysis whereby positive spatial correlation among condition observations is assumed is reserved for future study. Another fruitful area worthy of further research is to investigate the impact of certain factors on the M&R decision outcomes and the life-cycle cost. Factors including inspection cost, user cost, and the nature of the spatial correlation can be considered. The results would indicate which factors are critical in influencing the decision outcomes. In addition, considering spatial sampling in making decisions at the network level is another important problem to address. Network-level decisions focus on selecting facilities for maintenance or rehabilitation under network constraints, such as budget limitations. Simlowitz and Madanat (2000) utilized the LMDP framework to address the network level problem. However, spatial sampling is not considered. It is worthwhile to do so at the network level given that inspection resources need to be allocated across the network optimally as well.

Acknowledgements

This study was funded through a grant from the US National Science Foundation.

References

Casella G., and Berge R. (2002). *Statistical Inference*, Duxbury, Pacific Grove, CA.

Humplick, F. (1992). " Highway Pavement Distress Evaluation: Modeling Measurement Error." *Transportation Research*, Vol. 26B, pp. 135-154.

Madanat S. M. (1991). "Optimizing Sequential Decisions under Measurement and Forecasting Uncertainty: Application to Infrastructure Inspection, Maintenance and Rehabilitation." Ph.D. Dissertation, Department of Civil Engineering, MIT.

Madanat S. M. (1993). "Optimal Infrastructure Management Decisions under Uncertainty." *Transportation Res. –C*, Vol. 1, No.1, pp. 77-88.

Madanat S. M. and Ben-Akiva M. (1994). "Optimal Inspection and Repair Policies for Infrastructure Facilities." *Transportation Science*, Vol. 28, No. 1, pp. 55-62.

Mishalani, R.G., and Koutsopoulos H. N. (1995). "Uniform Infrastructure Fields: Definition and Identification." *Journal of Infrastructure Systems*, ASCE, Vol. 1, No. 1, pp.44-55.

Mishalani R. G., and Koutsopoulos H. N. (2002). "Modeling the Spatial Behavior of Infrastructure Condition." *Transportation Research Part B*, Vol. 36, pp. 171-194.

Smilowitz K., and Madanat S. (2000). "Optimal Inspection and Maintenance Policies for Infrastructure Networks." *Computer-Aided Civil and Infrastructure Engineering*, Vol. 15, No. 1, pp. 5-13.

Application of Shackle's Model for Highway Project Evaluation under Uncertainty

Zongzhi Li[1] and Kumares C. Sinha[2]

Abstract: This paper introduces application of Shackle's model for highway project evaluation under uncertainty. Shackle's model overcomes limitations of existing risk-based approaches for project evaluation by: i) using degree of surprise as a measure of uncertainty instead of probability distribution; ii) introducing a priority index to jointly evaluate each outcome and degree of surprise pair; and iii) establishing a focus gain-over-loss ratio as the basis of project comparison under uncertainty. Survey data were utilized to calibrate systemwide focus ratio functions to help establish focus ratios as project benefits under uncertainty. The calibrated focus ratio functions were applied to data on past candidate projects for Indiana state highway programming in 1998-2001 to establish focus ratios for individual projects. The computed focus ratios as project benefits were further utilized in an existing optimization model for project selection. It was found that consistency matching rate between lists of projects selected using proposed methodology and list of projects actually implemented was at a minimum of 84 percent. The model could be used by transportation agencies for enhanced highway project evaluation under uncertainty in long-range planning and programming.

CE Database subject headings: Shackle's Model; Highway Project; Project Evaluation; Evaluation under Uncertainty.

[1]Asst. Prof, Dept. of Civ. and Arch. Engrg, Ill. Institute of Tech., 3201 S. Dearborn St., Chicago, IL 60616. Phone: (312)567-3556, Fax: (312)567-3519, Email: lizz@iit.edu.

[2]Prof, School of Civ. Engrg, Purdue Univ., 550 Stadium Mall Dr., West Lafayette, IN 47907-2051. Email: Phone: (765)494-2211, Fax: (765)496-7996, sinha@ecn.purdue.edu.

INTRODUCTION

In most cases, highway project evaluation is conducted with the assumption that investment outcomes occur under certainty. Uncertainty characteristics of project benefits such as decrease of life-cycle agency costs and reduction in travel time are not considered. In recent years, researchers began to incorporate risk-based analysis. This was done by considering a probability distribution to possible investment outcomes. The expected benefit based on a weighed average was then obtained and used as the basis of project evaluation. The underlying assumption of risk-based analysis was that the weight associated with a particular outcome was independent of degree of uncertainty of the outcome. Notable examples included expected utility approach for project benefit analysis by Ford and Ghose (1998) and probabilistic life-cycle cost analysis in pavement design proposed by the Federal Highway Administration (FHWA) (1998). However, it might not be possible to assign a probability distribution to possible outcomes or even the full range of outcomes is unknown. That is, the decision-maker is faced by uncertainty cases where an expected value cannot be established. This paper introduces a new methodology that uses Shackle's model (Shackle, 1949; Young, 2001) for project evaluation in highway asset management practice.

BACKGROUND INFORMATION

Highway Asset Management System Goals

A system goal is a general statement of a desired state or ideal function of a highway system. Highway asset management focuses on system goals from both agency and user perspectives. Transportation agencies aim to preserve physical asset conditions at or above a desired level to protect public investments and deliver highest level of service with minimum investment. On the other hand, highway users wish to appreciate lowest cost while using the highway system, maintain basic mobility and safety to employment and educational opportunities, and minimize energy use and adverse impacts to the environment. System goals considered in current study include system preservation, agency cost, user cost, mobility, safety, and the environment.

Highway Asset Management Programs

Highway asset management programs are classified for preserving and expanding physical highway assets and sustaining levels of service of a highway system. General programs considered in current study include bridge and pavement preservation, safety and roadside improvements, system capacity expansion, forest highways and state facilities, Intelligent Transportation Systems (ITS) installations, multi-modal facilities, and maintenance activities.

Performance Measures

Performance is defined as execution of required function and performance measures are quantitatively or qualitatively indicators that directly or indirectly assess degree to which outcomes of investments meet system goals (Poister, 1997). In highway investment decision-making process, typical sources of uncertainty in project evaluation involve those related to improvement of physical asset conditions, reduction in life-cycle agency and user costs, enhancement of mobility and safety, and mitigation of vehicle emissions. These benefits are computed by changes in the values of *different performance measures.*

PROPOSED METHODOLOGY

Shackle's Model for Project Evaluation under Uncertainty

For the case under uncertainty with unknown probability distribution or unknown range of possible investment outcomes, the expected value of benefits cannot be established. Shackle's model overcomes this limitation by: i) using degree of surprise as a measure of uncertainty instead of probability distribution; ii) introducing a priority index to jointly evaluate each outcome and degree of surprise pair; and iii) establishing a standardized focus gain-over-loss ratio for a given range of deviations from the expected outcome. The focus ratio is then used as a means of project comparison. Higher preference is given to a project with a higher focus ratio. This model requires calibrating, degree of surprise function, priority function, and focus ratio function for each performance measure.

Degree of Surprise Function

Degree of surprise indicates the decision-maker's reaction to degree of uncertainty regarding different outcomes of a performance measure from any investment alternative, with gains and losses from the expected outcome being considered separately (Figure 1). Degree of surprise function for a performance measure can be established in the following:

- Assume a range of possible outcomes x_i from an investment alternative ($i = 1, 2, ..., n$).
- Assign a value as degree of surprise y, ranging from 0 for no surprise to 10 for highest surprise, to reflect the decision maker's degree of belief for a given outcome.
- Establish degree of surprise function $y = f(x)$.

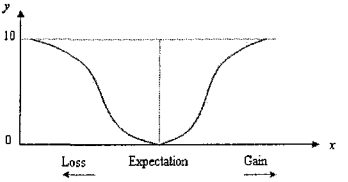


FIG. 1. Diagram of a Degree of Surprise Function

Priority Function and Focus Gain and Focus Loss Values

Priority function indicates the weight that any outcome and degree of surprise pair (x, y) is given, or in Shackle's terminology, the power of any pair to attract the attention of the decision maker (Figure 2). Priority function for a performance measure can be developed as follows:

- Determine a priority index ϕ by jointly considering each outcome of a performance measure and degree of surprise pair, using an index of 0 for lowest priority and 2, 3, 4, ... for higher priorities.
- Denote the decision-maker's priority function by $\phi = \phi(x, y)$ and the function possesses following properties: $\frac{\partial \phi}{\partial x} \rangle 0$; $\frac{\partial \phi}{\partial y} \langle 0$. Priority function for a performance measure can be defined in the following function forms $\phi = \alpha_1 x^{0.5} - \alpha_2 y^2$, $\phi = \alpha_1 x - \alpha_2 y^2$, or $\phi = \alpha_1 x^{0.5} - \alpha_2 y$, where α_1 and α_2 are coefficients with respect to outcome x and degree of surprise y.
- Priority function ϕ is a saddle shaped curve that maintains a maximum priority index on the gain from expected outcome side and a maximum priority index on the loss from expected outcome side. The two outcomes corresponding to the maximum priority indices are called focus gain (x_{FG}) and focus loss (x_{FL}).

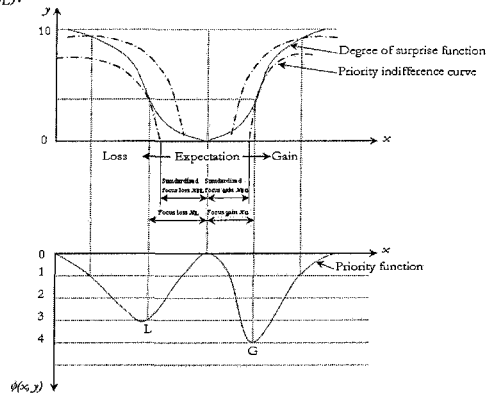


FIG. 2. Diagram of a Priority Function

Standardized Focus Gain-over-Loss Ratio and Focus Ratio Function

As seen from Figure 2, focus gain x_{FG} and focus loss x_{FL} are associated with certain degrees of surprise, i.e., on the gain side $\phi(x_{FG}, y(x_{FG}))$ is maximized and $y(x_{FG}) \neq 0$; and on the loss side $\phi(x_{FL}, y(x_{FL}))$ is maximized and $y(x_{FL}) \neq 0$. It is therefore desirable to use priority indifferent curves with constant $\phi(x, y)$ values to derive respective standardized focus gain x_{SG} and standardized focus loss x_{SL} which are associated with no degrees of surprise. That is, to find x_{SG} and x_{SL} such that $\phi(x_{SG}, 0) = \phi(x_{FG}, y(x_{FG}))$ and $\phi(x_{SL}, 0) = \phi(x_{FL}, y(x_{FL}))$. Ranking of any project against that of another is achieved by ratio of standardized focus gain and loss pair $R = x_{SG} / x_{SL}$. The larger the ratio R is, the higher likelihood to select a project.

Standardized focus gain x_{SL}, standardized focus loss x_{SL}, and their ratio R vary for different ranges of possible outcomes. The functional relationship between the ratios R and different ranges of possible outcomes can then be established. Using the standardized focus gain-over-loss ratio function (focus ratio function hereafter), a focus ratio R can be determined provided with information on range of deviations from the expected outcome.

Development of a Systemwide Focus Ratio Function for Each Asset Management Program

For each asset management program, multiple performance measures are involved under each or across various system goals. This requires synthesizing focus ratio functions for individual performance measures into a systemwide focus ratio function for each asset management program. The synthesis process is highlighted as follows:

- Relative weights of asset management system goals
- Relative weights of multiple performance measures under each system goal
- Focus ratio functions for individual performance measures
- Focus ratio function under each system goal for each asset management program
- Systemwide focus ratio function for each asset management program.

Determining relative weights of system goals will help synthesis of an asset management program's goal-specific focus ratio functions into an overall systemwide focus ratio function. Where involving multiple performance measures under a specific system goal, relative weights of performance measures need to be established so that focus ratio functions for these performance measures can be combined into a goal-specific focus ratio function using additive rule.

$$R(X_i) = \sum_{k=1}^{K_i} w_{ik} \cdot R(x_{ik}) \tag{1}$$

where

$R(X_i)$ = Goal-specific focus ratio function
w_{ik} = Relative weight of performance measure k under system goal i
$R(x_{ik})$ = Focus ratio function for performance measure k under system goal i
K_i = Number of performance measures under system goal i.

Strong correlations exist between system goals such as system preservation, user cost, and environmental impacts. As such, additive rule cannot be applied when synthesizing the six goal-specific focus ratio functions into a systemwide focus ratio function for each asset management program. In an alternative formulation, multiplicative rule needs to be followed to establish a systemwide focus ratio function for each asset management program.

$$\left[c \cdot R(X_1, X_2, X_3, X_4, X_5, X_6) + 1\right] = \prod_{i=1}^{6} \left[c \cdot W_i \cdot R(X_i) + 1\right] \tag{2}$$

where

$R(X_1, X_2, X_3, X_4, X_5, X_6)$ = Systemwide focus ratio function for an asset management program

W_i = Relative weight of system goal
$R(X_i)$ = Goal-specific focus ratio function
c = Scaling constant.

The systemwide focus ratio function for a specific asset management program will help compute systemwide focus ratios for individual candidate projects under the asset management program as their benefits. Each project will result in changes in the values of individual performance measures. Given the expected outcome of a performance measure impacted by the project and range of deviations from the expected outcome, a focus ratio corresponding to the performance measure can be determined. Then, goal-specific focus ratios associated with the project can be computed by Equation (1) and a systemwide focus ratio for each project can be derived using Equation (2).

APPLICATION OF PROPOSED METHODOLOGY

Determination of Relative Weights of System Goals and of Performance Measures

Questionnaire surveys were conducted to derive relative weights of system goals and relative weights of multiple performance measures under a specific system goal. Each questionnaire was distributed to randomly selected employees of the Indiana Department of Transportation (DOT) and to highway users in Indiana, respectively. These survey participants represented the agency and user groups. Delphi technique (Dalkey and Helmer, 1963) with dual swaps was used to achieve consistent data handling. Twenty-nine people from agency group and twenty-eight people from user group participated in the surveys on relative weights of system goals. In addition, eighteen people from each group were involved with surveys on relative weights of performance measures. Raw data were further processed using Analytical Hierarchy Process (AHP) (Saaty, 1977) to obtain most compromised weights. The relative weights of system goals and performance measures are shown in Tables 1 and 2.

TABLE 1. Relative Weights of System Goals Derived Using AHP Technique

System Goal	Weights Assigned by Agency Group		Weights Assigned by User Group		Optimal Weights Derived by AHP Technique	
	Mean	STDEV	Mean	STDEV	Agency Group	User Group
System preservation	0.2259	0.0226	0.1857	0.0529	0.2267	0.1850
Agency cost	0.1922	0.0276	0.1625	0.0502	0.1920	0.1602
Highway user cost	0.1776	0.0336	0.1795	0.0605	0.1768	0.1778
Mobility	0.2112	0.0184	0.1955	0.0574	0.2119	0.1950
Safety	0.2319	0.0199	0.2295	0.0273	0.2322	0.2327
Environment	0.1716	0.0370	0.1911	0.0506	0.1708	0.1931

TABLE 2. Relative Weights of Performance Measures Derived Using AHP Technique

System Goal	Asset Management Program	Performance Measure	Agency Group	User Group
System preservation	Pavement Preservation	Road condition	0.5000	0.5000
		Remaining service life	0.5000	0.5000
Agency cost	Bridge Preservation	Construction cost	0.3542	0.3821
		Rehabilitation cost	0.3596	0.2982
		Maintenance cost	0.2863	0.3197
	Pavement Preservation	Construction cost	0.3528	0.3584
		Rehabilitation cost	0.3569	0.3140
		Maintenance cost	0.2903	0.3276
Highway user cost Mobility Environment	Bridge Preservation Pavement Preservation	Detour length Travel speed	0.4962 0.5038	0.5182 0.4818
Safety	Pavement Preservation	Travel speed	0.5072	0.5109
		Skid resistance	0.4928	0.4891
	Safety Improvements	Travel speed	0.4590	0.4596
		Collision rate	0.5410	0.5404
	Forest highways and state facilities under Department of Natural Resources	Travel speed	0.5071	0.5109
		Skid resistance	0.4929	0.4891

Calibration of Focus Ratio Function for Each Performance Measure

For the purpose of implementing Shackle's model, two sets of survey questionnaires were instrumented to collect information on degrees of surprise and priority indices for performance measures considered for uncertainty-based analysis. For each performance measure, ten different ranges of deviations from the expected outcome were used and, for a given range of deviations, a number of possible outcomes after project implementation were considered. Each questionnaire was separately distributed to randomly selected employees of the Indiana DOT and to highway users in Indiana, representing agency and user groups. Thirteen people from each group participated in the survey. Using the survey data, econometric methods were used to calibrate degree of surprise functions, priority functions, and focus ratio functions for individual performance measures. The calibrated focus ratio functions are presented as in Table 3.

TABLE 3. Calibrated Focus Ratio Functions for Individual Performance Measures

System Goal	Performance Measure	$R(x) = \alpha_0 + \alpha_1$ Range of Deviations x				Adjusted R^2
		α_0	(t-statistic)	α_1	(t-statistic)	
System preservation	Pavement condition	1.1785	(46.55)	-0.7683	(-9.42)	0.91
	Remaining service life	1.0380	(72.61)	-0.0759	(-4.12)	0.64
Agency cost	Construction cost	1.0382	(106.66)	-0.1847	(-5.89)	0.79
	Rehabilitation cost	1.0398	(106.12)	-0.2647	(-8.38)	0.89
	Maintenance cost	1.0132	(216.41)	-0.1538	(-0.19)	0.92
Highway user cost	Average travel speed	1.1535	(58.04)	-0.2331	(-9.10)	0.90
Mobility	Detour length	1.0251	(64.62)	-0.2070	(-7.68)	0.87
Environment	Intersection delay time	1.1414	(38.17)	-0.1572	(-3.45)	0.55
Safety	Average travel speed	1.1248	(56.36)	-0.1330	(-8.05)	0.88
	Bridge load inventory	1.0924	(69.61)	-0.1791	(-7.08)	0.85
	Pavement skid resistance	1.0834	(76.77)	-0.1887	(-8.30)	0.88
	Vehicle collision rate	1.0750	(88.34)	-0.1526	(-7.78)	0.87

Development of Systemwide Focus Ratio Function for Each Asset Management Program

Using relative weights of the six system goals derived and assuming equal weights between the agency and user groups, scaling constant c in Equation (2) was determined from $[c+1] = \prod_{i=1}^{6} [c \cdot W_i + 1]$, which is -0.33522. Hence, systemwide focus ratio function for each asset management program is in the following specification

$$R(X_1, X_2, X_3, X_4, X_5, X_6) = 2.98312 \cdot \left\{1 - \prod_{i=1}^{6} [1 - 0.33522 \cdot W_i \cdot R(X_i)]\right\}. \quad (3)$$

CASE STUDY

Data Source and Model Applications

The calibrated focus ratio functions based on Shackle's model were applied to data on past candidate projects for Indiana state highway programming in 1998-2001 to establish focus ratios. The computed focus ratios as project benefits were further utilized in an existing optimization model for project selection (Li and Sinha, 2004). A subset from all candidate contracts was selected using the optimization model that maximized total focus ratio values of selected contracts under budget and other constraints. Two project selection lists were derived from the optimization model according to two scenarios of budget constraints: with year-by-year constraints (budget scenario 1) and with a cumulative budget for all years combined (budget scenario 2). Since candidate projects were proposed in the past, the list of projects selected by Indiana DOT for actual implementation was available. This allowed us to make cross comparisons between lists of candidate projects selected using the proposed methodology and list of projects actually authorized. The robustness and consistency of the proposed methodology were evaluated.

Discussion of Case Study Results

Cross comparison results are listed in Table 4. The matching percentage in total number of contracts both selected using focus ratios as project benefits derived from the proposed methodology and authorized by the Indiana DOT was quite reasonable. At minimum, the consistency matching rates are 84 percent and 91 percent for budget scenario 1 and budget scenario 2, respectively.

TABLE 4. Number of Candidate Contracts Proposed and Both Selected Using the Proposed Methodology and Authorized by the Indiana DOT

Fiscal Year	Candidate Contracts	Indiana DOT Authorized	Case Study					
			Scenario 1 and Indiana DOT	% Match	Scenario 2 and Indiana DOT	% Match	Scenarios 1, 2 and Indiana DOT	% Match
1998	429	152	142	93%	142	93%	137	90%
1999	412	324	293	90%	306	94%	281	87%
2000	611	584	491	84%	532	91%	488	84%
2001	418	412	411	100%	394	96%	394	96%
Total	1,870	1,472	1,337		1,374		1,300	

Note: One match is counted if a contract is selected by the proposed method and authorized by Indiana DOT.

CONCLUSIONS

This paper presents a new methodology based on Shackle's model for highway project evaluation under uncertainty. Shackle's model was applied to calibrate systemwide focus ratio functions using data collected from a series of questionnaire surveys. The calibrated focus ratio functions were applied to data on past candidate projects for Indiana state highway programming in 1998-2001 to establish focus ratios as project benefits. The computed project benefits were further utilized in an existing optimization model for project selection. Results of project selection generated from the proposed methodology were compared with those actually implemented and high matching percentages were found. The proposed methodology could be adopted by transportation agencies for enhanced highway project evaluation under uncertainty in long-range planning and programming.

REFERENCES

Dalkey, N.C., and Helmer, O. (1963). An Experimental Application of the Delphi Method to the Use of Experts. *Management Science*, 9(3), 458-467.

Ford, J.L., and Ghose, S. (1998). Lottery Designs to Discriminate between Shackle's Theory, Expected Utility Theory and Non-Expected Utility Theories. *Annals of Operations Research*. Kluwer Academic Publishers, Norwell, MA.

FHWA. (1998). *Life-Cycle Cost Analysis in Pavement Design-In Search of Better Investment Decisions*. Federal highway Administration, U.S. Department of transportation, Washington, D.C.

Li, Z., and Sinha, K.C. (2004). A methodology for Multicriteria Decision Making in Highway Asset Management. *Transportation Research Record*, 1885, 79-87.

Poister, T.H. (1997). Performance Measurement in State Department of Transportation. *NCHRP Synthesis of Highway Practice 238*. National Academy Press, Washington, D.C.

Saaty, T.L. (1977). A Scaling Method for Priorities in Hierarchical Structures. *Journal of Mathematical Psychology*, 15, 234-281.

Shackle, G.L.S. (1949). *Expectation in Economics*, 2nd Edition. Cambridge University Press, Cambridge, United Kingdom.

Young, R.A. (2001). Uncertainty and the Environment: Implications for Decision Making and Environmental Policy. Edward Elgar, Cheltenham, United Kingdom.

A STOCHASTIC OPTIMIZATION MODEL FOR HIGHWAY PROJECT SELECTION AND PROGRAMMING UNDER BUDGET UNCERTAINTY

Zongzhi Li[1] and Murat Puyan[2]

Abstract: Federal, state, and local transportation agencies are faced by issues of steady increase in travel demand, deterioration of physical asset conditions, public demand for government accountability, and pressure of constrained budget. Responding to these challenges, transportation agencies have started to advance asset management concepts that cover all physical assets and usage of a highway transportation system, identify needs of the entire system, and determine investments to maintain and improve highway performance most cost-effectively. The existing models for project selection in investment decision process treat the optimization problem as a deterministic problem. The random nature of budget faced by the decision-maker is not considered, which limits achieving robust results. This paper introduces a stochastic optimization model for project selection that considers budget uncertainty. The model was formulated as the stochastic multi-choice multidimensional Knapsack problem with Ω-stage budget recourses. Multi-choice corresponds to multiple budget levels for different asset management programs, while multi-dimension refers to multiple years of analysis. The objective was to select a subset of candidate projects to achieve maximized system benefits under budget and other constraints. An efficient solution algorithm was developed using Lagrangian relaxation techniques. Data on candidate projects for Indiana state highway programming in 1998-2001 were used to apply the proposed stochastic model and solution algorithm. To assess the impact of budget uncertainty on project selection, the stochastic model was also applied to the same data set without considering budget recourses, namely, the model was treated as a single-stage deterministic model by only executing the first stage optimization using expected budget. The two sets of results generated were compared and it was revealed that the stochastic model and solution algorithm could assist transportation agencies for enhanced highway investment decisions for optimal system performance.

CE Database subject headings: Stochastic Model; Stochastic Optimization; Highway Investment Decision-Making; Budget Uncertainty.

[1]Asst. Prof, Dept. of Civ. and Arch. Engrg, Ill. Institute of Tech., 3201 S. Dearborn St., Chicago, IL 60616. Phone: (312)567-3556, Fax: (312)567-3519, Email: lizz@iit.edu.

[2] Ph.D., Postdoc., Dept. of Civ. and Arch. Engrg, Ill. Institute of Tech., 3201 S. Dearborn St., Chicago, IL 60616. Phone: (312)567-3896, Fax: (312)567-3519, Email: puyan@iit.edu.

INTRODUCTION

Highway System Goals, Asset Management Programs and Candidate Projects

Highway investment decision-making is a complicated undertaking due to its many intricate dependencies on a number of factors. Within a transportation agency, various system goals are established aiming at preserving physical asset conditions, maintaining highest mobility and safety, and minimizing energy use and adverse impacts to the environment with minimum investments. Different asset management programs targeting pavements, bridges, safety improvements, roadside improvements, Intelligent Transportation Systems (ITS), multimodal facilities, maintenance, and so forth are normally developed to fulfill system goals. A certain amount of budget is designated to each program per year and candidate projects are proposed to compete for funding within the same program category and across different programs.

Characteristics of Project Selection

In general, projects that need funds from different programs in multiple years are bundled together as contract packages for implementation. For instance, a highway contract may contain pavement, bridge, and ITS projects that use funds from different programs in multiple years. In the decision-making process, selection of any one such contract necessitates the selection of all constituent projects and vice versa. The decision-making described above can be treated as a multi-choice multidimensional Knapsack problem (MCMDKP) (Martello and Toth, 1990). Multi-choice characteristic is imposed by the separate budget limitations of individual programs, while multidimensional attribute is derived owning to multiple years of analysis. The objective aims to maximize benefits achieved from the selection of improvement contracts, while respecting budget and other constraints.

Issues of Risk and Uncertainty

With perfect information regarding project costs and benefits as well as available budget, project selection can be accomplished using the deterministic optimization model. However, the actual system is more complex owing to uncertainty of each factor in that full range of its possible values is not defined or the probability distribution of the values within the range is unknown. The deterministic model is unequipped to model these conditions or reach an optimal solution. The uncertainty of these factors can only be modeled when it is transformed into a measure of risk with determinable ranges of values and continuous or discrete probability distributions. Once the ranges and probabilities are obtained, associated uncertainty becomes tractable and can be represented in the stochastic model. This study was undertaken with the aim to incorporate budget uncertainty for highway project selection.

A STOCHASTIC MODEL WITH Ω-STAGE BUDGET RECOURSES

Stochasticity of Budget

Consider a stochastic formulation with Ω-decision stages due to budget uncertainty. Without loss of generality, assuming that a discrete probability distribution of budget possibilities is available, as illustrated in Figure 1.

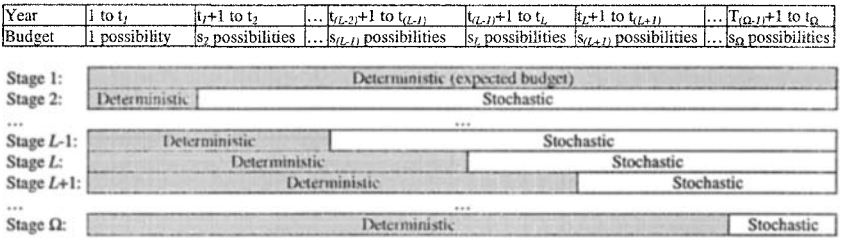


FIG. 1. Budget Attributes in a Ω-Stage Decision-Making Process

For the first stage decisions the budget matrix is comprised of the expected budget for all years. For the second stage decisions, budget for years 0 to t_1 is deterministic, and there are $(p_2=s_2.s_3.\ldots.s_{(L-1)}.s_L.s_{(L+1)}.\ldots.s_\Omega)$ possible combinations for years t_1+1 to t_Ω. For the generic stage L decisions, the budget up to year $t_{(L-1)}$ is deterministic and there are $(p_L=s_L.s_{L+1}.\ldots.s_\Omega)$ possible combinations for years $t_{(L-1)}+1$ to t_Ω. The final stage has a deterministic budget up to year $t_{(\Omega-1)}$ and $p_\Omega=s_\Omega$ budget possibilities for years $t_{(\Omega-1)}+1$ to t_Ω.

Scenarios of Budget Constraints

Budget for each program is not transferable across different programs. For instance, budget for pavement program is not supposed to be used for bridge program. However, multi-year budget for each program can either be constrained year-by-year or be treated as a cumulative budget for all years combined. Accordingly, system optimization can be conducted separately under the two scenarios. For year-by-year constrained budget scenario, a small amount may be left from a preceding year that is not sufficient to select any additional contract. The surplus can then be carried over to the following year so as to fully utilize available funds. For cumulative budget scenario, it is just a single-period analysis and no carryover is involved.

Model Formulation

In the case of a stochastic budget for a multi-year analysis period, actual budget in each year could be different from the expected budget. Stochastic formulation can reduce this gap by replacing the expected value with probabilistic values and replacing one-stage static decisions as in the deterministic model with multi-stage recourse decisions. This provides flexibility to decision-makers when the values of budget become manifested with passage of time. In this study, discrete probability distributions for budget possibilities in different years were considered. The stochastic model with Ω-stage budget recourses was formulated as a deterministic equivalent program. This was done by combining first stage decisions using the expected budget with the expected values of Ω-stage recourse functions for remaining stages (Birge and Louveaux, 2002).

Denote:

x_i = Decision variable of contract i, $i = 1, 2, \ldots, n$

a_i = Benefits of contract i, $i = 1, 2, \ldots, n$

c_{ikt} = Cost of contract i using budget from program category k in year t

ξ_L = Randomness associated with budget in stage L and decision space

$X_L(p)$ = Decision vector using budget $B_{kt}^L(p)$ in stage L, $X_L(p)= [x_1, x_2, \ldots, x_n]^T$

A = Vector of benefits of n contracts, $A = [a_1, a_2, \ldots, a_n]^T$

C_{kt} = Vector of costs of n contracts using budget from program category k in year t, $C_{kt} = [c_{1kt}, c_{2kt}, \ldots, c_{nkt}]^T$

$Q(X_L(p), \xi_L)$ = Recourse function in stage L

$E_{\xi_2}[Q(X_L(p), \xi_L)]$ = Mathematical expectation of the recourse function in stage L

$B_{kt}^L(p)$ = The p^{th} possibility of budget for program category k in year t in stage L

$p(B_{kt}^L(p))$= Probability of having budget scenario $B_{kt}^L(p)$ occur in stage L

$E(B_{kt}^L)$ = Expected budget in stage L, where $E(B_{kt}^L) = \sum_{p=1}^{p_L} \left[P\left(B_{kt}^L(p)\right) \cdot B_{kt}^L(p)\right]$

$p = 1, 2, \ldots, p_L=s_L.s_{L+1}.\ldots.s_\Omega$

$i = 1, 2, \ldots, n$

$k = 1, 2, \ldots, K$

$t = 1, 2, \ldots, M.$

A stochastic model with Ω-stage budget recourses with year-by-year budget constraints as a MCMDKP formulation is shown below:

Maximize $A^T.X_1 + E_{\xi_2}[Q_2(X_2(p), \xi_2)] + ... + E_{\xi_\Omega}[Q_\Omega(X_\Omega(p), \xi_\Omega)]$ (1)

Stage 1

Subject to $C_{kt}^T.X_1 \leq B_{kt}(p)^1 = E(B_{kt}^1)$ (2)

X_1 is a decision vector with 0/1 integer elements.

Stage 2

$E_{\xi_2}[Q_2(X_2(p), \xi_2)] = \max \{ {}^{A^T.X_2(p)} |_{B_{kt}^2(p)=E(B_{kt}^2)} \}$ (3)

Subject to $C_{kt}^T.X_2(p) \leq B_{kt}(p)^2$ (4)

$X_1 + X_2(p) \leq 1$ (5)

X_1 and $X_2(p)$ are decision vectors with 0/1 integer elements.

...

Stage L

$E_{\xi_L}[Q_L(X_L(p), \xi_L)] = \max \{ {}^{A^T.X_L(p)} |_{B_{kt}^L(p)=E(B_{kt}^L)} \}$ (6)

Subject to $C_{kt}^T.X_L(p) \leq B_{kt}{}^L(p)$ (7)

$X_1 + X_2(p) + ... + X_L(p) \leq 1$ (8)

$X_1, X_2(p), ..., X_L(p)$ are decision vectors with 0/1 integer elements.

...

Stage Ω

$E_{\xi_\Omega}[Q_\Omega(X_\Omega(p), \xi_\Omega)] = \max \{ {}^{A^T.X_\Omega(p)} |_{B_{kt}^\Omega(p)=E(B_{kt}^\Omega)} \}$ (9)

Subject to $C_{kt}^T.X_\Omega(p) \leq B_{kt}{}^\Omega(p)$ (10)

$X_1 + X_2(p) + ... + X_L(p) + ... + X_\Omega(p) \leq 1$ (11)

$X_1, X_2(p), ..., X_L(p), ..., X_\Omega(p)$ are decision vectors with 0/1 integer elements.

PROPOSED ALGORITHM FOR THE STOCHASTIC MODEL

Computational Complexity

Solution algorithms applicable to the stochastic model for the MCMDKP problem can be divided into two broad varieties: exact algorithms and heuristic algorithms. Exact algorithms generally include branch-and-bound, dynamic programming, and their variations. Heuristic algorithms pertinent to the problem mainly include Lagrangian relaxation techniques. The choice between exact and heuristic algorithms generally involves a tradeoff between solution quality and computational complexity. The Knapsack problem is NP-hard and time for exact algorithms increases exponentially with the size of problem instance. With this in mind, an efficient heuristic algorithm based on Lagrangian relaxation was developed.

Budget for Stage *L* Computation

Budget for each program category in stage L has the following characteristics: i) period from years 1 to $t_{(L-2)}$, it will be kept the same as that in stage L-1; ii) period from $t_{(L-2)}+1$ to $t_{(L-1)}$, it will inherit the budget used for computation in stage L-1 selected from $s_{(L-1)}$ possibilities; and iii) period from $t_{(L-1)}+1$ to last year t_Ω, it has $p_L = s_L \cdot s_{L+1} \cdot ... \cdot s_\Omega$ possible combinations. The choice of a specific budget combination is determined by the rule of minimum deviation between a budget possibility $B_{kt}{}^L(p)$ and expected budget $E(B_{kt}{}^L)$ in stage L, $\Delta B^L(p)$. This can be separately computed according the two budget scenarios, i.e., $\Delta B^L(p) = \sum_{k=1}^{K} \sum_{t=1}^{M} \left[\left(B_{kt}^L(p) - E(B_{kt}^L) \right)^2 \right]$ or $\Delta B^L(p) = \sum_{k=1}^{K} \left[\left(\sum_{t=1}^{M} B_{kt}^L(p) - \sum_{t=1}^{M} E(B_{kt}^L) \right)^2 \right]$.

Feasibility of Previous Stage Solution

First stage decisions in the proposed model are taken based on the expected budget. For the case of a lower budget level in the second stage, feasibility of first stage solution may already have been violated. This failure is corrected in the algorithm by checking at the beginning of each stage for

budget violations due to selections made in the previous stage. If a violation exists, contracts are removed in an optimal manner till feasibility of selection is satisfied.

Concept of Lagrangian Relaxation Techniques

The system optimization model for stage L operations can be redefined as follows:

Objective maximize $z(Y_L) = A^T.Y_L$ (12)

Subject to $C_{kt}^T.Y_L \leq B_{kt}^L$ (13)

Y_L is stage L decision vector with 0/1 integer elements.

Given non-negative, real Lagrangian multipliers λ_{kt}, the Lagrangian Relaxation of (12), $z_{LR}(\lambda_{kt})$, can be written as

$$\text{Objective} \quad z_{LR}(\lambda_{kt}) = \text{maximize} \left\{ A^T \cdot Y_L + \sum_{k=1}^{K}\sum_{t=1}^{M}\left[\lambda_{kt} \cdot \left(B_{kt}^L - C_{kt}^T \cdot Y_L\right)\right]\right\}$$

$$= \text{maximize} \left\{ \left(A^T - \sum_{k=1}^{K}\sum_{t=1}^{M}\left(\lambda_{kt} \cdot C_{kt}^T\right)\right) Y_L + \sum_{k=1}^{K}\sum_{t=1}^{M}\left(\lambda_{kt} \cdot B_{kt}^L\right)\right\} \quad (14)$$

Subject to Y_L with 0/1 integer elements.

As $\sum_{k=1}^{K}\sum_{t=1}^{M}\left(\lambda_{kt} \cdot B_{kt}^L\right)$ in (14) is a constant, optimization can just be concentrated on the first term, namely, maximizing $\left(A^T - \sum_{k=1}^{K}\sum_{t=1}^{M}\left(\lambda_{kt} \cdot C_{kt}^T\right)\right).Y_L$ (Volgenant and Zoon, 1990). (15)

The solution to (15) is Y_L^*, where $Y_L^* = \begin{cases} 1, \text{ if } \left(A^T - \sum_{k=1}^{K}\sum_{t=1}^{M}\left(\lambda_{kt} \cdot C_{kt}^L\right)\right) > 0 \\ 0, \text{ otherwise} \end{cases}$ (16)

Then, Y_L^* maximizes $z(Y_L) = A^T.Y_L$, subject to Y_L with 0/1 integer elements.

In order to obtain optimal solution by maximizing $z(Y_L) = A^T.Y_L$, only subject to Y_L with 0/1 integer elements, the following condition needs to be satisfied $\sum_{k=1}^{K}\sum_{t=1}^{M}\left[\lambda_{kt} \cdot \left(B_{kt}^L - C_{kt}^T \cdot Y_L\right)\right] = 0$ (17)

In this regard, stage L optimization operations need to focus on determining Lagrangian multipliers λ_{kt} such that i) Y_L^* obtained in (16) is feasible to original optimization model, i.e., $C_{kt}^T.Y_L \leq B_{kt}^L$ is valid, and ii) condition (17) is satisfied to maintain optimality to original optimization model.

Main Steps of Proposed Algorithm for Stage L Computation

Part I: Find a Feasible Solution prior to Budget Carryover from Year to Year
Step 0: Add all contracts and sort their benefits in descending order
Step 1: Initialize Lagrangian multipliers and normalize contract costs and budget
Step 2: Determine the least total cost constraint for each budget category for all years
Step 3: Compute increase of Lagrangian multipliers based on benefit-to-cost ratio by contract
Step 4: Remove smallest benefit-to-cost ratio contract one at a time until no budget violation
Step 5: Improve solution by adding removed contract, stop if no improvement

Part II: Improve the Feasible Solution by Carrying Over Remaining Budget to Next Year
Step 1: Hold solution from Part I for the first year
Step 2: Increase budget for the second year, re-optimize starting from the second year
Repeat the process for each year until the last time period.

CASE STUDY

Data Source and Model Applications

Data on candidate projects for Indiana state highway programming in 1998-2001 were used to apply the proposed stochastic model and solution algorithm. To assess the impact of budget uncertainty on project selection results, the stochastic model was also applied to the same data set without considering budget recourses, namely, the model was treated as a single-stage deterministic model by only executing the first stage optimization using the expected budget.

Project costs were directly extracted from the data set and project benefits were separately computed for cases under certainty, risk and uncertainty using existing value functions (Li and Sinha, 2004). Information on estimated budget and actual budget available for period 1998-2001 was used to establish low, medium, and high budget levels as well as respective probabilities for each year. Data analysis revealed that actual budget differs on average by 1 percent with 8 percent deviation as compared to the estimated budget for the four-year period.

Discussion of Results

Project selection results from the model applications are shown in Table 1. Regardless of whether considering budget recourses, the analysis based on cumulative budget scenario yielded selection of a higher number of contracts as compared to those based on year-by-year budget scenario. The results are not unexpected. Cumulative budget scenario did not have year-by-year restrictions as those imposed to year-by-year budget scenario. This enabled project selection in a more flexible and holistic manner, leading to an improved solution. For year-by-year budget scenario, the matching rate of selected contracts in the two model applications ranges from 89-95 percent. For cumulative budget scenario, this rate is 94-97 percent. Conversely, differences in project selection results are 5-11 percent and 3-6 percent, respectively.

TABLE 1. Project Selection Results Using the Stochastic Model with and without Ω-Stage Recourses

Year	Contract Benefits Expressed by	Candidate Contracts	Year-by-Year Constrained Budget				Cumulative Budget			
			SP_W	$SP_{W/O}$	Match Rate		SP_W	$SP_{W/O}$	Match Rate	
1998	1) Utility value	429	389	405	370	91%	410	411	395	96%
	2) Exp. Utility value		390	417	382	92%	411	412	397	96%
	3) Shackle's ratio		393	415	382	92%	411	412	397	96%
1999	1) Utility value	412	361	374	331	89%	385	391	370	95%
	2) Exp. Utility value		364	375	335	89%	385	392	370	94%
	3) Shackle's ratio		369	375	339	90%	384	391	368	94%
2000	1) Utility value	611	523	520	479	92%	552	548	516	94%
	2) Exp. Utility value		523	520	481	93%	563	554	527	95%
	3) Shackle's ratio		526	515	479	93%	561	558	530	95%
2001	1) Utility value	418	397	417	396	95%	404	401	390	97%
	2) Exp. Utility value		393	417	392	94%	405	399	388	97%
	3) Shackle's ratio		394	417	393	94%	406	400	389	97%
Total	1) Utility value	1,870	1,670	1,716	1,576		1,751	1,751	1,671	
	2) Exp. Utility value		1,670	1,729	1,590		1,764	1,757	1,682	
	3) Shackle's ratio		1,682	1,722	1,593		1,762	1,761	1,684	

Note: SP_W and $SP_{W/O}$ for the stochastic model with and without Ω-stage recourses, respectively.

CONCLUDING REMARKS

This paper presents a stochastic model and solution algorithm for highway project selection considering budget uncertainty. The model was applied in a case study for systemwide project selection using past candidate projects for state highway programming in Indiana. It was found that an average of 1 percent gap between estimated and actual budgets could result in as high as 11 percent change in project selection results. Budget uncertainty should therefore be taken into consideration in the investment decision process in order to achieve robust results. The proposed model and algorithm could be adopted by transportation agencies for enhanced highway investment decision-making in long-range planning and project programming.

REFERENCES

Birge, J.R., and Louveaux, F. (1997). *Introduction to Stochastic Programming*. Springer-Verlag, New York, NY.

Martello, S., and Toth, P. (1990). *Knapsack Problems: Algorithms and Computer Implementations*. John Wiley & Sons, Chichester, United Kingdom.

Li, Z., and Sinha, K.C. (2004). A methodology for Multicriteria Decision Making in Highway Asset Management. *Transportation Research Record*, 1885, 79-87.

Volgenant, A., and Zoon, J.A. (1990). An Improved Heuristic for Multidimensional 0-1 Knapsack Problems. *Journal of the Operational Research Society*, 41(10), 963-970.

Optimizing Facility Component Maintenance, Repair, and Restoration Investment Strategies Using Financial ROI Metrics and Consequence Analysis

M.N. Grussing[1], D.R. Uzarski[2], and L.R. Marrano[3]

[1]Engineer Research and Development Center - Construction Engineering Research Laboratory (ERDC-CERL), P.O. Box 9005, Champaign, IL 61826-9005; Phone: 217-398-5307; Fax: 217-373-3490; email: michael.n.grussing@erdc.usace.army.mil
[2]Unity Consultants, 2011 Barberry Circle, Champaign, IL 61821; Phone: 217-398-3984; email: d.uzarski@insightbb.com
[3]Engineer Research and Development Center - Construction Engineering Research Laboratory (ERDC-CERL), P.O. Box 9005, Champaign, IL 61826-9005; Phone: 217-373-4465; Fax: 217-373-3490; email: lance.r.marrano@erdc.usace.army.mil

Abstract

In order to proactively manage facility assets and allocate resources to optimize facility performance, infrastructure asset management professionals at all levels need improved decision support tools and processes. This paper addresses the business process framework for developing a comprehensive facility infrastructure investment plan using information about current and projected future asset condition. This process starts by identifying which building components are candidates for corrective repair or replacement using standards and policies linked to acceptable building performance requirements. Next, a prioritization schema determines which work actions are most important for funding when budgets are constrained. Priorities are developed through importance and condition metrics, risk assessment, penalty costs of not doing work, and return on investment (ROI) analysis and metrics. These financial metrics also help to determine the best option between component stop-gap repair, major repair, or total replacement alternatives by evaluating the economic return over the lifecycle of the asset. Finally, consequence analysis simulations determine the impact of difference standards, policies, and budget levels, further maximizing building performance and return on investment.

Introduction

Decision support technology for planning building facility maintenance, repair, and capital renewal has evolved over the past several decades from a reactive breakdown mode to a predictive, reliability centered approach (Moubray 2002). This evolution has been predicated on improved assessment techniques and metrics, which provides the basis for better decision support and justification for work requirements. However, a full and detailed building assessment program and work requirements formulation can be prohibitively expensive for many organizations, and they revert back to the short term costs savings (through reduced inspections) associated with reactive maintenance and repair. In the long term, this decision eventually leads to chaotic fiscal swings and inopportune downtimes for many maintenance organizations. So as the shifting focus towards proactive asset management calls for accurate accounting and assessment of building infrastructure, it also requires an improved knowledge-based process to balance repair and inspection resource planning based on acceptable risk and asset criticality. This involves just-in-time assessments and timely corrective action before advancing degradation leads to more costly repairs or impact on organizational goals and mission. Next-generation processes for condition assessment, condition prediction, work requirements generation, prioritization, and consequence analysis are all important features of this changing asset management environment.

State of the Practice in Building Facility Asset Management

The merit of proactive maintenance and inspection of buildings is not a new concept. Past and current practices employ trained engineers or specialty technicians to inspect and evaluate various aspects of a building. These inspections are designed primarily to 1) ensure the safety of the building for the occupants; 2) determine maintenance, repair, or modernization requirements. With each assessment, the inspector surveys the building and records deficiencies, which relate to corrective service calls or repair and restoration work. The inspector may also assign a priority code which reflects the importance or urgency of the deficiency. This puts the work prioritization process in the hands of several different inspectors, instead of under centralized, consistent, information driven process.

After the inspection identifies building deficiencies that need to be addressed, a project engineer or technician often will draw up a detailed plan for correcting the deficiencies, develop a bill of materials, and a cost estimator will determine the project costs. This labor intensive and expensive process results in a "job jar" of projects based on the often subjective input from many different inspectors. In addition, if the total work requirement exceeds the budgeted amount, (as is most often the case) wasted design and estimating effort results from unfunded projects that never get completed. Therefore, inspection programs can become questionable in value and many programs end up abandoned due to lack of funding or justification to support them.

Performance-Based Metrics

To improve on this process described above, facility performance metrics are needed which are 1) affordable to obtain within the organizational framework; 2) Consistent, objective, and repeatable across different inspectors and assessors; 3) engineering and science-based; and 4) correlated to physical condition, work requirements, and resource justification. One single metric does not address all these requirements. Instead, a collection of engineering and economic derived metrics provides a toolkit for facility managers to base investment decisions. These metrics include:

- A Condition Index (CI) metric that provides an objective measure of the physical condition of an asset on a 0-100 point scale from a standardized distress-based, not deficiency-based, inspection process (Uzarski and Burley 1997),
- A Functionality Index (FI) metric that provides an objective measure of built-in capability of an asset or facility to perform its required function based on defined organizational mission or goals, obsolescence, and codes/regulations. The functionality metric is measured on a similar 0-100 point scale based from a standardized functionality evaluation,
- A Performance Index (PI) metric that is a function of condition and functionality, and relates how well a building performs as required by mission as the building ages, degrades, and obsolesces,
- A Remaining Service Life (RSL) metric that relates the expected remaining amount of time a building component asset will perform at or above a required minimum level of performance,
- A Remaining Maintenance Life (RML) metric that relates the expected remaining amount of time a building component will perform above a desired condition standard,
- A Component Importance Index (CII) that measures the criticality of a component asset against risk of failure, and
- A Mission Dependency Index (MDI) that measures the importance of a building with respect to agency or organizational mission. (Mission Dependency Index)

These metrics describe the state of the building or component asset at a measured point in time based from assessment information. In addition, the Condition Index metric provides a means for predicting future asset condition and reliability using trend analysis. This allows for the development of long term work plan strategies and consequence analysis. The CI is the main metric discussed within the scope of this paper.

Determining Candidates for Corrective Action

The Condition Index describes the absolute condition of an asset on 0-100 scale (Table 1). It also establishes a means to compare all building component conditions on a relative scale. A condition standard specifies a threshold CI value for an asset or group of assets to establish a target condition below which the assets become candidates for corrective action. Not all components in a building are required to be maintained at the same standard level. Therefore, multiple condition standards may exist and a policy based on organizational goals and accepted risks will identify which components should be maintained at a high condition and which can be maintained at lower condition levels. These standards can be selectively applied based upon several attributes of the components, such as the type of building or building system the component resides in, and the effect of component failure on mission.

Table 1. Condition Index Definitions

Condition Index	Definition
100-85 Good	Slight or no serviceability or reliability reduction overall to component.
85-70 Satisfactory	Component serviceability or reliability is degraded but adequate.
70-55 Fair	Component serviceability or reliability is noticeably degraded
55-40 Poor	Component has significant serviceability or reliability loss.
40-25 Very Poor	Unsatisfactory serviceability or reliability reduction
25-10 Serious	Extreme serviceability or reliability reduction
10-0 Failed	Overall degradation is total.

Evaluation of Corrective Work Action Scenarios using ROI

The cost to accomplish the work can be estimated from a model. Each building component has an associated replacement value. When component repair is warranted, the estimated corrected work action cost is some percentage of total replacement cost. As illustrated in Figure 1 below this percentage is a function of CI. It also varies by component. Due to the curvilinear nature of the repair $ versus CI curve, deferring repair opportunities early in a component lifecycle can have a significant impact later on repair costs, making repair a less viable option than total replacement.

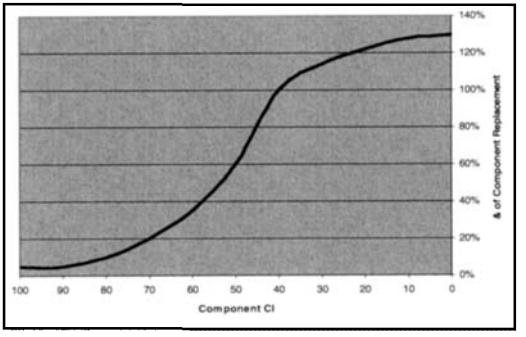


Figure 1. Component Repair Cost as a Function of CI.

The key to minimizing the total component lifecycle cost is by maximizing the benefit or return on any repair or capital investments over the lifecycle. Figure 2 shows the typical condition profile of a component over its expected service life, with the lifecycle response due to several corrective action alternatives. As the component operates in service and ages, its condition degrades, usually after several years of relatively high steady-state condition. When this degradation reaches the threshold limit set by the standard, corrective action alternatives are explored. These corrective actions include: 1) do nothing – allow the component to run until failure; 2) Stop-Gap repair; 3) major component repair; and 4) total component replacement.

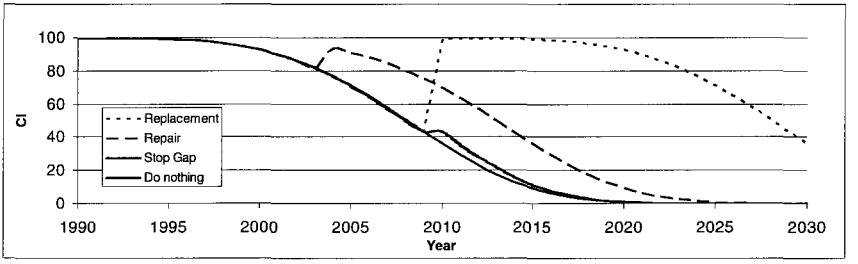


Figure 2. *Component Lifecycle Trends for Corrective Action Alternatives*

Run to Failure

If no corrective action is performed on the component, its condition, performance, and eventual reliability will degrade to such a point, termed the CI terminal value, where failure has essentially occurred. At this point, component replacement is required. The run to failure option is viable only if component failure does not have a significant impact on mission.

Stop Gap Repair

For some components where advancing degradation is evident, measures can be taken to slow or halt this degradation for some finite amount of time. These measures are limited in scope and do not improve or restore condition, but they can prevent the component from degrading below some minimum performance requirement until a more permanent solution can be accomplished. Stop gap repairs are usually less attractive from an economic perspective, because they do not improve condition and only defer larger major repairs or replacement for a short amount of time, but may be needed due to a lack of funds for a permanent solution.

Major corrective Repair

Major repair improves the condition of the component (not necessarily to 100) sometime prior to failure. Due to this increase in condition, the component's remaining service life is also extended. This additional service life, brought on by repair, defers the capital cost of replacement due from impending failure. Therefore, the monetary benefit of a repair can be calculated by taking the additional service life generated by the repair multiplied by the amortized expenditure of component replacement. This monetary benefit, divided by the cost of the repair, determined the return on the investment (ROI) (ASTM E833-02a). The repair of a component also has benefits due to improved operational performance and reliability which are more difficult to quantify in monetary terms.

Replacement/Capital Renewal

Complete component replacement essentially resets the component lifecycle clock. When replacement is performed, the component CI is restored to its maximum (CI=100) and the full expected service life is reset. Since replacement does not defer capital expense, its return of investment ratio based on the logic described for

repair is essentially 1. However, replacement of a component may involve some modernization, potentially resolving obsolescence issues. This, in turn, can lead to benefits of efficiency and lower maintenance, operations, or energy costs, which should be accounted for in any ROI calculation (ASTM E1074-93).

The procedures discussed above provide a logical framework for calculating an ROI for different corrective action alternatives. The analysis can also be used to determine the best time for scheduling and executing component repair or replacement. The extended service life gained from a corrective repair action, and the cost of the repair both depend on when that action takes place, and the condition of the asset at the time. Hence, there is an optimal point for each building component for corrective work to occur. The ROI analysis procedures can be used to pinpoint optimal work timing and establish the remaining time in service before maintenance is warranted.

Prioritizing Work

Typically, the total estimated costs for all corrective work item candidates will exceed the authorized budget. Hence, a ranked work item list is necessary. This ranking should reflect the importance of both the building and component and the potential ROI for doing work at the specified time plus other variables. In so doing, an objective prioritization scheme must be based on the priorities of the organization and the risk associated with investment alternatives.

The development of a prioritization scheme starts with the definition of organizational objectives and the evaluation of how well a given component or its work action meets those objectives. This is done by specifying attributes of the building, component and work item which can be related to importance measures. For example, one objective is accomplishing the most cost efficient work items. In this case, the main prioritization criterion is the calculated ROI metric. Another, sometimes competing objective, is repairing the most important component based on mission criticality. Here, the different measures associated with building use type, building systems, and the MDI and CII metrics are prioritization criteria. By assigning relative weights to the different measures and objectives, a consistent and repeatable importance score is calculated for each work item, which is then used to rank work items and establish the funding cut line based on the budget.

Consequence Analysis

A long term maintenance, repair, and capital renewal plan for an organization can involve a portfolio of several buildings with hundreds of building components in each, all at varying condition states. This plan may seek to identify work requirements over a five to ten year horizon. With the numerous asset elements involved, optimizing a strategy that incorporates current user requirements, budget constraints, and future performance sustainment can be a difficult challenge. Using a structured business process framework, and component degradation analysis, different investment decision scenarios can be explored, and consequences can be evaluated over a long-term horizon.

One such automated consequence analysis tool is the IMPACT simulation model (IMPACT) used in conjunction with the BUILDER® Engineered Management System (BUILDER). This model simulates the annual fiscal cycle of work planning/executing and displays building, system, and component conditions up to ten years into the future. Model inputs include the real property inventory, condition information and deterioration trends, current work projects, budget projections, and user defined standards, policies, and prioritization schemes to initialize the model. The simulation then 1) generates work requirements based from projected conditions, user defined standards and policies 2) prioritizes work actions 3) assigns funding to highest priority work items using set budget resources 4) simulates the execution and completion of funded work 5) predicts the future condition of component assets based on work completed and deferred and 6) updates the component inventory database to reset the cycle for each year in the simulated budget plan.

After simulation, a complete picture of the condition response of the portfolio of assets is known for the duration of the scenario horizon (Figure 3). In addition, the maintenance and repair (M&R) backlog due to deferred work requirements is available for each year in the plan (Figure 4). This gives facility managers and

decision makers the ability to see the effects of budget and policy on the condition of facility assets over time, and adjustments can be made accordingly to maximize performance and return on investment.

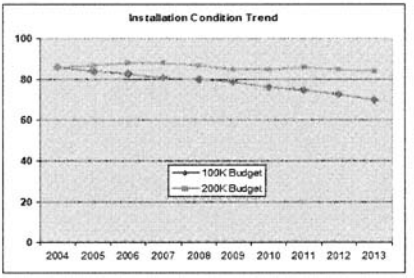


Figure 3. IMPACT Scenario Condition Trend

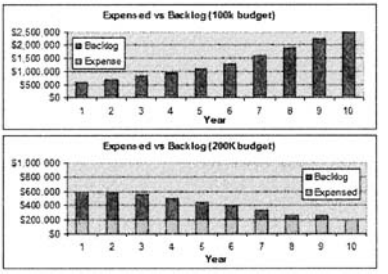


Figure 4. IMPACT Scenario M&R Backlog

Conclusions

Fundamental to any successful facility infrastructure investment strategy are the objectives to 1) minimize lifecycle ownership costs, 2) maximize facility performance, and 3) manage risk. The process described herein provides the framework and decision support to help facility managers achieve these objectives. A distress-based condition assessment process is standardized, affordable, and results in a Condition Index metric that provides performance-based information about condition state, condition prediction trends, and remaining service life that are objective, consistent, and meaningful across all building components and infrastructure domains. The CI establishes the foundation for risk-based condition standards, component reliability projections, and best cost work planning and timing, and presents a rational and streamlined approach to work generation, work estimation, financial analysis computation, and prioritization. Finally, the consequence analysis routines provide the critical ability to evaluate the effect of infrastructure investment policies against the fundamental strategic objectives. This supports proactive and accountable building infrastructure management. By implementing this approach, such as through the BUILDER® EMS, a scalable, enterprise-wide solution can support work planning activities, and achieve organizational organization or agency-level mission, goals, and performance benchmarks for facilities.

References

ASTM International (2002). Standard Terminology of Building Economics, ASTM E 833-02a.

ASTM International (1993). Standard Practice for Measuring Net Benefits for Investments in Buildings and Building Systems, ASTM E 1074-93.

BUILDER (2005). BUILDER Engineered Management System, http://dwww.cecer.army.mil/td/tips/product/details.cfm?ID=64&TOP=1

IMPACT (2005). IMPACT (Integrated Multi-Year Prioritization and Analysis Tool), http://dwww.cecer.army.mil/td/tips/product/details.cfm?ID=738&TOP=1

MDI (2005). Mission Dependency Index, http://www.nfesc.navy.mil/shore/mdi/index.html

Moubray, John. (2002). Reliability-Centered Maintenance, Industrial Press Inc., New York.

Uzarski, D.R. and L.A. Burley. (1997). "Assessing Building Condition by the Use of Condition Indexes," proceedings of the ASCE Specialty Conference Infrastructure Condition Assessment: Art, Science, Practice, pp.365-374.

A Fuzzy Multiple Objective Mixed Integer Programming Approach for Carrier Selection and Freight Volume Allocation with Revenue Based Discount Schemes

Chuanxu Wang

School of Economy and Management, Shanghai Maritime University, Shanghai, 200135, PH (8621) 58606090; FAX (8621) 58606090; email: cxwang@shmtu.edu.cn

Abstract

Based on fuzzy multiple objective mixed integer programming, a model is introduced to simultaneously determine the number of carriers to employ and freight volume to allocate to these carriers under a multiple- cargo, multiple- carrier scenario. The model jointly considers the objectives of transportation cost, cargo loss and damage, and cargo delivery performance as well as the freight rate discount based on total transportation revenue. Furthermore, the solution procedure for this model is illustrated. At last, a numerical example is provided to verify the availability of above model and perform the results analysis. It is shown from results analysis that the transportation cost, cargo loss and damage quantity, and cargo late delivery quantity are increasing in the degree of vagueness in carriers' transportation capacity.

Key Words:

Carrier selection, Freight allocation, Fuzzy multiple objective mixed integer programming, Revenue Based Discount schemes

1 Introduction

The selection of carriers and the determination of freight volume to allocate to these carriers have been an important area in the effective transportation operations management for many shippers and may greatly influence a shipper's competitive ability in transportation market. At present, many studies have been centered on carrier selection and carrier performance evaluation (Gentry, 1996, Pedersen and Gray, 1998). Previous studies have also examined the perceptional differences between shippers and carriers concerning carrier selection variables and importance of service attributes (Jerman et al, 1989, Abshire and Premeaux, 1991, Kent and Parker, 1999). Recently, Lu (2003) has specifically explored carrier performance within the context of its influence on partnering relationships. While many studies on carrier selection and performance evaluation exits, little research addresses the method that simultaneously considers carrier selection and freight allocation.

Carrier selection problem is complicated in nature and needs to be considered on multiple criteria. In practice, the input data or decision parameters, such as capacity,

cost and objective function are often imprecise because some information is uncertain. Therefore, the carrier selection has the characteristics of impreciseness and fuzziness. In this paper, we use fuzzy multiple objective programming model to formulate and solve the problem of carrier selection and freight allocation under a multiple- cargo, multiple- carrier scenario, and then demonstrate the application of the model through a numerical example. Further, our model considers the freight rate discount schedules based on transportation revenue. In traditional discount schedules, freight rate discounts are based on freight volume. Because of the advent of advanced technologies, cargoes demanded by customer are increasingly smaller, lighter and more value-added. Therefore, it is more meaningful to give discounts based on the total transportation revenue (Refer to Dahel 2003 for the Sales revenue based discount schemes considered in product purchasing decision).

2 A Fuzzy Programming Model for Carrier Selection and Freight Allocation

We consider a situation in which $i = 1,2,...,I$ items are to be transported by $m = 1,2,...,M$ carriers to $n = 1,2,...,N$ destinations. Carriers provide different levels of freight rate, service quality, and delivery performance and transportation capacity for each item they transport. Furthermore, based on the received total transportation revenue, carrier m offers a discount having $r = 1,2,...,R_m$ discount categories.

2.1 Notation

Parameters

M : Number of carriers; I : Number of items; N : Number of destinations; M_i : Number of carriers transporting item i ; I_m : Number of items transported by carrier m ; N_m : Number of destinations to which the cargo can be transported by carrier m ; N_i : Number of destinations demanding item i ; D_{in} : units of item i demanded by destination n ; p_{imn} : Freight rate of item i transported by carrier m to destination n ; q_{imn} : Loss and damage rate of item i transported by carrier m to destination n ; t_{imn} : Late delivery rate of item i transported by carrier m to destination n ; S_{im} : Transportation capacity for item i provided by carrier m ; u_{mr} : Upper bound of discount category r provided by carrier m ; d_{mr} : Discount rate associated with discount category r provided by carrier m .

Decision Variables

x_{imn} : units of item i transported by carrier m to destination n ; y_{mr} : transportation revenue received by carrier m in discount category r ; $z_{mr} = 0/1$, If the transportation revenue received by carrier m falls within category r , $z_{mr} = 1$; otherwise, $z_{mr} = 0$ 。

2.2 Mathematical Model

A mathematical programming model to formulate the above problem is given as follows:

$$\text{(P1) } \min C_1 = \sum_{m=1}^{M}\sum_{r=1}^{R_m}(1-d_{mr})y_{mr}, C_2 = \sum_{i=1}^{I}\sum_{m=1}^{M_i}\sum_{n=1}^{N_i} q_{imn}x_{imn}, C_3 = \sum_{i=1}^{I}\sum_{m=1}^{M_i}\sum_{n=1}^{N_m} t_{imn}x_{imn}, \quad (1)$$

$$\text{subject to: } \sum_{m=1}^{M_i} x_{imn} = D_{in}, \quad i = 1,2,...I, n = 1,2,...N_i \quad (2)$$

$$\sum_{n=1}^{N_i} x_{imn} \le S_{im}, \qquad i = 1,2...I, m = 1,2,...M_i \quad (3)$$

$$\sum_{i=1}^{I_m}\sum_{n=1}^{N_i} p_{imn}x_{imn} = \sum_{r=1}^{R_m} y_{mr}, \quad m = 1,2,...M_i \quad (4)$$

$$y_{mr} \le u_{mr}z_{mr}, \quad m = 1,2,...M, r = 1,2...R_m \quad (5)$$

$$y_{m,r+1} \ge u_{mr}z_{m,r+1}, \quad m = 1,2,...M, r = 1,2,..., R_m - 1 \quad (6)$$

$$\sum_{r=1}^{R_m} z_{mr} = 1 \qquad m = 1,2,...M \quad (7)$$

$$z_{mr} = 0/1, \quad m = 1,2,...M, r = 1,2,...R_m; \quad y_{mr} \ge 0, m = 1,2,...M, r = 1,2,...R_m \quad (8)$$

$$x_{imn} \ge 0, \quad i = 1,2,...I, m = 1,2,...M_i, n = 1,2,...N_m \quad (9)$$

Objective function (1) minimizes the total transportation cost, the cargo loss and damage quantity as well as the cargo late delivery quantity of carriers.

2.3 Model with Fuzzy Parameters

In our model, we capture the shipper's ambiguity about the fuzzy information related to the transportation cost, cargo loss and damage quantity, cargo late delivery quantity and carriers' capacities by transforming the (P1) model into the following (P2) model.

$$\text{(P2) } \sum_{m=1}^{M}\sum_{r=1}^{R_m}(1-d_{mr})y_{mr} \lesssim \tilde{C}_1, \quad \sum_{i=1}^{I}\sum_{m=1}^{M_i}\sum_{n=1}^{N_i} q_{imn}x_{imn} \lesssim \tilde{C}_2, \sum_{i=1}^{I}\sum_{m=1}^{M_i}\sum_{n=1}^{N_m} t_{imn}x_{imn} \lesssim \tilde{C}_3, \quad (10)$$

$$\text{subject to: } \sum_{n=1}^{N_i} x_{imn} \underset{\sim}{<} \tilde{S}_{im}, \qquad i = 1,2...I, m = 1,2,...M_i, \tag{11}$$

as well as conditions given in (2), (4), (5), (6), (7), (8) and (9).
where the symbol "$\underset{\sim}{<}$" indicates "essentially smaller than or equal to " and allows one reach some aspiration level, $\tilde{C}_1, \tilde{C}_2, \tilde{C}_3$ and $\tilde{S}_{im}$ denote fuzzy values.

In this paper, a linear membership function defined in Bellman and Zadeh (1970) has been considered for all fuzzy parameters in (P2) problem.

2.4 Equivalent Formulation

According to Zimmermann (1978), the fuzzy linear programming model (P2) can be transformed into the equivalent linear programming model with the objective function of "Maximize L" as well as with constrains (2), (4), (5), (6), (7), (8), (9),

$$L(C_i^{\max} - C_i^{\min}) + C_i(x) \le C_i^{\max} (i = 1,2,3), L(S_{im}^U - S_{im}^L) + \sum_{n=1}^{N_i} x_{imn} \le S_{im}^U \quad (i = 1,2..I, m = 1,2,..M_i),$$

and $0 \le L \le 1$. In this model, L represents the maximun degree of overall satisfaction with decision-making's multiple fuzzy goals. $C_i^{\min(\max)} = \min(\max) C_i(x^*)$, x^* is optimal solution. S_{im}^L and S_{im}^U represent, respectively, the lower bound and the upper bound of parameter S_{im}. $C_i^{\min}$ and $C_i^{\max}$ for each objective can be obtained from the individual optima using the method mentioned in Zimmermann (1978).

3 A Numerical Example

We consider a hypothetical numerical example in which two items are to be transported by three carriers to three destinations. The demands from destination 1 and 2 for Item1 are 1200 tons and 1000 tons respectively. The demands from destination 1, 2 and 3 for Item2 are 800 tons, 1000 tons and 2000 tons, respectively. The other parameter specifications are given in Table1—Table 3.

3.2 Solution

We first obtain the initial solutions for each of the objective functions using conventional linear programming model. The results are $C_1^{\max} = 154565$, $C_1^{\min} = 93241$, $C_2^{\max} = 14.2$, $C_2^{\min} = 8.2$, $C_3^{\max} = 36$, $C_3^{\min} = 24$. Then we formulate the fuzzy multiple objective linear programming and its equivalent linear programming model presented in Section 2.We use the computer to run the equivalent linear programming model which yields the following results: $C_1 = 135143.4$, $C_2 = 12.2998$, $C_3 = 32.1996, L^* = 0.7167, x_{131}^* = 1200, x_{132}^* = 1000,$

$x_{211}^{*}=500, x_{212}^{*}=1000, x_{221}^{*}=300, x_{223}^{*}=1600, x_{233}^{*}=400, y_{11}^{*}=2000, y_{23}^{*}=57200, y_{33}^{*}=50000,$ $z_{11}^{*}=z_{23}^{*}=z_{33}^{*}=1$. The other decision variables are zeros.

Table 1 Basic data for different items transported by different carriers to different destinations

Item		Destination1			Destination 2			Destination 3		
		Carrier1	Carrier 2	Carrier 3	Carrier 1	Carrier 2	Carrier 3	Carrier 1	Carrier 2	Carrier 3
1	Freight rate (RMB Yuan/ton)	20	21	15	25	19	18			
	Loss & damage rate (%)	0.2	0.2	0.3	0.3	0.2	0.3			
	Late delivery rate (%)	0.5	0.4	0.6	0.5	0.4	0.6			
2	Freight rate (RMB Yuan/ton)	12	20	24	14	28	30	21	32	35
	Loss & damage rate (%)	0.2	0.1	0.2	0.2	0.1	0.2	0.2	0.1	0.2
	Late delivery rate (%)	0.5	0.4	0.6	0.5	0.4	0.6	0.5	0.4	0.6

Table 2 Transportation capacity provided by carriers for different items (tons)

Item	Carrier1		Carrier2		Carrier3	
	Lower bound	Upper bound	Lower bound	Upper bound	Lower bound	Upper bound
1	80000	84000	100000	105000	150000	157500
2	80000	84000	100000	105000	150000	157500

Table 3 Discount category and discount rate provided by carriers

Carrier	1		2			3		
Discount category (Tons)	0-2000	2000 and over	0-8000	8000-48000	48000 and over	0-10000	10000-50000	50000 and over
Discount rate (%)	0	10	0	5	10	0	5	10

3.3 Sensitivity Analysis

We perform the analysis on the impact of vagueness degree in carriers' transportation capacities. It can be shown that the increase in allowed changed interval $S_{im}^{U}-S_{im}^{L}$ will lead to the increase of vagueness level in transportation capacity We keep the value of S_{im}^{L} unchanged and increase the value of $S_{im}^{U}-S_{im}^{L}$ with the increment of 5% of S_{im}^{L}. Then we solve the fuzzy model for each $S_{im}^{U}-S_{im}^{L}$. The results can be shown in Table 4. It can be shown from Table 4 that the transportation cost, cargo loss and damage quantity and cargo late delivery quantity are increasing in the value of $S_{im}^{U}-S_{im}^{L}$. This indicates the value of each objective is increasing in the degree of vagueness in carriers' transportation capacity.

4 Conclusions

This paper has presented a model in selecting carriers and determining the carriers' freight allocation with uncertain carrier transportation capacity. It has considered the discount schemes offered by carries based on transportation revenue. Furthermore, This paper has also illustrated the solution procedure for this model by transforming the fuzzy multiple objective programming into equivalent conventional linear programming. At last, a numerical example analysis is provided to demonstrate the

effectiveness of proposed model. Meanwhile, the results analysis through above example shows that the value of the each objective is increasing in the level of vagueness in carriers' transportation capacity.

Table 4 The value of each objective at different vagueness degree in transportation capacity

Objective Values	The increment percentage of the lower bound of transportation capacity (%)						
	5	10	15	20	25	30	35
Transportation cost (RMB Yuans)	135143	137305	141424	145809	152224	160597	171517
Cargo loss and damage (Tons)	12.300	12.595	13.162	13.820	14.649	15.674	16.944
Cargo late delivery (Tons)	32.200	32.715	33.794	35.146	36.903	39.118	41.934

Acknowledgements

This paper is supported by the National Natural Science Foundation of China (Grant No. 70573068), Shanghai Social Science Foundation (Grant No. 05BJB014), Shanghai Leading Academic Discipline (Grant No.T0602) and Shanghai Pujiang Program.

References

Abshire, R.D., Premeaux, S.R., (1991). Motor carrier selection criteria: perceptual differences between shippers and carriers, *Transportation Journal,* 31(1), 31-35.

Bellman, R.E., Zadeh, L.A., (1970), Decision making in a fuzzy environment, *Management Sciences*, 17, B141-B164.

Dahel, N.E., (2003). Vendor selection and order quantity allocation in volume discount environment, *Supply Chain Management,* 8(4), 335-342

Gentry, J.J., (1996). The role of carriers in buyer-supplier strategic partnership: a supply chain management approach, *Journal of Business Logistics,* 17(2), 35-55

Jerman, R.E., R.D. Anderson, J.A. Constantin, (1989). Shipper versus carrier perceptions of carrier selection variables, *International Journal of Physical & Material Management,* 9(1), 29-38.

Kent, J.L., Parker, R.S., (1999). International containership carrier selection criteria: shippers/carriers differences. *International Journal of Physical Distribution and Logistics Management*, 29(6), 398-408.

Lu, C.S., (2003), An evaluation of service attributes in a partnering relationship between maritime firms and shippers in Taiwan, *Transportation Journal*, 42(5), 5-12.

Pedersen, E.L., Gray, R., (1998). The transport selection criteria of norwegian exporters, *International Journal of Physical Distribution and Logistics Management*, 28(2), 108-120.

Zimmermann, H.J., (1978). Fuzzy programming and linear programming with several objective functions. *Fuzzy Sets and Systems,* 1, 45-56.

A Non-Contact System to Detect and Quantify Segregation in Hot Mix Asphalt Pavements

By Edgar de León[1] and Gerardo W. Flintsch[2]

[1]Graduate Research Assistant, Virginia Tech Transportation Institute, 3500 Transportation Research Plaza, Blacksburg, VA 24061-0536, voice (540) 231 1021, fax (540) 231 1555, email: edeleoni@vt.edu

[2]Associate Professor, Transportation Fellow, The Via Department of Civil and Environmental Engineering, 200 Patton Hall, Virginia Tech, Blacksburg, VA 24061-0105, voice (540) 231 9748, fax (540) 231 7532, email: flintsch@vt.edu

Abstract

Current Hot Mix Asphalt (HMA) pavements are designed to withstand expected traffic and environmental conditions for a particular roadway section. Segregation, or lack of homogeneity (uniformity) in the in-place HMA constituents, has been identified by the asphalt industry as one of the most common problems associated with premature failure of HMA pavements. Segregation leads to accelerated pavement distresses, which translates in a reduction in the level of performance and service life of the pavement. Therefore, HMA Quality Assurance procedures usually include provisions to detect and quantify the presence of this problem in newly constructed pavements.

Traditionally, visually identified areas of non-uniform surface texture have been classified subjectively as segregated HMA and, therefore, as bad construction. To reduce the subjectivity involved in the quality assurance, an automatic measurement-based process is necessary. This paper presents a new methodology to detect and quantify HMA segregation using digital image analysis. The main objective of the non-contact system presented in this paper is the detection of segregated HMA areas and the identification of the locations of these areas along a road for HMA quality assurance purposes. The system uses relatively low cost off-the-shelf components for capturing images of pavement and innovative image processing and analysis software to automatically detect changes in the surface appearance due to segregation. The system can be used stand-alone or in combination with current laser-based systems for pavement quality assurance and has the potential to reduce the subjectivity in the HMA inspection process in addition to providing a permanent and periodic record of captured images, thus enhancing pavement quality assurance practices.

Introduction

Modern life has made our travel and freight patterns dependant on roads, which have to provide an appropriate travel surface under all-weather conditions. Hot Mix Asphalt (HMA) Pavements, are designed and constructed to withstand the expected traffic and environmental conditions for a particular road section. The goal is to build roads that are durable, safe, and comfortable to ride.

To achieve this, contractors have implemented detailed quality control procedures in their operations, and road owners, on the other hand, make sure that they deliver the specified quality through quality assurance procedures. As in any industrial process, present HMA road construction methods involve many potential variability sources that could result in

unacceptable end product quality. Variability is responsible for creating the lack of homogeneity (uniformity) in the in-place HMA constituents that is commonly referred to as segregation. Segregation is a serious problem because it leads to accelerated pavement distress (*1*). A segregated mix does not conform to the original job-mix formula in gradation and/or asphalt content. This nonconformity affects the density and air void content of the mix. The resulting non-uniform composition of aggregates, asphalt binder, and/or air voids results in uneven aggregate distribution in the surface.

Objective

The main objective of this paper is to discuss the application of imaging techniques in a system for the detection of segregated HMA areas and identification of these areas along a road for HMA quality assurance purposes. The proposed system uses relatively low cost off-the-shelf components for capturing images of new pavements and applies innovative image processing and analysis software for detecting and quantifying segregated areas.

HMA Segregation

Traditionally, visually identified areas of non-uniform surface texture have been classified subjectively as segregated HMA and, therefore, as bad construction. Many studies and agencies have tried to develop a reliable and independent method to define, detect, and quantify segregation, but none have eliminated the need for the visual inspections. The reduction in pavement performance life due to segregation can be up to 50% (*2, 3*). Considering that the industry produces more than 500 million tons of asphalt a year (*4*), there is an urgent need for better technologies to guarantee the effective return of all funds invested in road construction.

Measuring and Detecting Segregation

Most state DOT's base their segregation detection on visual identification made by construction inspectors, followed by density measurements to define boundaries of the segregated area. Upon its detection and delimitation, the usual action taken is to completely remove and replace the segregated area for the full depth of the course, sparking angry disputes with contractors. NCHRP Report 441 (*1*) investigated state-of-the-art technologies for detecting and measuring segregation. This project investigated the use of infrared thermography, ground penetration radar, thin-lift nuclear asphalt content/density, dynamic (laser-based) surface macrotexture, and seismic pavement analyzers. Infrared thermography was recommended for quality control (it can only be used during construction while the mix is hot). Macrotexture measurements were recommended for quality assurance purposes (*5*).

Regardless of all the technological methods now available, pavement inspectors still play a vital role in the supervision of all construction projects. Visual inspections have been the baseline against which any quantitative approach used to detect and measure segregation is tested and compared (*1, 6*).

Because there are different types of segregation (namely, gradation, temperature and asphalt-aggregate segregation) the lack of uniformity found on new pavement projects exhibit different characteristics inherent to its specific cause. However, non-uniform areas usually appear as spots where it is apparent that the aggregate particles have either become dislodged or are intricately closer than the ones in neighboring places. The human eye can detect these non-uniformities by observing the texture of the mix in these areas and comparing it with a normal area. Previous research has shown that visual observations by experienced inspectors could adequately identify non-uniform surface texture indicators of segregation (*7*).

Computer Vision

The field of computer vision is concerned with the automatic interpretation of pictorial information. Automatic is, however, an overstatement. The problem of using a two dimensional visual image to infer information from three-dimensional objects is very complicated. Human vision can overcome many of the difficulties with a blink of the eyes. Human vision has the advantage of using two eyes which allow it to detect depth (stereo vision), analyze shapes, and recognize objects automatically. Machine vision systems, on the other hand, don't have this capability but they offer other advantages. For example, machine vision systems can help eliminate some problems associated with the subjectivity of human visual inspections, fatigue, inconsistencies, repeatability, etc. Although there are challenges associated with the capture of digital images that will be explained later, the application of the approach proposed herein provides a more objective quality assurance methodology.

Typically in computer vision systems, an image-processing phase is followed by a higher-level image analysis phase. Image processing commonly modifies the pictorial appearance of the image and deals with noise removal, edge detections, or contrast improvements of an image. Image analysis provides information concerning the content of the image. Common applications of this second phase deal with problems involving object recognition or distance estimation. Both of these phases were used in this research.

Digital imaging has been utilized in various applications that include pavement crack detection, aggregate shape properties, aggregate shape characterization, and pavement distress surveys. For example, Clemeña and Howe (*8*) studied the feasibility of using digital image analysis to evaluate pavement distresses. Another project used image analysis to quantifying the surface condition of seal coated pavements, utilizing the two-dimensional Fourier Transform to compare images of the different pavements analyzed (*9*).

System Framework

The system development included an extensive testing plan, which combined preliminary field measurements, hardware and software development, and field and laboratory calibrations for developing a novel method to try to detect and quantify segregation. The system (Figure 1) consists of one camera that acquires images of the pavement at a user-selected interval. The camera is controlled by a computer connected to a distance measuring device (DMI). The computer controls the acquisition rate, stores the images, and analyzes them. The digital camera has a resolution of 640 x 480 pixels (picture elements) and can record up to 80 frames per second. The system can record images every foot at highway speeds of up to 55 mph. The volume of images used can be adjusted by increasing/decreasing the data collection frequency

Figure 1 System Back and Lateral View

Image Processing and Analysis

Experienced inspectors effectively detect non-uniformities by observing the texture of the mix in these areas and comparing it with a normal "uniform" area. However, in difficult situations, human subjectivity, eye fatigue, and environmental factors can alter the observations. In these cases, an automatic process is preferred.

The procedure presented in this paper uses digital grayscale images, which consist of an array of pixels, with spatial coordinates in the two dimensions facing the camera and an associated brightness level at each pixel location. The array forms a matrix that has an (x, y) configuration with a relative brightness level in each of the cells of the matrix. Most brightness levels are recorded as integer powers of 2, usually using the byte (or 2^8) = 256 shade scale, with 0 = black, 255 = white, and 254 shades of gray in between.

Using a histogram of pixel values, it is often possible to perform fast statistical analysis of some image properties. However, one limitation of standard histogram analysis is that spatial information is ignored. An alternative to this is the Gray Level Co-Occurrence Matrix (GLCM), which is commonly used for visual texture analysis (*10*). This two-dimensional matrix captures statistical information, similar to that of a histogram, based on the relative locations of pixel values in the image. The distribution of values within the GLCM depends on the level of coarseness of the visual texture.

Human perception of texture is qualitative (coarse, fine, regular, irregular, granular, etc). The relationship between texture and image intensity is complex, but a quantitative analysis can be done by identifying texture primitives (texels) and describing the spatial relationships between them. Pixel values are independent of other "First-order statistic" like marginal probability:

$$P_i = \Pr(I(r,c) = i) \tag{1}$$

Images having second order statistics have the following joint probability

$$P_{ij} = \Pr(I(r,c) = i \cap I(r+\Delta r, c+\Delta c) = j) \tag{2}$$

The Gray-level Co-occurrence Matrix (GLCM) gives intensity information in pixel "pairs", with a displacement vector d = (Δr, Δc) where the GLCM is equal to:

$$C(i,j) = \frac{1}{M}\left[a_{i,j}\right] \tag{3}$$

where,
i = intensity value of the pixel in row "r" and column "c"
j = intensity value of the pixel in row "r+Δr" and column "c+Δc"
M= # different pixel pairs and
$a_{i,j}$ is the number of times I(r,c)=i and I(r+Δr, c+Δc)=j.

Coarse textures have large diagonal values, fine textures values spread out from the diagonal. Using the GLCM results, the following parameters have been derived by machine vision researchers to identify different types of texture parameters depicted in the images:

$$\text{Contrast} = \sum_{ij} P_{ij}^2 \qquad \text{Energy} = \sum_{ij} |i-j|^2 P_{ij}$$

$$\text{Correlation} = \sum_{ij} \frac{(i-\mu)(j-\mu)P_{ij}}{\sigma^2} \qquad \text{Homogeneity} = \sum_{ij} \frac{P_{ij}}{1+|i-j|} \tag{4}$$

where,
μ = gray level mean
σ^2 = gray level variance

Field Testing

Testing of the system has been conducted at the Virginia Smart Road, which is a controlled traffic facility with different surfaces. These surfaces types included five different SuperPave mixtures, a 12.5 mm stone mastic asphalt (SMA), a 12.5 mm open-graded friction course (OGFC), and texturized continuously reinforced concrete (Figure 2). The HMA mix types are presented in Table 1. The results of the analysis for all the surfaces are summarized in Figure 3.

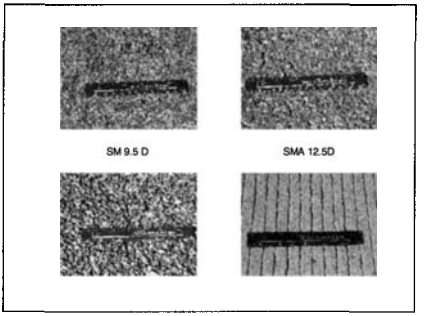


Figure 2 Smart Road Surface Mixes

Table 1 Smart Road Mix Types

Section	Mix	Binder
A	SM-12.5D	PG 70-22
B	SM-9.5D	PG 70-22
C	SM-9.5E	PG 76-22
D	SM-9.5A	PG 64-22
E -H	SM-9.5D	PG 70-22
I	SM-9.5A(h)	PG 64-22
J	SM-9.5D	PG 70-22
K	OGFC	PG 76-22
L	SMA-12.5D	PG 70-22

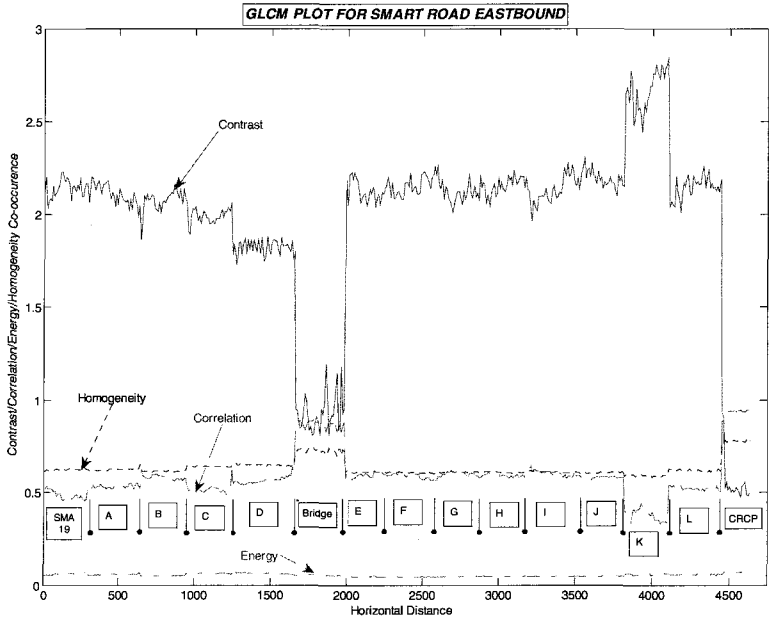


Figure 3. Analysis of the Smart Road Surface Textures

Figure 3 shows the changes on the four texture parameters along the eastbound lane of the Virginia Smart Road, clearly showing the differences in texture of the different mixes. Future testing will include newly paved projects in the up-coming paving season to test the methodology for intended quality assurance purposes. A visual survey will be performed to compare the results and other field and laboratory test results will be used to verify the potentially segregated areas identified by the prototype vehicle. In addition, during this stage, the software will also be revised to optimize the user interfaces as suggested by the inspectors of the projects. The end-user application will have user-friendly graphical interfaces to hide complicated algorithms and processes from users.

Summary and Conclusions

This paper discussed the development of a non-contact system for the detection of segregated HMA areas and identification of these areas along a road for HMA quality assurance purposes. The system uses relatively low cost off-the-shelf components for capturing images of new pavements and applies innovative image processing and analysis software for detecting and quantifying segregated areas. The system also has the advantage of providing a permanent record of captured images that could assist in resolving disputes between pavement owners and contractors. The proposed quality assurance method is expected to help automate the construction inspection process, reduce the subjectivity associated with an inspector performing this process, and enhance quality-assurance procedures.

References

1) Stroup-Gardiner, M. and E. R. Brown. *NCHRP Report 441, Segregation of Hot-Mix Asphalt Pavements*. TRB, National Research Council. Washington, DC, 2000.
2) Brock, J.D. and Jakob, H. *Temperature Segregation/Temperature Differential Damage.* Astec Industries, Inc., Technical Paper T-134. Chattanooga, TN, 1996.
3) Wilson, R.L. *Commentary on Special Provision Concerning HMAC Segregation.* Texas: Texas Department of Transportation, 1999.
4) Acott, Mike. *How the U.S. Asphalt Industry is Affected by HS&E Regulations, and How it is Responding to SHRP Developments.* 2nd Eurasphalt & Eurobitume Congress, Barcelona, 2000.
5) McGhee, K.K., Flintsch, G.W., de León, E., *"Using High-Speed Texture Measurement to Improve the Uniformity of Hot-Mix Asphalt"*, VTRC 03-R12, Virginia Transportation Research Council and Federal Highway Administration, 2003.
6) Khedaywi, T. S. and T. D. White. "Effect of Segregation on Fatigue Performance of Asphalt Paving Mixtures". In *Transportation Research Record 1543*, TRB, National Research Council. Washington, DC, 1996, pp. 63-70.
7) Cross, S. A. and E. R. Brown. Effect of Segregation on Performance of Hot-Mix Asphalt. In *Transportation Research Record* 1417, TRB, National Research Council. Washington, DC, 1993, pp. 117-126.
8) Clemeña, G. and Howe, R. *An Assessment of the Feasibility of Developing and Implementing an Automated Pavement Distress Survey System Incorporating Digital Image Processing.* Virginia Transportation Research Council Research, 1998.
9) Gransberg, D.D., Karaca, I., and Burkett, W.R. Quantifying Seal Coat Surface Condition Using Digital Image Processing Based on Information Theory. In *International Journal of Pavement Engineering*. Taylor and Francis, 2003, Vol. 3, No. 4/25, pp. 197-205.
10) Haralick, Robert, and Shapiro, Linda, *Computer and Robot Vision, Volume 1.* Addison-Wesley, 1992.

Comparative Analysis of Using Laboratory Testing and Nondestructive Testing to Obtain Subgrade Resilient Moduli

Jianming Ling[1] , Runhua Guo[2] and Jie Yuan[3]

[1]Department of Road and Airport Engineering, Tongji University, Shanghai 200092, China; PH (86-21) 6598-1431; Email: jmling01@yahoo.com.cn
[2]Department of Civil Engineering, The University of Texas at Austin, Texas 78703, USA; PH (512) 228-9108; Email: guog75@mail.utexas.edu
[3]Department of Highway and Airport Engineering, Tongji University, Shanghai 200092, China; PH (86-21) 6598-1431; Email: yuanjie@vip.sina.com

Abstract

Pavement subgrade characterization in terms of resilient modulus (M_R) is essential for pavement design. This paper presented a new methodology to monitor and evaluate seasonal variation of subgrade resilient modulus. First, a kind of typical cohesive subgrade soil was tested to determine the relationship between tension (Ts) and saturation degree (Sr) utilizing soil tensiometer in laboratory, as well as the relationship between resilient modulus and saturation degree based on repeated triaxial test. The majority of the analyses had R2 values above 0.98, which indicated that a good correlation exists between tension, saturation degree and resilient modulus. Thus it is allowed to predict the resilient modulus of cohesive soil from the tension. In addition, a road section was instrumented, coupled with Falling Weight Deflectometer (FWD) deflection tests monthly over a period of a year. Further, AASHTO equation method and MODULUS program were both used to back-calculate the subgrade modulus based on the FWD deflection data. Finally, a corrective coefficient was suggested for pavement design when using the MODULUS program to back-calculate the sub grade modulus. A comparative analysis was conducted throughout the above process.

Introduction

Background. Pavement subgrade soil characterization in terms of resilient modulus (M_R) is essential for pavement design. There are usually two main methods to obtain this value: one is through the in-door strain-stress testing on the soil, according to accepted standards or manuals, to obtain it directly (ASTM, 1997) or test dependent parameters which are directly related with the modulus; the other one is so-called Nondestructive Testing (NDT). Ali and Khosla (1987) employed VESYS, ELMOD, MODCOMP2, and OAF models to determine the modulus of the pavement structural layers and noted that the first two methods were more appropriate for back-calculation. It was also established by the Strategic Highway Research Program that, MODCOMP3, MODULUS, and WESDEF models were more appropriate tools for back-calculation analysis, compared with other three programs ELCON ILLI-BACK, and ISSEM4 (PCS/Law Engineering, 1993).

For the reason of the high variation property of subgrade soil modulus due to environmental conditions, especially the moisture, and the short of straightforward measuring technology, it is difficult to accurately characterize the modulus of cohesive soil subgrade with changing water table (Hossain et al., 2000), especially in Shanghai soft clay area.

Objective. The objective of this study was to demonstrate and validate a new methodology to determine subgrade resilient modulus, and then compare the testing results with other back-calculation data. Finally the authors conducted a comparative analysis of using laboratory testing and nondestructive testing to obtain subgrade resilient moduli.

Methodology and Scope. First, a kind of typical cohesive subgrade soil was tested to determine the relationship between tension (Ts) and saturation degree (Sr) utilizing soil tensiometer in laboratory, as well as the relationship between resilient modulus and saturation degree based on repeated triaxial test. Thus it is allowed to predict the resilient modulus of cohesive soil from the tension. In addition, a road section was instrumented, coupled with FWD deflection tests monthly over a period of a year. Further, AASHTO equation method and MODULUS program were both used to back-calculate the subgrade modulus based on the FWD deflection data. Finally, a corrective coefficient was suggested for pavement design when using the MODULUS program to back-calculate the sub grade modulus. The comparative analysis will be conducted throughout the above process.

Road Section Selected and Field Instrumentation

Section Characteristics.
(1) A one-km long road section of Shanghai Outer Ring Highway (Phase 2-1), extending from k+3.2 to k+ 4.2, was selected for this study.
(2) 8-lane divided asphalt pavement with 6m wide median strip planting, 2% transverse grade and 0% longitudinal grade.
(3) This section was constructed in 2000 and the pavement structure comprised 150mm asphalt concrete surface, 400mm lime-flyash concrete (lime-flyash stabilized crushed stone) base, and 300mm 10% lime stabilized soil subbase on a silty clay subgrade soil, within a fill area varies from 0.89m to 1.21m in height.
(4) Filled subgrade soil classification (in MUCS): Silty clay. Wop=19.1%, γmax=17.30kN/m3. The compaction degree required is greater than 98%.

Field Instrumentation and FWD deflection Testing. The section selected for the study were instrumented with two kinds of monitoring sensors, coupled with FWD deflection testing points, shown in Figure 1 and Figure 2. In which two water table gauges were installed within median strip planting and outer shoulder; 6 soil tensiometers were calibrated in laboratory and installed at different depths; 40 FWD testing points were located and marked in the wheel paths of lane 2 and lane 3.

Laboratory testing and data analysis

Soil Samples Preparation and Routine Tests.
(1) All the soils used in the laboratory testing program were retrieved from the instrumented site and the tensiometer used in the laboratory tests was the same type as in the field tests.
(2) Routine soil tests were performed on the soil, including grain size distribution, Atterberg limits, soil classification, standard compaction tests, and soak test. Summarized results:
Silty clay: Wop=19.1%, γmax=17.30kN/m3
Saturation capacity at standard compacted state Wmax=26.4%

(3) Soil samples for tension test and resilient modulus test (triaxial repeated loading test) were compacted in optimum moisture content.

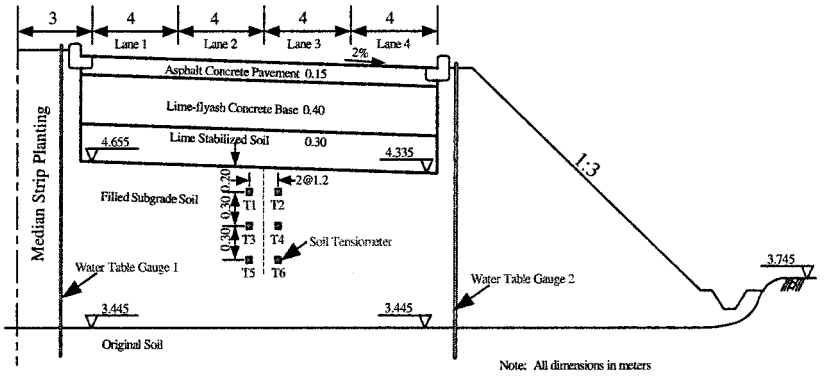


Figure 1. Road Cross Section and Monitoring Sensors Locations

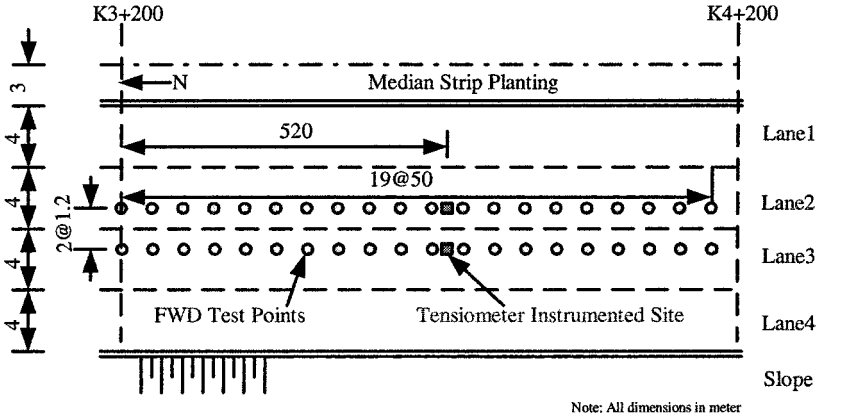


Figure 2. Plan View of Locations of FWD Test and Monitoring Sensors

Subgrade Soil Tension Test. To measure the soil tension stresses Ts in different saturation degree Sr with advanced tensiometers, and subsequently to determine the Ts-Sr relations as shown in figure 3 (Sr vs. Ts). These were performed to enable the inference of saturation degree from measurements of soil tension values.

Resilient Modulus Test. Triaxial repeated loading tests on the subgrade soil samples were performed to obtain its resilient characteristics. In this testing program, three main factors were under consideration: Moisture contents, Confining stress level and Deviator stress. A 40kPa of deviator stress and 16kPa of confining stress level was selected for the test to simulate the real situation.

The influence of soil saturation degree on the resilient modulus was depicted in figure 3 (Sr vs. Er). This figure showed that good correlations were obtained, which indicated that Sr had a very pronounced effect on the resilient modulus. For the range of the Sr investigated, i.e. 63.1% to 95.2%, the Er decreases by 9 orders of magnitude (from 78MPa to less than 8MPa), especially in the range near optimum moisture content (Sr=72.3%).

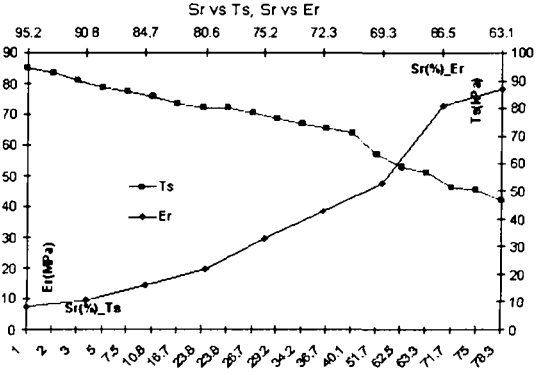


Figure 3. Laboratory testing of E_r, Ts vs. Sr

Correlation between Er, Ts and Sr. Significant linear relationships between Er and Sr, Ts and Sr were presented in figure 3. Moreover, the majority of the analyses had R^2 values above 0.93, which indicated that a good correlation existing between tension, saturation degree and resilient modulus, so did Er and Ts. Based on the correlation between Er and Sr, Ts and Sr resulted from the test, the correlation between Er and Ts for cohesive subgrade soils could be obtained as shown in figure 3, applying linear regression techniques, a predictive model for monitoring and evaluating subgrade resilient modulus from measured soil tensions was proposed as follows:

$$Er = 8.0e0.0437Ts,\ R^2 = 0.9846 \tag{1}$$

Field Testing and Nondestructive Testing (NDT)

FWD Test and Back-calculation of Modulus. The FWD tests of the field pavement were monthly conducted with a PHOENIX FWD (100kN load, 30cm diameter plate) at the same time as soil tension measurement. The summarized statistics of FWD deflection test results were listed in Table 1. Two methods were selected for back-calculation of resilient modulus of subgrade:

(1) Equation recommended by AASHTO Guide for Design of Pavement Structures (AASHTO, 1993)

$$E_{rA} = c\frac{0.24P}{l_r r} \tag{2}$$

In consideration of the greater stiffness of the pavement, here takes the corrective coefficient c=0.33, lr=D150 when r=150cm.

(2) MODULUS Program (Uzan, and Lytton, 1989) without corrective coefficient

The resilient moduli of subgrade back-calculated with two methods above were listed in Table 2, coupled with those derived from soil tension data. In viewing table, it was apparent that an unfavorable season really existed within a period of one year. The FWD deflections of the pavement in this season were much larger than that in other seasons. It was corresponding to the water table, the soil tensions and the derived resilient modulus of subgrade. It was also worth noting that the shapes and slopes of the FWD deflection basins were different in different seasons. As a result, not only the subgrade modulus but the overall stiffness of the road varied seasonally.

Table 1. Statistics of FWD Test Results (P=100kN, d=30cm)

Test Date	Deflection at Various Sensor Points (um)						Cv (%) of	
	D0	D30	D60	D90	D120	D150	D0	D150
06/05/00	318	179	126	99	80	67	17	12
07/06/00	323	207	164	137	116	99	11	8
08/05/00	490	348	296	258	225	197	31	17
09/07/00	666	510	451	403	360	321	19	11
10/08/00	745	610	551	501	454	411	27	10
11/05/00	526	411	371	323	287	256	10	7
12/05/00	447	327	279	243	212	186	11	10
01/07/01	421	303	260	227	198	173	14	10
02/04/01	473	365	322	286	254	225	9	12
03/08/01	410	310	272	239	211	186	12	9
04/05/01	414	305	263	231	201	177	24	11
05/08/01	461	338	293	255	223	195	13	15
06/05/01	440	315	271	235	204	175	16	11

Table 2. Resilient Moduli Back-calculated and Derived with Different Methods

Date	Resilient Moduli Back-calculated		Resilient Moduli Derived
	ErA(MPa)	ErM(MPa)	ErT(MPa)
2000/6/5	79	285	76
2000/7/6	53	188	53
2000/8/5	27	92	18
2000/9/7	16	54	8
2000/10/8	13	40	10
2000/11/5	21	68	21
2000/12/5	28	97	37
2001/1/7	31	105	31
2001/2/4	23	78	21
2001/3/8	28	96	28
2001/4/5	30	105	25
2001/5/8	27	93	35
2001/6/5	30	102	33

Note: ErA, ErM, ErT respectively for AASHTO equation, MODULUS program and Soil Tension deriving method.

Analysis and Comparison

From Table 2, it was found that ErA approximated to ErT, while ErM was much greater than ErT. The regressive relation of ErM and ErT was displayed as in the following models.

$$ErT = 1.0077ErA - 1.0093, R^2 = 0.9143 \quad (3)$$
$$ErT = 0.2722ErM + 1.0893, R^2 = 0.9119 \quad (4)$$

The majority of the above analyses had R values above 0.9, which indicated that good correlations exist between ErT, ErA and ErM.

These equations maybe provided a guideline for taking value of back-calculated resilient moduli of cohesive subgrade soils in pavement design.

Conclusions

The following conclusions seem warranted:

(1) Based on the laboratory test results, relation between Er and Ts was rationally established, and then a new procedure to monitor the resilient modulus of subgrade soil using soil tensiometer was proposed and proved to be reliable.

(2) Different modulus back-calculation methods were checked and compared with the testing results from soil tension method. It suggests that resilient modulus of subgrade soil determined from MODULUS program should be adjusted by a shift factor (in this case 0.27). While results from AASHTO equation tend to be applicable.

References

AASHTO. (1993). "AASHTO Guide for Design of Pavement Structure." Chapter 3: Guides for Field Data Collection. AASHTO, Washington, D.C..

Ali, N. A., and Khosla, N. P. (1987). "Determination of layer moduli using a falling weight deflectometer." Transportation Research Record 1117, Transportation Research Board, Washington, D.C...

ASTM. (1997). "Annual Book of Standards." ASTM, Section 4, Construction, Vol. 4.08, West Conshohochen, PA.

Hossain, M., Romanoschi, S. and Gisi, A. J. (2000). "Seasonal and Spatial Variation of Subgrade Response. Pavement Subgrade. Unbound Materials, and Nondestructive Testing." Proceedings of Sessions of Geo-Denver 2000, ASCE, Denver, pp. 150-166.

PCS/Law Engineering. (1993). "SHRP's layer moduli backcalculation procedure." Rep. SHRP-P-655, Strategic Highway Research Program, Transportation Research Board, Washington, D.C...

Uzan, J., and Lytton, R. L. (1989). "Experiment design approach to nondestructive testing of pavements." Journal of . Transportation. Engineering., 115~5!, 505–520.

Characterization of Hamburg Wheel Tracking Device Testing Results

Runhua Guo[1] and Jorge Prozzi[2]

[1] Department of Civil Engineering, The University of Texas at Austin, Texas 78703, USA; PH (512)-228-9108; Email: guog75@mail.utexas.edu
[2] Department of Civil Engineering, The University of Texas at Austin, Texas 78703, USA; PH (512)-232-3488; Email: prozzi@mail.utexas.edu

Abstract

The Hamburg Wheel Tracking Device was developed as specification requirement for some of the heavy-duty road to measure rutting (potential) and stripping (moisture susceptibility). TxDOT adopted cylindrical specimens for testing with the HWTD (Hamburg Wheel Tracking Device) manufactured by PMW, Inc. Traditionally, the rut depth has been represented by the largest impression point which occurs randomly along the 11 points. The maximum impression is randomly located and more likely to be in the middle part of specimen. It is difficult to characterize with a single theoretical distribution such as a normal or other such function that defines one central tendency.

This paper detailed the spatial distributions of Hamburg Wheel Tracking Device Test Results and demonstrated its predictive capability using statistics tools. In addition, the author conducted statistical analysis of all the test points in four patterns, and thus a comparison between all patterns were carried on. Finally, a more reasonable representative pattern is proposed. Central to the approach developed in this research was utilizing a statistics tool to better characterize the deformation.

It is certainly more straightforward to represent them with maximum deformation rather than other averaged points. But in case where small amounts of observed data are available, large variance could prove prohibitive to reliable analysis, which will finally result in great uncertainty in predicting performance. Average of the next four maximum deformations presents a relative smaller variance compared to other method. In case observed data are limited, it will provide a relatively better choice.

Introduction

Background. There have been past efforts to predict HMA (hot mix asphalt) rutting potential through accelerated laboratory test, and the return to these efforts are relatively inexpensive but useful quality control/quality assurance (QC/QA) of great benefit. Loaded wheel testers (LWT) are the most common type of laboratory test tools for this purpose, in fact there are a number of types of LWT equipment in use throughout the world such as the Georgia Loaded Wheel Tester (GLWT), Asphalt Pavement Analyzer (APA), Hamburg Wheel Tracking Device (HWTD), the French Laboratoire Central des Ponts et Chaussées (LCPC) Wheel Tracker, Purdue University Laboratory Wheel Tracking Device (PURWheel),

one-third scale Model Mobile Load Simulator (MMLS3), and the Superfos Construction Rut Tester (SCRC). Most of them are currently being used in the United States (Cooley).

The Hamburg Wheel Tracking Device, shown in Figure 1 (Gogula), was originally developed as specification requirement for some of the heavy-duty road to measure rutting (potential) and stripping (moisture susceptibility). The HWTD test slab is usually 260 mm wide, 320 mm long and 40 mm deep with an approximate mass of 7.6 kg and 7 ± 1 % air voids. Testing is conducted underwater and the temperature is controlled at approximately 50°C (122°F) ranging from 25°C to 70°C (77°F to 158°F). This temperature was selected based on HMA's severing temperature submits to rutting. The steel wheel is 47 mm wide loaded by 705 N (158 lb) and travel at a speed of approximately 1.2 km/h (53±2 passes /minute) until the slab is loaded for 20,000 cycles or 20 mm of impression, whichever occurs first. Typically, a pair of two samples is tested simultaneously with two steel wheels moving concurrently. It usually will take more than 6 hours to perform the 20,000 passes. Impression is measured at 11 different points (marked as RTD1 through RTD11 in order) along the length at the center of each pair of specimens by a Linear Variable Differential Transformer (LVDT). Results from the testing include rut depth, creep slope, stripping slope, and stripping inflection point, which are theoretically and statistically correlated with potential rutting and moisture susceptibility (Cooley).

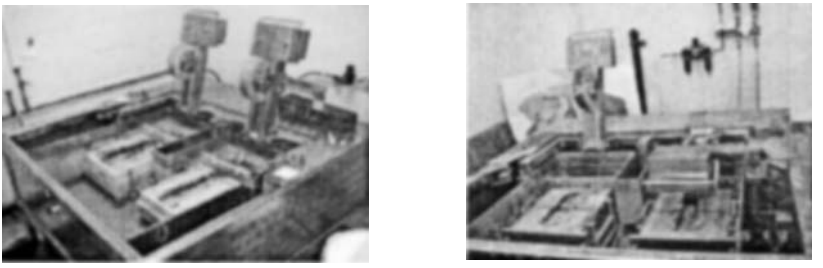


Figure 1 Hamburg Wheel Tracking Device: (a) slab; (b) cylinder

As Superpave Gyratory Compactor gains greater acceptance and use, there is a need to accurately characterize the shape expected to be applied on specimen. TxDOT adopted cylindrical specimens for testing with the HWTD manufactured by PMW, Inc. (Figure 1b). As described in the manual, the purpose of TxDOT Hamburg Wheel-tracking Test is as follows: "Use this test method to determine the premature failure susceptibility of bituminous mixtures due to weakness in the aggregate structure, inadequate binder stiffness, or moisture damage and other factors including inadequate adhesion between the asphalt binder and aggregate. This test method measures the rut depth and number of passes to failure." The cylinders are 150 to 300 mm in diameter and 62 mm deep, also compacted to 7±1 percent air voids like slabs. To provide required wheel track length, the two specimens are attached together by cutting edges. The configuration of the test device is almost the same as slabs; furthermore, four cylindrical samples are tested at the same time (Gogula).

Traditionally, the rut depth has been represented by the largest impression point which occurs randomly along the 11 points. It must be noted that typically this point should be located in the middle part of the specimen, but the real results are quite complex and historically have not been well-characterized by theoretical statistical distributions. This

largest approach impression has been used by numerous agencies; it is under developing as part of our new research and is currently being updated by using other points.

Objective. The objective of this study is to demonstrate and validate the applicability and effectiveness of using other representative test points for better predicting potential rutting and moisture susceptibility of HMA, instead of using the largest deformation points.

This paper detailed the spatial distributions of Hamburg Wheel Tracking Device Test Results and demonstrated its predictive capability using statistics tools. In addition, the author conducted statistical analysis of all the test points in four patterns, and thus a comparison between all patterns were carried on. Finally, a more reasonable representative pattern is proposed. Central to the approach developed in this research was utilizing a statistics tool to better characterize the deformation.

Spatial Distributions of Hamburg Wheel Tracking Device Test Results

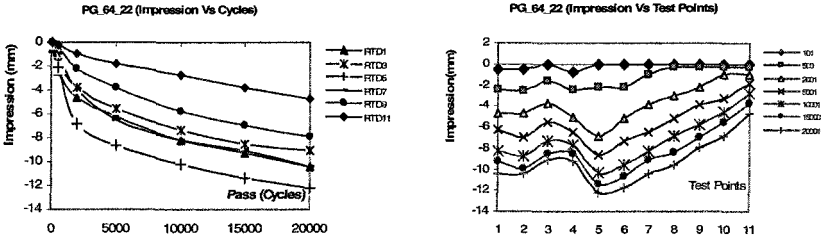
Gogula et al. presented the effect of PG binder grade on performance (TxDOT) and considering that, the author classified all the test specimens into three categories according to the PG binder grade: PG-76, PG-70 and PG-64. Table 1 lists the Aggregates and binders used for HWTD and the number of specimens with confirmed data sources. Four main kinds of aggregates were used to prepare the specimens and altogether 169 sample results were collected for further analysis as shown in the table.

TABLE 1. Aggregates and binders used for HWT

	Gravel	Limestone	Igneous	Sandstone	Total
PG 64	3	83/11*	0	0	86
PG 70	0	21/4*	2/1*	2	25
PG 76	3/1*	38/3*	13	4/2*	58
Total	6	142	15	6	169

Impression throughout the wheel pass. Figure 2 illustrates sample configurations of spatial distributions of TxDOT Hamburg Wheel Tracking Device test results. In the figure, (a) represent the impression of each test points throughout the maximum passes, accordingly, which illustrating how the deformation progresses.

The curves shown in Figure 2(b) represent the bowl-shape deformation of the specimens at selected pass (cycle), which illustrate what happens to the specimen surface after the wheel pass. Different series stand for different test points or passes.



(a) (b)

Figure 2 Spatial Distributions of Impression

Maximum Impression distribution (Maximum). Figure 3 illustrates the distribution of maximum impression location among the 11 test points. In the figure, different series stands for different PG binder. It can be observed in this figure that in most cases the distribution is not the same. A normal tail fit the lower tail fairly well, but does not characterize the entire distribution if we put all the binders together. Moreover, if considered separately the distribution of different binder is significant different from each other. Consequently, it is difficult to characterize with a single theoretical distribution such as a normal or other such function that defines one central tendency.

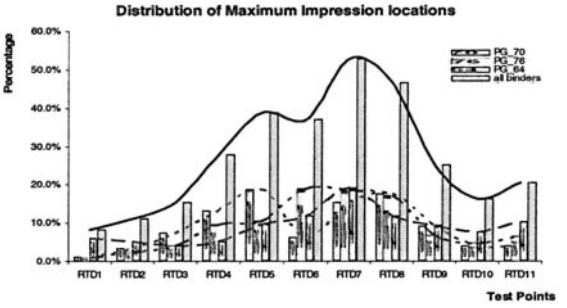


Figure 3 Distributions of Maximum Impression Locations

The maximum impression is randomly located and more likely to be in the middle part.

Statistical Analysis of Hamburg Wheel Tracking Device Test Results

Proposed method. As described previously, the maximum impression is randomly located and this uncertainty will be exaggerated when used in predicting potential rutting and stripping (moisture susceptibility). From a practical and theoretical standpoint, it would be useful to develop substitution of the maximum deformation point. In other words, finding a more reliable point would be advantageous.

On the basis of a numerical analysis of the experimentally determined results, taking the average of test points will greatly reduce the variance; hence increase the reliability of predicting. In this study, three other patterns excluding the maximum method were proposed as follows: average of the RTD2, RTD3 and RTD4 (RTD234); average of the RTD8, RTD9 and RTD10 (RTD890); and average of the next four maximum (RANK2345) excluding RANK1. It was demonstrated that the proposed methods could analytically reduce variance, hence offering a more reliable point to predict.

t-Test: Paired Two Sample for Means. At this point it must be noted that, when using different method, the difference between them should be significant. In other words, proposed method should be significantly different from the original one.

The results from the t-Test of Paired Two Sample for Means are described in Table 2 together with the analysis of the t Critical one-tail and two-tail value. Ideally, the hypothesis of significant difference between RANK2345 and RANK1 was validated (Table 2a), and as well as between RTD890 and RANK2345 (Table 2b), no matter you put all binders together

or consider them separately. However, the difference between RTD234 and RTD890 was probably insignificant as of PG 70 (Table 2c). This is reasonable, since the test device, specimen and applied load are all symmetrical. Moreover, the difference between RTD234 and RTD890 as of PG 64 and PG 76 is significant, which illuminates the uncertainty of deformation for all the eleven test points.

Table 2 t-Test: Difference between Different Representative Points

Representative Points	t Stat	P(T<=t) one-tail	t Critical one-tail
RANK2345 & RANK1	10.12	2.31E-23	1.65
RANK2345 & RTD890	33.78	1.28E-171	1.65
RTD234 & RTD890	7.16	7.45E-13	1.65

Average, Standard Deviation, and Coefficient of Variance (CV). Numerically, the variance and the mean are calculated and analyzed. Two main facts have effects on the results: PG grade and number of passes. This is reasonable because these two affect the specimen performance directly.

Average . The result shows a difference in average depending on the PG grade: for the same wheel pass number, averages of all the test points with PG 64 present the largest result and the smallest for PG 76. However, for the same PG grade, the difference between RANK2345, RTD890 and RTD234 is not so significant before 10,000 wheel passes. But after that, RANK2345 changes to be the smallest.

Standard Deviation (SD). The Standard Deviation of all the proposed points according to the PG grade and wheel pass number. It is apparent that in most scenarios RANK1 presents the highest variance, which agrees with the previous study, while RANK2345 presents to be with smallest variance in most cases. Similarly like the average, points with PG 64 present the largest value compared with other PG grade. Moreover, the Standard Deviation of all the points presents an increase as wheel pass number increases.

Coefficient of Variance (CV). Most of the Coefficient of Variance of RANK1 were slightly smaller than those of other points. This result differs considerably from what we expected. But for all the other points, the difference is not significant.

Discussions

As described previously, none of the proposed method can best satisfy all the criteria at the same time. However, there are two important differences. First, all the proposed method are not assumed to be the same as others. In the second term, the performances are different. In terms of presenting performance, it is certainly more straightforward to represent them with RANK1 rather than other averaged points. But in case where small amounts of observed data are available, large variance could prove prohibitive to reliable analysis, which will finally result in great uncertainty in predicting performance.

RANK2345 presents a relative smaller variance compared to other points. In case observed data are limited, it will provide a relatively better choice.

Performance of RTD234 and RTD890 not only present highest variance in most cases, but also the highest Coefficient of Variance, which proves the randomly properties of

the test points. In other words, it is no good to treat them as representative test points in any case.

Since there could be numerous combinations of the 11 test points, the selection of proposed methods is somewhat subjective. The present study has been limited to special points. Even though this makes the study less general, the result gives an idea. Results of this study lend credence to and confidence in the use of other points.

Aggregate is also closely related with the rutting performance of HMA, which means that the three categories can be expand to more to achieve accurate analysis. The analysis of these three included PG grade cannot stand for the other PG grade.

CONCLUSION

Based on the procedure and the data presented, the following conclusions may be made:

The maximum impression is randomly located and more likely to be in the middle part of specimen. It is difficult to characterize with a single theoretical distribution such as a normal or other such function that defines one central tendency.

It is more straightforward to represent them with maximum deformation rather than other averaged points. But in case where small amounts of observed data are available, large variance could prove prohibitive to reliable analysis, which will finally result in great uncertainty in predicting performance.

Average of the next four maximum deformations presents a relative smaller variance compared to other method. In case observed data are limited, it will provide a relatively better choice.

REFERENCES

Cooley, LA Jr; Kandhal, P.s., Buchanan, M.S. (2000). "Loaded Wheel Testers in the United States: State of the Practice." Transportation research. E-circular. Number E-C016.

Gogula, A., Hossain, M., Boyer, J., and Romanoschi, S.. (2003). "Effect of PG Binder Grade and Source on Performance of Superpave Mixtures under Hamburg Wheel tester." Proceedings of the 2003 Mid-Continent Transportation Research Symposium, Ames, Iowa.

TxDOT. (2004). "TxDOT Manual System, Hamburg Wheel-tracking Test." Chapter 38 - Tex-242-F, 200-F, Bituminous Test Procedures Manual.

Analysis of Simulated Variation in Dielectric Properties of Reinforced Concrete Systems

K. Belli[1], S. Wadia-Fascetti[2], and C. Rappaport[3]

[1]PhD Candidate, Dept. of Mech. & Ind. Eng., Northeastern Univ., 360 Huntington Avenue, Boston, MA 02115; PH (781) 789-8498; kbelli@coe.neu.edu

[2]Associate Professor, Dept. Civil & Envir. Eng., Northeastern Univ., 360 Huntington Ave., Boston, MA 02115; (617) 373-4248; (617) 373-4419 (fax); swf@coe.neu.edu

[3]Professor and CenSSIS Assoc. Dir., Dept. Elect. & Comp. Eng., Northeastern Univ., 302 Stearns Bldg., Boston, MA 02115; (617) 373-2043; rappaport@ece.neu.edu

Abstract

Health monitoring systems that monitor location, characterization, and quantification of damage can benefit from model-based assessment to relate better the measured response to physical indicators of damage. Subsurface sensing technologies, one of many sensing tools available for health monitoring systems, produce wave-based responses that traverse the subsurface and can be converted to a subsurface image for further interpretation. An integrated modeling environment is under development by the authors. The authors use this modeling environment to simulate responses from a ground penetrating radar investigation of a reinforced concrete bridge deck. The focus of this paper is to analyze the differences in the simulated responses due to variability in the dielectric properties of concrete and the presence of air voids. Calculated dielectric properties can vary 10% - 15%. Quantification of the dielectric variability is necessary to develop robust solutions of the inverse problem that will characterize subsurface conditions.

Introduction

Ground penetrating radar (GPR) has become a popular and effective technology for subsurface assessment of reinforced concrete bridge decks. The technology is based on the transmission of an electromagnetic wave into the medium of interest, here a reinforced concrete bridge deck. The electromagnetic wave refracts and reflects as it comes into contact with different materials. The signal that returns to the surface is rich with information about the subsurface. GPR provides information about changes in the electromagnetic properties of materials, which may include layer boundaries, rebar, voids and areas of contaminated concrete. After having collected the data, it needs to be interpreted to identify and ideally quantify anomalies in the

subsurface. A major step toward being able to characterize the bridge deck is determination of the electromagnetic properties of the materials.

Electromagnetic Property & Layer Thickness Calculation

The intrinsic impedance (η) of a media is a function of the permittivity (ε), conductivity (σ) and permeability (μ) of the material and of the angular frequency (ω) of the probing signal. The permittivity of a material is the measure of the material's ability to resist formation of an electric field within it. The conductivity is a measure of a material's ability to carry an electric current and can have a significant affect on the attenuation of a radar signal. A material's conductivity and permittivity are greatly influenced by water and chloride content. Permeability is a measure of the magnetic polarization of a material, which is generally assumed to be the same in reinforced concrete systems as for free-space. The intrinsic impedance of the material can be approximated by Equation (1) (Balanis 1989).

$$\eta = \sqrt{\frac{\mu}{\varepsilon}} \tag{1}$$

According to Fresnel's Equation, the reflection coefficient (ρ) expresses the relationship between the reflected and incident energy of a plane wave. The reflection coefficient at the interface of an air/asphalt halfspace can be expressed as a function of the intrinsic impedances of the air (η_{air}) and asphalt overlay (η_{asph}). Assuming a normally incident plane wave on a planar interface the reflection coefficient is calculated using Equation (2).

$$\rho_{asph} = \frac{\eta_{asph} - \eta_{air}}{\eta_{asph} + \eta_{air}} \tag{2}$$

When computing the relative permittivity of the underlying concrete layer, the wave interaction with asphalt and the spreading of the wave as it leaves the point source needs to be considered. It is assumed that at the point of normal incidence, the wave is locally considered to be planar. The composite reflected signal from the air/asphalt/concrete combination is the product of the two transmission coefficients into and back out of the air/asphalt interface and the reflection coefficient of the asphalt/concrete interface Snell's Law is used to determine a virtual source point above the air/asphalt interface. If the real transmitter/receiver is located h meters above the interface, rays in the asphalt appear to be emanating from $h_v = h\ (\varepsilon_{r\ asph})^{1/2}$ where $\varepsilon_{r\ asph}$ is the relative permittivity of asphalt. The wave is reduced in amplitude due to the spreading of the wave through the asphalt (of thickness d) on the way to and from the asphalt/concrete interface. The composite reflection coefficient is given by Equation (3).

$$\rho_{conc} = \frac{4\,\eta_{asph}\,\eta_{air}\left(\eta_{conc} - \eta_{asp}\right)}{\left(\eta_{asph} + \eta_{air}\right)^2\left(\eta_{conc} + \eta_{asp}\right)\sqrt{1 + \frac{d}{h_v}}} \tag{3}$$

In order to compute the intrinsic pavement material impedances it is necessary to have the observed reflection coefficient. This reflection coefficient is computed for the simulated cases similar to how it is obtained in the field where a metal plate is

placed on the deck surface to obtain a 'perfect' reflection. The reflection coefficients are computed from the signal response amplitudes as follows:

$$\rho_{asph} = \frac{\text{deck surface amp}}{\text{metal plate surface amp}} \qquad \rho_{conc} = \frac{\text{composite air/asphalt/concrete interface amp}}{\text{metal plate surface amp}} \tag{4}$$

where the early time signal response is assumed to be from the air/asphalt interface, and the second, later time response is due to reflection from the deepest asphalt/concrete interface.

The reflection coefficient at the air/asphalt interface (ρ_{asph}) is extracted from the data and can be used in (2) to solve for η_{asph}. Using the relationship in (1) and assuming the asphalt permeability is the same as freespace, the relative permittivity of asphalt ($\varepsilon_{r\ asph}$) can be calculated. From the permittivity, the virtual source point location (h_v) and velocity of the wave through the asphalt (v, Equation 5) can be calculated.

$$v = \frac{c}{\sqrt{\varepsilon_{r\,asph}}} \tag{5}$$

The data recorded from a GPR system is voltage of the received signal versus time. If the velocity of propagation through a material is known, the depth to an object can be calculated from the recorded two-way travel time. The thickness of the asphalt layer can be calculated using Equation 5 and the previously computed asphalt permittivity.

$$d = v\frac{t}{2} = \frac{c}{\sqrt{\varepsilon_{r\,asph}}}\frac{t}{2} \tag{6}$$

where c is the velocity of a wave through freespace, and t is the observed time between the first and second peaks. Notice that since the wave travels from the transmitter to the object and back to the receiver, the two-way travel time is halved to obtain the actual time to the layer.

The reflection coefficient from the asphalt/concrete interface is extracted from the data, and, since the virtual point source location and the intrinsic impedance of asphalt are known, the permittivity of concrete can be computed by solving (3) for η_{conc} and applying the relationship in (1).

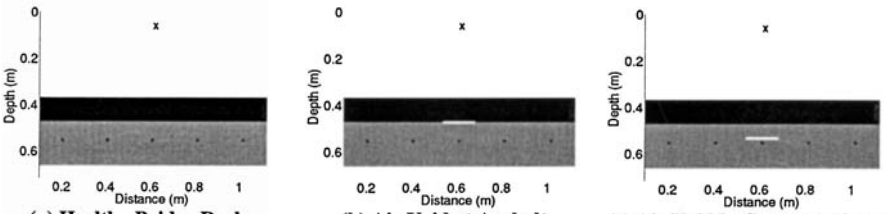
Electromagnetic Simulation

An integrated modeling environment (IME) is being developed by the authors as an analysis tool capable of sophisticated system simulations for the US NSF funded Engineering Research Center for Subsurface Sensing and Imaging Systems (CenSSIS). The IME extends the capabilities of currently available modeling tools. Capabilities include representation complex subsurface features and the randomness of these features, sensor models, and the integration of the sensors with the subsurface environment. Benefits of the IME lay in modeling, simulation, and interpretation capabilities, and it supports improved understanding of system behavior and the ability of a particular modality to detect defects.

The electromagnetic model (EM) for ground penetrating radar incorporated into the IME is the finite difference time domain (FDTD) model. FDTD models efficiently discretize Maxwell's Equations in a robust and flexible manner, and handle multiple inhomogeneities (Sadiku 2001). The simulations presented herein

are performed using the 2D transverse magnetic (TM) mode ($E_X = E_Y = H_Z = 0$). The simulations use a temporal resolution of 5.95 ps through a 12 ns duration. The 3D physical model is discretized using a uniform spatial resolution of 2.54 mm. For all simulations the excitation is a spherical wave and is modeled as a soft source. This excitation signal produces a record of the transmitted signal as it passes through air similar to measured GPR signals collected in the field.

In the examples presented herein, the simulated healthy bridge deck is a 188.0 mm thick reinforced bridge deck with reinforcing steel (rebar) buried 73.7 mm under the surface of the concrete, spaced 203.2 mm on center with a diameter of 12.7 mm. There is a 104.1 mm asphalt overlay. For examples of an unhealthy bridge deck, a 12.7 mm thick and 142.24 mm wide air void has been introduced into the model at the asphalt and concrete interface and in the concrete above the rebar. The simulated bridge deck models are presented in Figure 1. The transmitter and receiver are collocated 307.3 mm above the bridge deck as indicated by the 'X' in Figure 1. For ease of discussion, a baseline simulation is defined using the healthy bridge deck model. The relative permittivities of the asphalt and concrete are 5 and 10 respectively and the conductivity of both materials is 0. This model is a reasonable representation of sensor behavior in the field.



(a) Healthy Bridge Deck (b) Air Void at Asphalt Concrete Interface (c) Air Void in Concrete Above Rebar

Figure 1. Simulated Bridge Deck Models.

Variation of Electromagnetic Properties in a Healthy Bridge Deck

Simulations were performed varying the relative permittivity of asphalt at 4, 5 and 6 and the conductivity at 0 and 0.02 S/m. The relative permittivity of the concrete was simulated at 9, 10 and 11 and the conductivity at 0 and 0.01 S/m. The responses due to the changes in permittivity are shown in Figure 2a and 2b. These are typical in situ parameter ranges. Probabilistic distributions of these basic variables are unavailable. The responses, as expected, indicate lower amplitude when the values of permittivities between two materials are more similar, and a slower wave velocity through a material as the permittivity of that material increased. The responses due to the changes in conductivity are shown in Figure 2c. As one would expect, an increase in conductivity results in greater attenuation of the response (Balanis, 1989).

The relative permittivities of the asphalt and concrete and the asphalt layer thickness have been computed as previously discussed. The results are summarized as Cases 1 through 8 of Table 1. Recall that for all cases the actual thickness of the asphalt layer is 104.1 mm.

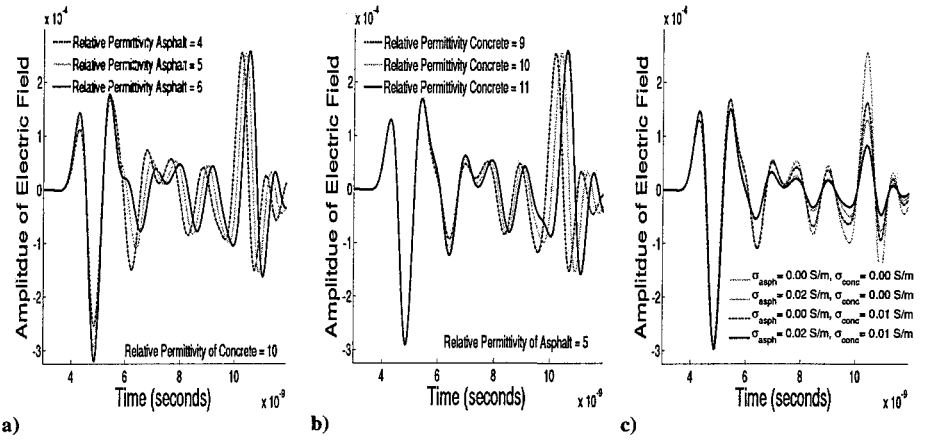


Figure 2. Response due Variation in Baseline Case (Figure 1a) for: a) asphalt permittivity, b) concrete permittivity, and c) conductivity.

Variation of Electromagnetic Properties in a Bridge Deck with an Air Void

To simulate a compromised situation, an air void was placed in the bridge deck model at two different locations as shown in Figure 1b and 1c. Parameters are simulated in the same manner presented in the previous section. The simulations were performed assuming no conductivity and assuming the asphalt and concrete have conductivity (0.02 S/m and 0.01 S/m respectively). Responses with the air void are plotted against the baseline (healthy) response in Figure 3. Additionally, relative permittivities and asphalt thickness have been computed and are presented as Cases 9 through 12 in Table 1.

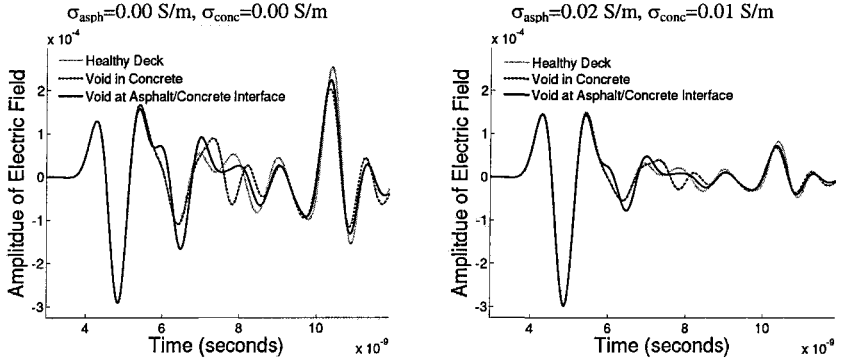


Figure 4. Responses in Bridge Deck with Air void (Figure 1).

Table 1. Summary of Computed Electromagnetic Parameters

	Actual Asphalt		Computed Asphalt		Actual Concrete		Computed Concrete
	$\varepsilon_{r\,asph}$	σ_{asph} (S/m)	$\varepsilon_{r\,asph}$	d_{asph} (mm)	$\varepsilon_{r\,conc}$	σ_{conc} (S/m)	$\varepsilon_{r\,conc}$
Healthy Deck Model							
1	5	0	4.95	104.1	10	0	10.16
2	4	0	3.96	104.1	10	0	10.40
3	6	0	5.95	104.1	10	0	10.05
4	5	0	4.95	104.1	9	0	9.14
5	5	0	4.95	104.1	11	0	11.19
6	5	0.002	5.15	99.1	10	0	7.39
7	5	0.002	5.16	99.1	10	0.001	7.37
8	5	0	4.95	106.7	10	0.001	10.20
Deck with Air Void Above Rebar							
9	5	0	4.95	106.7	10	0	10.28
10	5	0.02	5.16	99.1	10	0.01	7.39
Deck with Air void Between Asphalt and Concrete							
11	5	0	4.95	109.2	10	0	15.24
12	5	0.02	5.15	104.1	10	0.01	8.62

Conclusion

Analytical parametric studies of a GPR investigation of a bridge deck provide important information about the variability of computed dielectric constant. Variation in the actual bridge deck dielectric constants results in variation of the amplitude and time peaks of the response. Amplitude and time information is used to compute velocity of a wave through the material (and therefore depth to subsurface features) and provide an indication of the material condition such as presence of chlorides. Errors in the computation of electromagnetic properties will manifest as discrepancies in the calculation of wave velocity through the media and depth to subsurface features. The process of simulating various controlled conditions with a forward model is integral to developing inverse methods to better characterize and quantify suspect areas in the sub-surface. One example of an inverse method being investigated by the authors is a model-based assessment approach that uses iterative computational models to optimize electromagnetic properties of the subsurface materials and the subsurface layer thicknesses to generate a healthy bridge deck model (Belli et al. 2006).

References

Balanis, C. (1989). *Advanced Engineering Electromagnetics*. New York, NY: John Wiley & Sons.

Belli, K., Wadia-Fascetti, S., & Rappaport, C. (2006). Bridge deck condition assessment using gpr: a model based approach. *Computer-Aided Civil and Infrastructure Engineering*. Submitted for review.

Sadiku, M. (2001). *Numerical Techniques in Electromagnetics* (2nd ed.). Boca Raton, FL: CRC Press.

INVESTIGATION ON THE VIBRATION EFFECT INDUCED BY RAIL TRANSIT SYSTEM ON URBAN ENVIRONMENT

Xiangbai Zhang[1] and Daisy Le Cao[2]

[1]Ph.D. Registered Professional Engineer, Senior Engineer, Dept. of Civil Engineering, Tsinghua University, Beijing, 100084, P.R.China; PH (0086)10-68156899; e-mail: zxb02@mails.tsinghua.edu.cn

[2]M.ASCE Professional Engineer, Dept. of Civil Engineering, Beijing Institute of Civil Engineering and Architecture; No.7 MinZuUniv SouthRd Beijing, 100081, P.R.China; PH (0086)10-86370697; e-mail: architecture001@hotmail.com

Abstract

Rail transit circuit generally passes through the downtown area in a city. The ambient vibration effect induced by rail transit systems has aroused a great deal of public attention. It is necessary to make a comprehensive evaluation on ambient vibration effect to protect a preferable environment. This paper presents a series of detailed analysis of vibration effects on urban environment as wave propagation by different attenuation factors with numerical methods. They make a contribution to study on how vibration effect cause harm to the human body, and damage to building structures. Moreover, this paper put forward some suggestion of countermeasures.

Keywords

Vibration effect, rail transit system, wave propagation, building structure, urban environment.

Introduction

Rail transit circuit generally passes through the downtown area in a city. It is the main mode of public transportation and the integral part of transportation network in many countries. Nowadays, different types of rail transit systems are developed fast as metropolitan subways and high-speed railways. As a matter of fact, problems of ambient vibration effect induced by rail transit systems have come close to urban environment and aroused a great deal of public attention as environmental pollution. Although micro-vibrations may not result in the collapse of buildings as earthquakes do, they actually cause the malfunction and uncomfortable. Therefore, a general demand for increased densities of transportation systems requires a series of satisfied prediction of vibration effects.

Generation & Transmission of Vibration Effect Induced by Rail Transit System

The vibration effect induced by rail transit system is the direct result of strong ground shakings for irregularities of wheel and rail interface with small relative displacement excitation in the vertical direction by track dynamic interaction while trains are running on the track. Such vibrations are transmitted through track structures, including rail pads, sleepers, sub-ballast, and propagated as waves through the soil medium. The consequences of vibration are the damage of buildings and overloaded of the perception threshold for people exposed to the shakings. With a maximum speed of 80 to 120 km/h and an average of over one hundred daily trips including both directions of travel, people living alongside the railway or over the tunnel virtually feel vibration levels high enough to cause rattling of windows, dishes, and resonances of decorative items and furniture in severe cases. Likewise, in general area, people may not feel the vibration, but are aware of it by hearing rumbling noises, formed of low frequency noise waves. Such ambient vibration effects can interact with other parts of human body and cause healthy problems. In addition, adverse vibration effects in long-term will cause severe damage to historic buildings, particularly those in a weak condition.

Performance of Vibration Effect

Wave propagation in solid medium is more complex than the case for air. There are various factors, involved the running speed, the weight of train, the conditions of wheel, rail, tunnel foundation, embankment, and local geology, the distance to the track, and building construction, each of them constitutes a major influence on the final vibration to different degrees. However, those factors can be modeling by a series of simplified methods based on their natural frequencies.

From a historical perspective, the measurement has played a major role in the prediction implemented without some important aspects, such as the emission spectrum of rail vibration. Thus, an integrated design approach combined with measurements analytical formulae and numerical modeling, has turned out to be the best way forward to complex issues. Following the study of underground railway with FINDWAVE® model by a final report of Validation of Ground-borne Noise and Vibration (Cross London Rail Links Company, 2004), this approach has been described by properly using Finite Difference Element Method (FDEM) for predicting vibration of structures at frequencies and simulating dynamic behaviors of track and tunnel systems and the propagation of vibration wave through the ground. Two groups of results comparison in prediction and single measurement are plotted by secondary peaks in one-third octave band spectrum shown as Figure 1 and Figure 2. These procedure data represent that vibration intensity in the vertical direction is almost larger than performance in the horizontal direction.

The vibration intensity is expressed by the peak particle velocity in vibration decibel dB re 10^{-6} mm/sec by motion. Each vector of three dimensions is defined as the synchronized velocity component. Results for the x, y, and z axes describe performance on the lateral, vertical, and longitudinal directions.

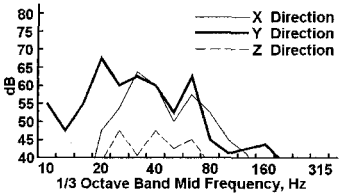


Figure 1. Comparison with predicted results of 3-dimensions vectors.

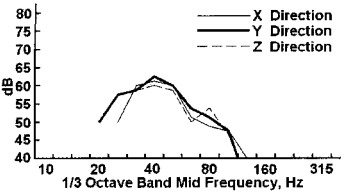


Figure 2. Comparison with measured results of 3-dimensions vectors.

Comparing with each component performance in the prediction and measurement, it has been recognized that vibration effect in the vertical direction is the major mode. Because of the rigidity of a concrete tunnel in the z-axis and its complex modal responses in the x and y axes, in fact, the excitation by forces of gravity and roughness act largely vertically. To a lesser extent laterally, field results normally make only a small contribution to vibration at the ground surface from the z-axis.

Propagation of Vibration Effect in Building Structures

Propagation of vibration effect through the ground is attenuated before it gets through the building foundation, but degrees of attenuation depend on local geological condition such as the depth to bedrock and existence of soil layers. When ground vibration excites a building foundation, it will set the building into a vibration motion and starts vibration waves propagating throughout building structures as walls and floors with amplification of levels, particularly to upper floors. It definitely depends on the nature of building foundation. For the slab foundation, there is little attenuation as slab floor contacts with surrounding soil so closely that vibrations are virtually the same as that which would exist without the slab. For the piled or spread foundation, vibration levels can be reduced much more as compared to the prediction on the ground. Yet buildings with timber floors, vibration can be even amplified, especially as natural floor resonating frequencies coincide with peaks in the ground-borne vibration spectrum.

Prediction of Ground Vibration Effect by Height Influence in Building

Propagation of ground vibration is unconventionally from the foundation to receiver storey in building. The vibration amplitude is supposed to be decreased for vibration energy propagates through the remainder of building. Normal assumption is approximately by 1 to 2 dB for vibration attenuation in each floor. Nevertheless,

building floors, walls and ceilings are actually subjected to significant amplification of vibration by height influence, because of the resonance. For a typical wood frame residential structure, the fundamental resonance is usually in the 15 to 20Hz range. Reinforced concrete slab floors in modern buildings will have fundamental resonance frequencies in the 20 to 30Hz range. As both Australian Standards and British Standards have compliant limits on ground vibrations, the potential disturbance to people has been defined in the range of 8 to 80Hz.

Prediction of ground vibration effect by height influence demonstrates that vibration amplitudes at the ground level are lower than performance at the upper storey. The study of an extensive detailed evaluation with Finite Element Methods (FEM) by Hal Amick's Report (11th ICSDEE, 2004), properly employing measurements for site transfer of vibrations into a multi-storey building, considering the coupling between building foundation and soil, has developed modeling on a site component in the attenuation of ground vibration. Attenuation factors are obtained from the frequency response functions in two interior locations, the 1st and 5th floors, by dividing the motion into the horizontal and vertical directions at a point, given in Table 1.

Table 1. Frequency response functions of vibration intensity.

Frequency (Hz)	Horizontal Vibration Intensity (dB)		Vertical Vibration Intensity (dB)	
	1st Floor	5th Floor	1st Floor	5th Floor
10	/	/	/	/
20	24	/	16	8
30	48	43	10	8
40	40	40	4	18
50	34	37	8	14
60	58	54	21	39
70	28	49	11	29
80	30	43	14	20
90	36	46	10	28

As stated above, most of representative data have been focused on frequencies greater than 20Hz. These actual data produce that vibration intensity in the vertical direction is much lower than performance in the horizontal direction, with different characteristics due to the coupling between deep building foundation and large amounts of soil. Anyway, the 1st floor spectrum data prove out the amplification of vibration in fundamental resonance frequencies, coincidentally, the 5th floor spectrum data also represent the similar resonance peaks at those same frequencies. Moreover, both horizontally and vertically inclined shakings, vibration intensities exhibit the same characteristics of amplification by height influence in building structures.

Attenuation Influence of Vibration Effect by Distance

The attenuation of vibration level decreases by distance from the track as a result of geometric spreading of vibration energy and dissipation by friction nonlinearity of viscosity in soil. According to the vibration intensity of $70\,dB$ in buildings for the perception threshold of human body authorized in the United States, the source of Federal Transit Administration (FTA Manual) developed the criteria for vibration effect assessment based on existing land use activities. The distance criteria for the institutional should be within 100 *feet* ($30.5\,m$). For the residential, it should be within 150 *feet* ($45.8\,m$). For the high sensitivity, it should be within 450 *feet* ($137.2\,m$). The propagation data provide an estimate of perceptible vibration attenuation as a function of distance from the track.

Following a tunnel section project of Beijing Metropolitan Railway with $114\,km$ long (total length of Beijing Metropolitan Railway will reach $200\,km$ by 2008), shown as Figure 3, the numerical modeling and analytical prediction with ANSYS Software Package (FEM-based) are successfully carried out in this paper. The finite element model is generated with specified structural geometry, material properties, and local ground dynamic properties identified in the investigation. Considering the longitudinal dynamic behavior in vibration effect, final attenuation influence by distance in several groups of low frequency bands are calculated associated with general track system and floating slab track system. Figure 4 indicates some results at typical frequency bands that modal characteristics have been predicted analytically with dynamic responses.

Figure 3. Typical tunnel section of Beijing Metropolitan Railway.

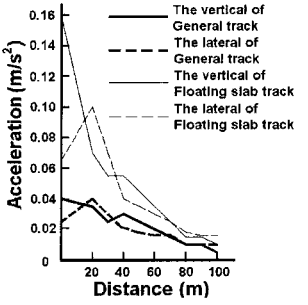


Figure 4. Attenuation influences by distance with general track and floating slab track system.

The vibration intensity is expressed by the peak acceleration m/sec^2 in motion. Typically, attenuation factors are still obtained from the frequency response functions by dividing the motion into the lateral and vertical directions at a point. Results of prediction illustrate exactly: (i) Attenuation influence by distance in the vertical direction is more conspicuous than performance in the lateral direction; (ii) Attenuation influence by distance with floating slab track system is more conspicuous than performance with general track system. These practical data of

attenuation influences by distance for vibration effect make a contribution to optimize design schemes on track structures.

Suggestion of Vibration Isolation

Traditional engineering approaches to reduce vibration remain cost effective as making stiffness, mass changing, and modifying the building section. Some new solutions are still price competitive. Among various countermeasures for controlling ground vibration effect induced by rail transit systems, the installation of floating slab track system, vibration reduction wall, wave isolation by trenches and impeding blocks, are common and effective ways. Furthermore, the ambient vibration caused by construction of rail transit systems could excite resonances. Within 50 to 100 m of such works with heavy plant equipments, vibrating rollers, rock hammers and pumps, it can make areas disturbed at those sensitive locations close to the work activities. It is recommended to develop a vibration management plan to monitor and assess potential vibration impacts on a basis as each phase of work implemented at nearby residential areas to ensure compliance with goals.

Conclusion

Vibration effects are challenging almost every design state, including new rolling stock designs, better track structures, improved trackside facilities and optimized building structures. This trend has forced the structural engineering community to adapt its historical approach to include a significant focus on vibration in general and on methods of reducing the vibration impact. In this paper, detailed analysis of vibration effect as wave propagation by different attenuation factors with numerical methods are introduced, and conclusions are made as following: (i) Rail vibration intensity in the vertical direction is almost larger than performance in the horizontal direction; (ii) Rail vibration intensity is higher at the upper storey than performance at the lower because of resonances of floor and ceiling systems; (iii) Attenuation of vibration level decreases by distance from the track. The planning of rail transit system construction should be established with an appropriate distance to residential areas. Overall, to avoid disturbance and damage by vibration effects, well managed work in construction process is necessary, particularly in those sensitive locations.

References

Cross London Rail Links Company. (2004). *Validation of Ground-borne Noise and Vibration Final Report*, London

International Conference on Soil Dynamics & Earthquake Engineering (11th ICSDEE) (2004). *The Role of Buildings and Slabs-on-Grade in the Suppression of Low-Amplitude Ambient Ground Vibrations*, Hal Amick, Calif.

Real-Time Three-Dimensional Object Detection and Tracking in Transportation

J. Teizer[1], C.T. Haas[2], C.H. Caldas[3], and F. Bosche[4]

[1] Jochen Teizer, ASCE Student Member, Ph.D. candidate, Dept. of Civil, Arch. and Env. Eng., Univ. of Texas at Austin, 1 University Station C1752, Austin, TX 78712; PH (512) 554-7857; FAX (512) 471-3191; email: teizer@mail.utexas.edu
[2] Carl T. Haas, ASCE Member, Professor, Dept. of Civil Eng., Univ. of Waterloo, 200 University Ave West, Waterloo, Ontario, Canada N2L 3G1; PH (519) 888-4567 x5492. FAX (519) 888-4300; email: chaas@civmail.uwaterloo.ca
[3] Carlos H. Caldas, ASCE Member, Assistant Professor, Dept. of Civil, Arch. and Env. Eng., Univ. of Texas at Austin, 1 University Station C1752, Austin, TX 78712; PH (512) 471-6014; FAX (512) 471-3191; email: caldas@mail.utexas.edu
[4] Frederic Bosche, ASCE Student Member, Ph.D. candidate, , Dept. of Civil Eng., Univ. of Waterloo, 200 Univ. Ave West, Waterloo, Ontario, Canada N2L 3G1; PH (519) 888-4567 x3872. FAX (519) 888-4300; email: fbosche@civmail.uwaterloo.ca

Abstract

Flash LADAR (Laser Detection and Ranging) devices are one of the recent technology developments which allow rapid spatial data acquisition of scenes. Algorithms that can process and interpret the output of such enabling technologies into three-dimensional models have the potential to significantly improve work processes. A particular important application in transportation is modeling the location and path of objects to make construction processes safer and more secure. Once people and equipment are detected and mapped into a three-dimensional computer model, their path can be analyzed and access to hazardous areas can be restricted. This paper presents experiments and results of a real-time three-dimensional modeling technique to detect static and moving objects within the field of view of a high-frame update rate laser range scanning device. Applications related to transportation and construction are specified.

Introduction

The ability to locate, describe in 3D, control, and track objects within a field of view has become an important factor in intelligent transportation systems, construction, maintenance, and asset management. Assets usually are somehow scanned and then modeled in 3D at varying frequencies, depending on the application. Asset management may require no more than bi-annual updates, whereas vehicle navigation or construction activities may require real-time updates (Kim and Haas, 2002). An overall framework of data acquisition and 3D model building is presented in Figure 1. It illustrates that in practice, 3D modeling requires combinations of top down design, bottom-up data acquisition, and comparison of both sources of information in many cases for individual assets. Typically, design processes provide

well defined information including perfectly parallel, perpendicular, flat, etc. forms (strong forms) like pipes, beams, columns and floors, whereas weak and non-parametric forms are produced from existing infrastructure conditions. Hirschberg (1996) defines "weak forms" as forms related to strong forms, but previous design information was improperly documented. Over time, for example, a rectangle may become an irregular four sided polygon to fit a distorted wooden beam or a cylinder which may grow a joint to represent a bent pipe. Non-parametric forms include wire nets that may represent contour data, polylines that can represent cracks and occupancy arrays or octrees that can represent amorphous volumes or deformed objects. These forms can generally be derived from range point data which contains the distance information in an array of pixels of the original scene image.

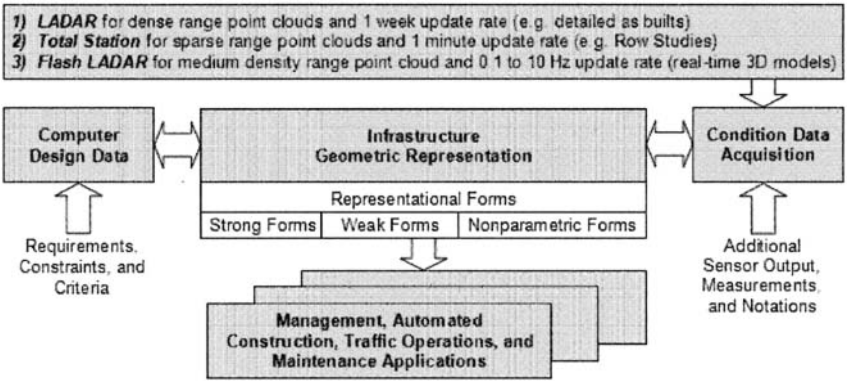


Figure 1. Real-time 3D Modeling Framework (Teizer et al. 2005)

Existing and emerging laser ranging technologies like LADAR (Laser Detection and Ranging), Total Stations, and Flash LADAR can help in the need for fast and accurate geometric modeling. New sensor technology innovations now allow us to address problems of the highest priorities in the transportation and construction area.

Background

Range scanning methods in 3D modeling can be categorized in three different categories. All are able to capture range information of as-built environments which is later on processed to 3D models. Namely they are dense point cloud approach, sparse point cloud approach, and Flash LADAR approach (Teizer et al. 2005). A dense point cloud approach uses LADAR technology that is able to scan high details of static environments. Due to a complex mirror based scanning system, the acquisition phase of millions of points requires anywhere from 20 seconds up to several hours and does not include the processing phase to convert range data into a 3D model. Moreover, laser range scanning equipment purchase cost is high and easily can reach several hundred thousand US dollars (Stone et al. 2004). A sparse point cloud approach uses selected points to model static environments. Major edges, corners, and lines may already describe an object in a significant manner. Thus, the

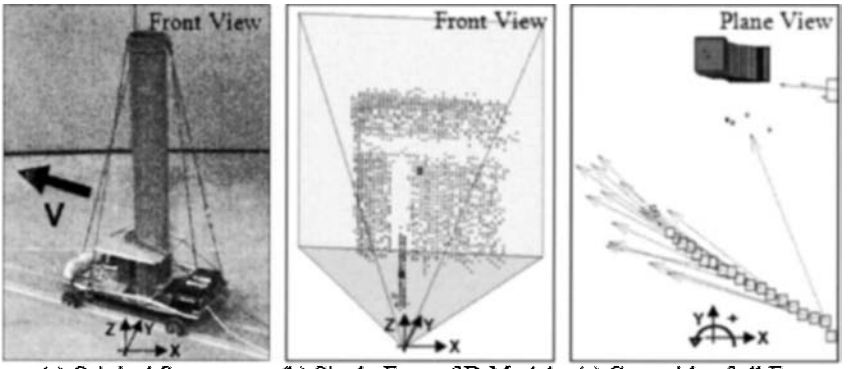
advantage of manually selecting fewer points with lower purchase cost of mass manufactured commercially available laser range finders and faster range data acquisition in the magnitude of seconds, reduces the overall modeling approach to minutes (Kwon et al. 2004). To overcome the limitations of not modeling in real-time at frame rates above 1 Hz and to advance the applicability to dynamic environments where objects are moving, technology is needed that can address static and moving situations in a rapid manner. Emerging Flash LADAR technology offers to scan static or dynamic environments at high frame update rates as well as provides a dense point cloud required for accurate real-time three-dimensional modeling.

Flash LADAR Approach, Experiments, and Results

Conventional laser range scanning methods are based on interference and triangulation techniques as well as of time-of-flight measurements. We used emerging technology based on the time-of-flight principle that uses a CMOS sensor and pulsed active illumination to generate a 3D array of pixels (image). To each pixel on-board the sensor chip, the system measures the phase shift between a reference signal sent directly to the detector's smart pixels and light reflected by the scene. Since the measured wave is cyclic, the distance measurement is not unambiguous. Under the current prototype development of the sensor, each pixel stores information values of range and intensity to objects contained in a scene at the unambiguous range of 7.5m. Changing the modulation frequency of 20 MHz increases the acquisition range, but reduces the illuminated power send out into the scene in form of an intensity-modulated wave front, ultimately limiting the accuracy of range measurements. Technological improvements are under way and will use higher power laser diodes instead of conventional near-infrared LEDs, or focus on applying two different modulation frequencies. As a result, the current inexpensive technological development stage limits the potential use to important feasibility studies rather than commercial end product driven solutions in real-time 3D modeling.

This research focuses on the application of detecting and tracking objects in the field of view of a 3D range camera. Rapid range measurements in transportation construction are needed, to navigate vehicles safely in paths were potential hazards and obstacles are avoided. We have conducted 27 indoor experiments to validate the accuracy of range measurements as well as to test our algorithm development for processing range information. The sensor's field of view (horizontal 42° and vertical 46°) has a resolution of 124x160 pixels, thus achieving an angular accuracy of 4.6cm at the furthest distance (7.5m). Once range data is acquired and stored in an array, our processing algorithm based on occupancy grid system is applied to the range information. Grouping range points into voxels, filtering noise measurements, and applying a hierarchical clustering technique reduced data significantly to define features in each frame. Comparing features in each cluster such as volume, speed, direction, and location in consecutive frames, allowed separating static from moving objects. Since the location to all centroids in the clusters where known, velocity vectors were applied. Figure 2 demonstrates the process of building an occupancy grid from an original scene. In the example, we used a cart with a mounted pipe to

simulate a moving vehicle (Figure 2a). Guided on rails at angles of 0/90/150 degrees to the X-axis, range frames were collected from static and moving boxes and pipes at frame update rates of 15.2 Hz. The processing algorithm converted each range frame in real-time (greater than 1 Hz) into voxel based models. In Figure 2b the pipe is represented by one cluster and also the background wall; the ground and cart are out of the sensor's field of view. Further analyzes classified features such as centroid locations, direction of movement, and velocities (Figure 2c). On average the errors in accuracy were: Object position +4.5% (positively off in all axis directions X, Y, and Z), object dimension +30% (half of which is due to the fact that line-of-sight prevents from modeling the depth Z, thus only 2½D instead 3D), direction -1.2° (further away than in reality), and speed +6.5% (faster than in reality). Errors most likely stem from measurement errors, but can also partially come from systematic errors. The results are promising and show that real-time 3D detection and tracking of objects is feasible.



(a) Original Scene (b) Single Frame 3D Model (c) Centroids of all Frames

Figure 2. Real-time 3D Shape and Path Modeling of a Simulated Vehicle

Future research steps will be extensive testing of the algorithm in outdoor environments. Initial tests indicate that the algorithm works successfully, however, due to the prototype stage of the sensor, it physical working principle does not allow full sunlight on the reflecting scene. Since sunlight is 400 times brighter than existing LEDs, power emitted by the sensor will not be reflected. For this reason, we successfully tested our algorithm in cloudy weather. We add at this point that the next generation Flash LADAR (to be released in early 2006) will significantly increase the range of non-ambiguity by the use of two or more frequencies and supply active background light suppression that will permit measurements in full sunlight.

Applications in Transportation

Many application areas for 3D cameras exist. One predominant area of current interest in range cameras for transportation is their use in safety. In 2003 the US experienced 42,643 deaths from motor vehicle crashes. The mileage death rate dropped from 4.88 deaths per million miles traveled to 1.56, and deaths from 100,000

population dropped from 26.8 to 15.4 (O'Neill, 2005). Nevertheless, the overwhelming bulk of serious crashes stems from drivers breaking the traffic laws or not paying attention. Brian O'Neill argues that scientific evidence has shown that more education does not change the drivers' behavior. Instead progress can be achieved through stricter law enforcement and political action on state level to create incentives for the use of existing or emerging technologies which can improve safety. The Federal Highway Administration (FHA) and National Highway Safety Administration (NHSA) have eased the requirements to remove hazardous road design features and upgraded safety on roadsides under renovation, by establishing new standards for roadside clear zones, breakaway lights and poles (O'Neill, 2005).

Continental Automotive Systems (2001) states, that 40% of all road casualties can be prevented once the entire response time of a vehicle breaking system can be reduced by 0.5 seconds. The human judgment of analyzing a road scene and applying force to the break system of a vehicle is at around 1 second. We believe that assisting the human eye with a real-time 3D sensor based vision system can help make aware of upcoming hazardous situations, since the response time can be reduced to the minimum time a sensor needs to acquire, process, and analyze a frame shot. Once a danger appears in the vehicle's path, a Flash LADAR device could reduce the entire response time to better than 0.1 seconds (greater than 10Hz). This largely depends on developing range sensing technologies to a commercial level where longer distances in outdoor environments and faster processing algorithms of dense point cloud data are fully developed. The car industry, for example, is working on a system for the robust measurement of the distance between two vehicles.

More applications exist in the vehicle guidance, such as the assistance of the drivers' perception in rear or side parking situations by automated scanning of confined spaces, slower inflation of airbags when persons are close to the dashboard, steering wheel, or airbags that should not inflate once baby seats are installed. Similar to belt use, three-dimensional laser scanning technology and processing can help to reduce injuries and fatalities in pedestrian/car related accidents. Running red lights becomes one of the most dangerous situations at intersections when pedestrians or bicycles are involved. Range scanning of entire or significant street areas might detect and track the movement of pedestrians or bicyclists who willingly, by fault, or by ignorance enter crosswalks or change lanes. Alarms can be set off (in cars or at pedestrian traffic lights) to prevent impacts or help in making a decision to react for a best-case scenario. Other application areas are in the robotics industry where automated robots need to navigate autonomously and, in the security industry, inexpensive measurement systems are being planned for the surveillance of complete rooms. Observation systems may even assist the prioritization and scheduling of elevator transportation (Opto & Laser Europe, 2004)

In construction related to transportation, the industry reported in 2005 that 23% of all 1,124 fatalities were related to transportation and 22% had contact to objects and equipment (BLS, 2005). These figures indicate that close to one half of all accidents in construction are related to transportation where vehicles, such as cars, trucks, or

heavy equipment, apply deadly forces to workers. By modeling workspace in real-time and in 3D, the operator's perception can be improved, safety work zones can be defined and restricted access can be granted. Equipment or robots that track or handle materials can be equipped with Flash LADAR devices to safely navigate construction sites. Flash LADAR sensing can be applied in simulating work processes by comparing 3D as-built models vs. as-planned data. Real-time 3D modeling further can help to monitor health progress in life-cycle management for transportation infrastructure, such as inspection processes for defect detection.

Conclusion

We have presented in our research the general feasibility of using real-time three-dimensional modeling as a detection and tracking tool for static and moving objects. Efficiently and rapidly built 3D models after the occupancy grid principle allowed to acquire, convert, and store range information data from a high frame update rate laser range camera. Since this research approach includes the dimension of time, tracking and path planning for static and moving objects from a dynamic platform or vehicle becomes feasible. This real-time 3D modeling technique offers a wide field of potential applications in transportation applications related to safety, such as vehicle navigation and pedestrian protection. Other application areas in the field of transportation exist and may be pursued in the future by modeling infrastructure, tracking materials, detection of structural defects, or life-cycle monitoring of assets.

References

Bureau of Labor Statistics (2005). http://www.bls.gov/iif/oshcfoi1.htm (Dec. 8, 2005).

Continental Automotive Systems (2005). "Nach dem Bremsweg wird der Reaktionsweg kürzer." http://www.conti-online.com (June 10, 2005)

Hirschberg, U., and Streilein, A. (1996). "CAAD meets Digital Photogrammetry: Modeling 'Weak Forms' for Computer Measurement." *Automation in Construction*, 5, 171-184.

Kim, Y.S., and Haas, C.T. (2002). "A Model for Automation of Infrastructure Maintenance Using Representational Forms." *Automation in Construction,* 10(1), 57-68.

Kwon, S., Bosche, F., Kim, C., Haas C.T., and Liapi, K.A. (2004). "Fitting Range Data to Primitives for Rapid Local 3D modeling Using Sparse Point Range Clouds." *Automation in Construction*, 13, 67-81.

National Highway Safety Administration (2005). http://www.nhtsa.gov (Dec 1, 2005).

Stone, W.C., Juberts, M., Dagalakis, N. Stone, J., and Fronczek, C. (2004). *Performance Analysis of Next-Generation LADAR for Manufacturing*, Construction, and Mobility, NISTIR-7117, Gaithersburg, MD.

Teizer, J., Kim, C., Haas, C.T., Liapi, K.A., and Caldas, C.H. (2005), "A Framework for Real-time 3D Modeling of Infrastructure," Trans. Res. Rec. 1913, 177-186.

O'Neill, B. (2005). "Improving U.S. Highway Safety – Have We Taken the Right Road?" TR News, 239, 24-27.

Opto & Laser Europe (2004). http://optics.org/articles/ole/9/12/2/1 (June 9, 2005).

The Case Study of TQC Controlling in Logistics Project Management

JIE WU[1] and QIYUAN PENG[2]

[1]School of Economics & Management, Southwest Jiaotong University, P.O. Box 610031. Chengdu, Sichuan Province, China; ST North Yihuan 111; PH (86) 13008162972; email: jie.wu@yeah.net
[2]College of Traffic & Transportation, Southwest Jiaotong University, P.O. Box 610031. Chengdu, Sichuan Province, China; ST North Yihuan 111; PH (86) 028-87600750; email: qiyuan-peng@home.swjtu.edu.cn

Abstract

At present, there are many questions regarding the application of 3PLs (the 3rd part logistics) in the logistics industry. 3PLs relate to meeting the needs of client sufficiency and efficiency, as well as improving the service benefits provided to clients. It has been determined that project management can offer the solution to 3PLs' developing logistics service projects. The 3PLs are also intent on implementing PM (project management) in spite of the absence of PM tools, which combine with logistics practice. This paper is based on analyzing the trade-off between time, quality, and cost, by using a multi-variable math optimization model in logistics service PM. In addition, a case study is provided to support the theory that the PM procedure can apply to 3PLs' service projects with the used of one 3PLs' true data.

Introduction

Projects touch people's lives in many ways. The types of projects which are especially influential within society are construction projects, engineering projects, computer science, organizational change projects, and even social service industry, such as logistics service projects. Even though project management (PM) has been developing over fifty years, it is a relatively new discipline. As much as the logistics revolution is a global phenomenon, it is taking place in countless locations around the world, where traditional structures are being re-formed, and individual companies have to adjust or perish. As the fastest growing economy in the world, the Asia-Pacific region is facing new challenges and opportunities in the logistics revolution (Feng and Chia, 2000). Many firms have relocated their transportation facilities, warehouses in order to transform themselves from forwarding to 3PLs service providers. In China, the application of PM in the 3rd logistics service projects (LSPs) is integral to the further development of China's industries. This is

especially challenging due to China's recent accession into the WTO. This paper intent to create a deeper understanding of PM within 3PLs, by bridging the gap between the operational logistics projects in practice and Project Management theories within 3PLs. The first section of this paper focuses on why industries apply common PM mechanism in LSP, and how they use the PM tool in 3PLs. The second Section introduces one 3PL in China, and provides a case study on its true data in the management area of LSP. Finally, the general condensed summary is presented.

PM mechanism application in LSP

Define or specify the Logistics service project (LSP). In China, as more solution-oriented 3rd logistics service providers are entering the market, it is more difficult to analyze logistics project. This difficulty arises mostly because responses from 3PLs are not directly comparable and can suggest quite different solutions.
The contract which is the project object to be achieved between buyer and supplier is the only authority. According to Turner, a project is: an endeavor in which human, material and financial resources are organized in a novel way, to undertake a unique scope of work, of given specification, within constraints of cost and time, so as to achieve beneficial change defined by quantitative and qualitative objectives. Similarly, the business implementation in 3PLs is managing the contracting process—from pre-sale to contract closeout. While the PMBOK (Project Management Body of Knowledge) knowledge areas of scope, time, cost, risk, quality, human-resources, communications, and procurement management must be considered by LSP manager. Moreover, the manager must master the following three areas at a minimum: time, quality and cost. Therefore the key of logistics service project contracts belong to these areas as common projects.

An important conceptual step is to define or specify LSP. The definition is a unique series of logistics services activities within constraints of limited resource in order to provide the specifiable services to customers in this application research. As long as it has the conditioned category of unique, these logistics services in the special purpose belong to LSP, no matter how long and how much they are involved.

The difference between managing the LSP and logistics routine operation. Within the business of 3PLs, there are three tasks: the unique, the provisional, and the routine. Generally, they deal with the provisional or the unique measures, but not with mature ways. At first the managers usually do not understand any of the aspects during the process of PM. They only know some isolated chains of logistics. As firms gain experience with the LSPs, they transform the simple business into advanced logistics business. The firms, which adhere to the principles of LSP, can survive within the competitive environment. Furthermore, firms can operate the

management process in professional ways. That is to say they transform the unique or the provisional operations to routine operation with the use of some mature measures. It is obvious that the trend and development of LSP is the key to 3PLs within developing industries. As a result of analytical comparison, the difference between the unique project is observable. So the managers and the firms can apply the tools of PM under LSP conditions upon realizing the characteristics of LSP.

The 3PL service projects, which include transportation, warehouse and information systems in a wide scope, have many characteristics. Especially high-risk projects are those with fluctuating schedules, changing locations and unstable staffs. In general, services differ from goods. The services are (Zeithaml et al., 1985): intangible, heterogeneous (not standardized), inseparable (meaning difficult to separate production of the service from the consumption), and perishable (not possible to stock) Logistics services still have these same characteristics as common services described in service literature. In logistics projects, services are regarded as the source of increment. In order to analyze Logistics Service Project's speciality to define project management's application, we spread out them as following Table 1:

	Logistics Service Project	Logistics Routine Operation
Activity property	Repetition work, single & basic services	Multiple and bundled services on condition of uniqueness
Activity environment	Relatively close & tangible	Relatively open & intangible outcome requirements
Activity style	Divided the work by function	Cooperation by the team
Organization mode	Organized by function	Organized by project team
Management manner	Base on function & department	Base on activity & process
Focus on	Emphasis on function management & command	Emphasis on cooperate & integrated management

Table 1. Compare the two style of logistics operation.

Define or specify the PM of LSP. Project management is the process by which a project is completed successfully. The PMBOK Guide defines project management as: the application of knowledge, skills, tools and techniques in order to meet or exceed stakeholder requirements from a project. When we combine PM theory with LSP practices, the characteristics of LSP must be emphasized. They focus on: the high professional team in PM, the special requirement of individual project, the

man-made terminal of project, and the integrated wholesale.

We define the PM of LSP on the basis of the above characteristics as: the management makes use of the PM knowledge, tools and methods to meet the requirements of every aspect in LSP relationship. The firms thereby gain the benefits of the project as well as the riches within society increase. In order to achieve all LSP's goals, the first step is to develop the PM in LSP. The bottleneck is how to integrated manage Logistics Service Project basing on the trade-off multi-aspect, such as time, quality and cost of a logistics service project (TQC of LSP). So far as the three key variables in the project are concerned, and the integration of PM is successful. Moreover program management is resorted to some industries in the form of PMBOK. In practice, it is impossible to be successful if LSP develop on condition that isolated aspect manages. Though it is important to deal with time, quality and cost of project in professional ways, we cannot separate the three aspects from whole. Moreover it is more important to control the relation of every factor and their variable trend in the mass. In the interest of every interrelated part from firms to society, they expect to benefit from value chain of it.

Demonstration

When we analyze the equilibrium and variable profit and loss characteristic of the three key factors (TQC) which are trade-off with each other, we combine project characteristics with 3PLs of LSP, bring up three type models. For simplification, we assume that Cost and Quality of activity vary as linear function of completion Time.

Elucidation of the model factor. The data for this study were obtained from one of the largest 3PLs in Korea of Chinese branch offices. The firm is selected because it has focused on Korean company and Outsourcing trade companies which advanced into China. The operation has the characters of representative PM. For instance, most logistics project teams always include participation by two-to-three external organizations in cooperation with the other functional logistics organization.

For predigesting, the model must involve some assumptions. Every main factor has an optimal data opposite to the other two factors under bounded conditions. Approval measuring crash quantity data is difficult. We suppose the most satisfaction degree is 110% according to returning of customer approval and investigation etc. over a long period of time. Therefore, crash quantity data can be calculated by compare satisfaction degree with added cost.

The elucidation of the adoption's data. Each LSP transport location includes land and sea transportation, destination is Inchon harbor. The activities of each

logistics project include ***warehouse & request shipping, inland transport, field station, customs clearance, load cargo, sea transportation,*** and ***pick-up the cargo.*** The pack spec of container is NO 20'or NO 40', three models aim at different pack respectively: Gather NO 20' only, gather NO 40' only, and gather NO 20' mix with NO 40'. Quantity unit of the model is unit 1. Because the cost of each ***customs clearance*** is lone and the cost of ***field station*** is fixed cost, as the two activities are not key path, the PM manager do not bring in to establish models instead of adding it in the end. The enactment of parameter is specified at Table 2.

Parameter	Enactment
NT	Normal time for activity(i, j)
CT	Crash time for activity(i, j) (NT≥CT≥0)
NC	Normal cost for activity (i, j)
CC	Crash cost for activity(i, j) (NC≥CC≥0)
NQ	Normal quality level for activity(i, j)
CQ	Crash quality level for activity(i, j)

Table 2. Parameter enactment.

Time (unit is hour), cost (unit is RMB Yuan) and quality data of different pack spec of container are shown as form below Table 3.

20'	NT	CT	NC	CC	NQ	CQ
Warehouse & request shipping	24	12	540	650	1	0.9
inland transport	24	12	1060	1270	1	0.9
load cargo	24	12	280	440	1	0.7
sea transportation	28	24	2240	3520	1	0.7
pick-up the cargo	48	24	280	440	1	0.7
total	148	84	4400	6320		

Table 3. Data of NO 20' only.

Modeling.

Model 1. Optimal Project time objective when cost and average quality are bounded
Model 2. Optimal Project cost objective when time and average quality are bounded
Model 3. Optimal Project quality objective when time and cost are bounded

Model 1. Each activity includes two variables (completion time and earliest time). There are 10 variables and 17 constraints for each model. Optimal solution is solved using software LINDO. We can gained optimal completion time aim at different enactment of bound constraint for completion costs and average quality. It is specified at Figure 1. Optimal solution of cost and quality can be obtained via mode 2 and model 3.Data of NO 40' only, and NO 20' mixing NO 40' are in the same way.

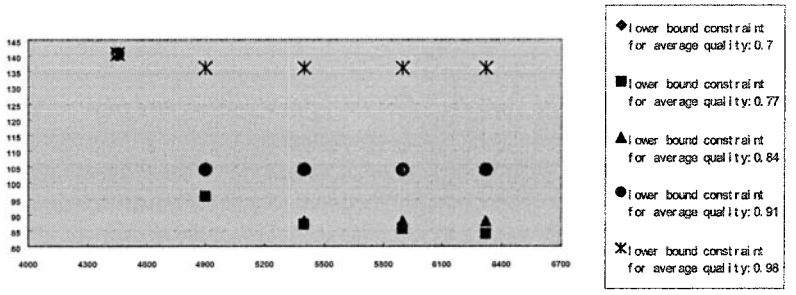


Figure 1. Optimal completion time.

Conclusion

This paper provides the methods and mechanisms used within trade-off factors of Time, Quality and Cost. These factors reflect the equilibrium and variable profit and loss characteristics within the logistics service project. In addition, this paper provides a guide for future research in this area. It also provides a new view for studying and applying logistics management within 3PLs. The three types of models have a stronger direct significance to project decisions than experience ways in traditional logistics firms. The methods discussed in this paper are applicable to all 3PL logistics services that utilize projects as their dominant business strategy.

References

Turner JR. (1993). "Improving the processes for achieving strategic objectives." *The handbook of project-based management,* McGraw-Hill, London.

Gardner, R. W., and Johnson, C.L. (1994). "Third Party Logistics. " *In: Robson, J.F., Copacino, W.C. (Eds.), The Logistics Handbook,* The Free Press, New York.

Frederick, L. Ayer., and William Bahnmaier. (1995). "Toward a defense extension to the project-management body of knowledge." *International Journal of Project Management.,* Vol. 13 No.2.

Babu A, Suresh N. (1996). "Project management with time, cost, and quality considerations." *European Journal of Operational Research,* (88): 320—327.

(2000). "Project Management Institute, A Guide to The Project Management Body of Knowledge." *PMI Standard Committee.*

Gert Wijnen, and Rudy Kor. (2000). "Managing Unique Assignments: A Team Approach to Projects and Programmes." *Twynstra Management Consultants,* Originally published in English by Gower Publishing Limited

Shapiro R, and Rosenfield D, and Bohn R. "Implication of Cost-Service Trade-Offs on Industry Logistics Structures." *Harvard Business School Working Paper*

Determination of Safe Distance between Railway Out-of-gauge Goods and Structure on Straight Lines

Mei Han[1] Boling Han [2] Liang Pang [3] Yanhui Han[4]

[1]Associate Professor, [3]Master Candidate, [4]PH.D Candidate School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China; Email: han_mei2005@126.com
[2]Senior Engineer, Transportation Bureau, Chinese Ministry of Railway, Beijing 100083, China

Abstract

Out-of-gauge goods are more prone to lead to train operation accidents than ordinary goods. Therefore, it is very necessary to keep safe distance between out-of-gauge goods and structure. According to this, the paper analyzes the constitutions of the safe distance and puts forward the mathematic model to determine the minimum safe distance. Also an example is applied to calculate the minimum safe distance.

Introduction

Most out-of-gauge goods are large equipment, such as stator and rotor of dynamoelectric machine, metallurgical equipment, chemical facilities and large bridge components. Most of these goods are loaded on railway flat cars. Whether they can be transported safely is not only very important to the receivers but also influential to the key projects. The past research and experience indicates that out-of-gauge goods are more prone to lead to transportation accidents than ordinary goods are. Therefore, more attention should be paid to guaranteeing safety of out-of-gauge goods transport.

In order that the out-of-gauge goods could be safely delivered by railway, the goods must be carried without being touched by structures whenever the car loaded with out-of-gauge goods runs on straight lines or on curves.

The necessary distance between out-of-gauge goods and structure must be kept properly. Otherwise, the safety of car loaded with out-of-gauge goods cannot be guaranteed, or the structure clearance cannot be fully utilized. According to Chinese *Regulations on Transportation of Railway Out-of-gauge Goods*, Distance of at least 150mm must be kept between out-of-gauge goods and structure when the car loaded with out-of-gauge goods passes the structures nearby with normal speed (China's Ministry of Railway 2003). In fact, the distance of 150mm is determined with reference to relative rules of the former Soviet Union. It is lack of theoretical basis according to Chinese railway fact, and also the distance is determined without experiments. Therefore, it is very necessary to study the minimum safe distance between out-of-gauge goods and the structure to secure the safety of running car and

to fully utilize the structure clearance.

The minimum safe distance mainly depends on offsets of the car loaded with out-of-gauge goods.

Offsets of the car loaded with out-of-gauge goods

Offsets of the car loaded with out-of-gauge goods include static offsets and rolling-swaying vibration offsets (figure 1). The minimum safe distance between out-of-gauge goods and structure cannot be less than offsets of the car and preparatory value.

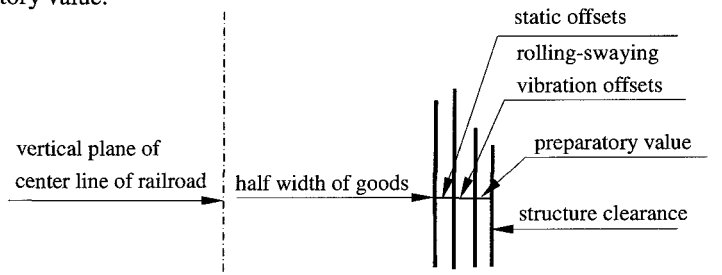


Figure 1. Minimum safe distance between out-of-gauge goods and structure

Static offsets of the car loaded with out-of-gauge goods

Static offsets of the car means the deviation of the longitudinal center line of the car from the center line of railroad. There are some factors affecting the deviation amount, such as lateral clearance of running parts of cars, free play between rail and wheel and the distance between two center pins of the car.

Ⅰ. Lateral clearance of running parts of cars

There is clearance among different parts of car such as free play between body center plate and bogie center plate. And the clearance will lead to deviation of longitudinal center line of car from center line of railroad. The clearance is decided by the bogie type. At present, there are five types of bogie of the car used to load out-of-gauge goods, which include Bogie 8A, Bogie 8AG, Bogie 8G, Bogie K2 and Bogie K4. According to Chinese relative rules on car maintenance, the possible maximum clearance is listed in table 1 (China's Ministry of Railway 2004; China's Ministry of Railway 2003).

From table 1 it can be seen that the total maximum clearance of vehicle structure is 16.25mm. It can be taken as 16.5mm.

Ⅱ. Free play between rail inner side and wheel edge outer side

According to Chinese relative rules on car maintenance, the permitted minimum distance between the wheel pair inner sides is 1350mm and the permitted minimum thickness of the wheel edge is 23mm, as shown in figure 2(China's Ministry of Railway 2003). Therefore, the permitted minimum distance between outer sides of the wheel pair can be:

Table 1. The Lateral Clearance of Car Structure

No.	Origin of clearance	the Maximum (mm)			
		Bogie 8A	Bogie 8AG, Bogie 8G	Bogie K2	Bogie K4
1	clearance between body center plate and bogie center plate	4.5	2.5	2.5	2.5
2	clearance between side frame and bogie bolster	4	4	1.8	4
3	clearance between roller and pedestal	6	7	7.25	9
4	clearance between roller and journal box	0.75	0.75	0.75	0.75
	sum	15.25	14.25	12.3	16.25

$$1350+23\times 2=1396 \quad \text{(mm)} \tag{1}$$

Chinese railway track gauge is 1435mm. According to Chinese *Regulations of Railway Technical Operation* (China's Ministry of Railway 1999), the permitted maximum gauge on straight lines is:

$$1435+6=1441 \quad \text{(mm)} \tag{2}$$

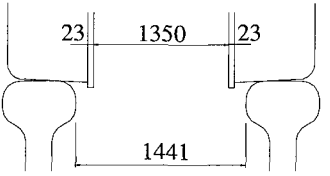


Figure 2. Free Play between Rail inner Side and Wheel Edge outer Side

The possible maximum free play between rail inner side and wheel edge outer side is:

$$\frac{1441-1396}{2}=22.5 \quad \text{(mm)} \tag{3}$$

III. Static offsets of the car on straight lines

The possible maximum deviation amount of center pins of the car from center line of railroad can be:

$$e=16.5+22.5=39 \quad \text{(mm)} \tag{4}$$

When the car runs on straight lines, the most unfavorable deviation is illustrated in figure 3 and the maximum offsets occur.

In general, distance between center pins of two bogies of railway flat car is not more than 9.35m (i.e. $l\leq 9.35$m), and the quotient of length of car divided by distance between center pins is 1.41 or so (i.e. $L/l\approx 1.41$).

Therefore, the maximum static offsets of can will be (Mei Han 2002):

$$S_s=\frac{L}{l}\times e=\frac{39L}{l} \quad \text{(mm)} \tag{5}$$

Among them,

L—length of railway car, m;

l—distance between center pins of two bogies of railway car, m;

e—maximum deviation amount of center pin of car from center line of railroad, taken as 39mm.

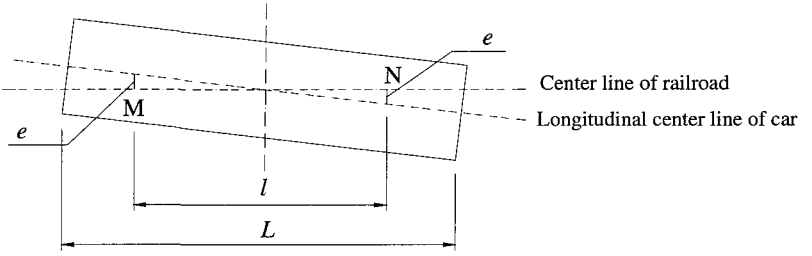


Figure 3. the Most Unfavorable Deviation of Car on Straight Lines

Rolling-swaying vibration offsets of the car loaded with out-of-gauge goods

When the car runs on railroad, the rolling-swaying vibration offsets includes:

(1) offsets caused by track cross-level irregularity;

(2) offsets caused by different diameters of wheel pair;

(3) offsets caused by clearance change of side bearings;

(4) offsets caused by lateral vibrating inertial force and wind force acting on the goods.

The offsets above can be calculated as follows.

Ⅰ. Offsets caused by track cross-level irregularity

Real tracks are not absolutely rigid, straight and even, therefore, there must be some irregularities among which track cross-level irregularity (referring to different height of two tracks) will lead to inclination and lateral offsets of the car. According to Chinese *Regulations on Railway Technical Operation*, the height difference of two tracks cannot exceed 8mm (China's Ministry of Railway 1999). Therefore, the possible maximum offsets of the car at the height of h from top of track will be:

$$y_1 = \frac{8h}{S_0} \text{ (mm)} \tag{6}$$

Among them,

h—height of computed point of goods from top of track, mm;

S_0—distance between rolling circles of wheel pair, taken as 1500mm.

Ⅱ. Offsets caused by different diameters of wheel pair

According to Chinese *Regulations on Workshop Maintenance of Railway Freight Car*, the permitted maximum diameter difference of wheel pair can be 2mm (China's Ministry of Railway 2003). Therefore, the possible maximum offsets of the car at the height of h from top of track will be:

$$y_2 = \frac{h}{S_0} \quad \text{(mm)} \tag{7}$$

III. Offsets caused by clearance change of side bearings

There is some clearance between upper side bearing and lower side bearing in order that the car can run on curves smoothly. According to Chinese *Regulations on Railway Technical Operation*, sum of left side bearings clearance and right side bearings clearance can be 2~20mm (China's Ministry of Railway 1999). If the maximum 20mm is taken, the possible maximum offsets of the car at the height of h from top of track will be:

$$y_3 = \frac{20(h-h_0)}{b_1} \quad \text{(mm)} \tag{8}$$

Among them,

h_0—height of down side bearing from top of track, mm;

b_1—distance between side bearing centers on left side and right side, mm.

IV. Offsets caused by lateral vibrating inertial force and wind force

These forces will give rise to deformation of spring. If the dynamic coefficient of spring is λ, and the spring deflection of loaded car is η, the possible maximum offsets of the car at the height of h from top of track will be:

$$y_4 = \frac{\lambda\eta}{b_2}(h-h_c) \quad \text{(mm)} \tag{9}$$

Among them,

h_c - height of rolling-swaying center from top of track, mm;

b_2 - half distance between spring centers of bogie, mm.

V. Rolling-swaying vibration offsets

The possible maximum offsets caused by rolling-swaying vibration will be:

$$S_r = y_1 + y_2 + y_3 + y_4 = \frac{8h}{S_0} + \frac{h}{S_0} + \frac{20(h-h_0)}{b_1} + \frac{\lambda\eta}{b_2}(h-h_c) \quad \text{(mm)} \tag{10}$$

Determination of minimum safe distance between out-of-gauge goods and structure

In order to secure the safety of car loaded with out-of-gauge goods, enough distance must be kept between out-of-gauge goods and structure clearance when the car runs on railroad. From above, it can be inferred that the safe distance should be:

$$S = S_s + S_r + S_p = \frac{39L}{l} + \frac{9h}{S_0} + \frac{20(h-h_0)}{b_1} + \frac{\lambda\eta}{b_2}(h-h_c) + S_p \quad \text{(mm)} \tag{11}$$

Among them,

S_p—preparatory value for safety, taken as 50mm.

In order to guarantee the safety of transportation, the distance between out-of-gauge goods and structure cannot be less than the minimum safe distance which can be calculated with formula (11).

Example

Among the cars loaded with out-of-gauge goods, flat car type N_{17A} constitutes the majority and it is the major type of flat car at present. The car has Bogie 8A, unfavorable spring stiffness and static loaded deflection. Hence, it is the most unfavorable condition to use N_{17A} to calculate the safe distance.

The relative parameters of N_{17A} car are listed in table 2.

Table 2. Parameters of N17A Car

L(m)	l(m)	h_0(mm)	b_1(mm)	b_2(mm)	λ	η(mm)	h_c (mm)
13	9	750	1520	978	0.408	35.8	464

When the parameters are cited into formula (11), the minimum safe distance with the height of computed point from top of track can be calculate as table 3.

Table 3. Minimum Safe Distance

H(m)	1250	2000	3000	3600	4800	5300
S(mm)	132.1	157.7	191.8	212.2	252.1	270.2

Conclusion

The mathematical models are put forward based on analysis on factors affecting the maximum safe distance between out-of-gauge goods and structure. And the formulae are presented to calculate the safe distance. The models for determination of safe distance will be of meaning to Chinese railway out-of-gauge goods transportation safety.

References

China's Ministry of Railway. (2003). *Regulations on Transportation of Railway Out-of-gauge Goods*. China Railway Press, Beijing, 21-29.

China's Ministry of Railway. (1999). *Regulations on Railway Technical Operation*. China Railway Press, Beijing, 31, 149.

China's Ministry of Railway. (2004). *Regulations on Workshop Repair of Railway Freight Car*. China Railway Press, Beijing, 143-150.

China's Ministry of Railway. (2003). *Regulations on Operation and Maintenance of Railway Freight Car*. China Railway Press, Beijing, 51-53.

Mei Han. (2002). "Evaluation on Safety of out-of-gauge Goods Transport by Railway." In: Ping Huang, Yajun Wang, Shengcai Li. Progress in Safety Science and Technology. Science Press, Beijing/New York, 1029-1033.

Algorithms of the Container Feeding Form between Railway Container Junction Station and Railway Container Transaction Station

Yong YIN[1], Qiyuan PENG[2], Haifeng YAN[3] and Xuecai XU[4]

[1]School of Traffic and Transportation, Southwest Jiaotong University, Chengdu, China; PH: 610031; Tel: 028-87600757; E-mail:yin_yong163@163.com
[2]School of Traffic and Transportation, Southwest Jiaotong University, Chengdu, China; PH: 610031; Tel: 028-87600757; E-mail:qiyuan-peng@263.net
[3]Institute of Transport and Economy, China Academy of Railway Sciences, Beijing, China; PH: 100081; E-mail: yanhaifeng@rails.com.cn
[4]School of Management, Xi'an Jiaotong University, Xi'an, China; PH: 710049; E-mail: xuecai_xjtu@163.com

Abstract

The paper presents three container feeding forms between railway container junction station and its transaction station: Project I —Use usual container train directly to feed into junction station; Project II—Use collecting and distributing train directly to feed into junction station to rearrange; Project III—Use collecting and distributing train to feed into marshalling station to rearrange, and for container wagon suspended, use terminal transfer train to feed into junction station. The operation procedure of each project is analyzed, and the total consumption of each one is calculated. Finally, conclusions are reached by comparing consumption and feasibility with each other: Project I is regarded as the developing orientation of container transportation organization while Project II or III can be used as transition measure under the current situation.

Introduction

According to China railway container transportation development plan, fifteen container junction stations and two hundred container transaction stations will be established within the railway network [1, 2, and 3]. The container flow between junction stations can resort to block container trains to solve transportation organization problem, while containers between junction station and transaction stations around them need to be fed by using suitable trains. Container junction

stations are basically located in main railway transportation terminals, optional train form is more than one for container feeding between junction station and transaction stations. Different container feeding forms have different advantages and disadvantages, which requires concrete analysis and comparison with each project till the optimal container feeding form is chosen.

Analysis of Container Feeding Form

The main train form between transaction stations and adjacent junction station is usual container train or collecting & distributing train. If usual container train is adopted, container flow can access the junction station directly [4, 5]. If collecting & distributing train is adopted, the selection of rearranging location still exists: ①Firstly collecting & distributing train enters into marshalling station to rearrange, and then the rearranged container wagon feeds to the junction station by terminal transfer train. ②Firstly collecting & distributing train enters into junction station to rearrange, and then the usual wagon with the collecting & distributing train feeds to marshalling station by terminal transfer train.

Seen from the above analysis, the container feeding form between transaction station and junction station lies in three possible projects: Project I. Use usual container train directly to feed into junction station and for the usual wagon along the way use collecting & distributing train to feed into marshalling station; Project II. Use collecting & distributing train directly to feed into junction station to rearrange and deliver the usual wagon suspended into the marshalling station with terminal transfer train; Project III. Use collecting & distributing train to feed into marshalling station to rearrange, and for the container wagon suspended use terminal transfer train to feed into the junction station.

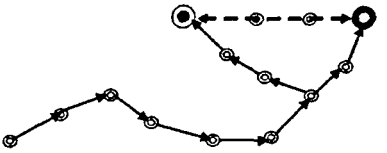
Optimization Selection of Container Feeding Form

The optimal project is the one with fastest container feeding speed and the least train operating cost. The container's consumption per hour can indicate the container feeding speed properly and the train operating cost include the wagon consumption per hour, train energy consumption and operation costs. Therefore, the optimal container feeding form can be chosen by comparing the total consumption with each project.

Marginal Assumption

1. There exists and only exists one junction station and one marshalling station.
2. Whatever operating pathway of the train is determined, the stations along the railway are not repetitive.

In this way, junction station and marshalling station are respectively located in the destinations of two different pathways, and the divergence is only produced at several stations close to the destination (See Fig 1)



Junction station ◉ Marshalling station ◎ Transaction station

Pathway of terminal transfer train

Pathway of usual container train and collecting & distributing train

Fig1. Train pathway sketch

3. For the three projects, the process that the container waits for collecting in the transaction station and the operation after arriving at the junction station, as well as process that usual wagon waits for collecting along the railway stations and the operation after arriving at the marshalling station, are uniform.

So it is supposed that the computation origin consumed by each container per hour is the time in which the farthest station begins to work and the destination is the time in which the container arrives at the junction station; while the computation origin consumed by each usual wagon per hour is time in which the farthest station begins to work and the destination is the time in which the train arrives at the marshalling station.

Related Parameters Definition

The length of operating pathway of usual container train (collecting & distributing train) is L_X, traveling through n_X (n_P) stations. The distance between the j^{th} station and junction station is l_j, container flow is m_{Xj} (usual freight flow is m_{Pj}) and the average operation time is t_{Xj} (t_{Pj}) .The running speed of the train is v, and the pathway from certain station K of junction n'_X (marshalling station n'_P) to the farthest one is matched together, including k ($k \le min\{n_X, n_P\}$) stations. It is obvious that there exists the following relationship:

$$L_{\overline{K}} = L_X - l_{n'_X} = L_P - l_{n'_P} \tag{1}$$

$$M_X = \sum_{j=1}^{n_X} m_{Xj} \tag{2}$$

$$M_P = \sum_{j=1}^{n_P} m_{Pj} \tag{3}$$

$$n = \max\{n_P, n_x\} \tag{4}$$

Consumption Calculation

Project I. The container consumption per hour T_{X1} produced by usual container train includes two parts: interval operating T_{XR1} and transaction station operation T_{XW1}.

$$T_{X1} = T_{XR1} + T_{XW1} = \sum_{j=1}^{n_X} \frac{l_j}{v} m_{Xj} + \sum_{j=1}^{n_X} \sum_{i=1}^{j} t_{Xi} m_{Xj} \tag{5}$$

The wagon consumption per hour T_{P1} produced by collecting& distributing train also includes station operation T_{PW1} and travel operating T_{PR1}, that is to say,

$$T_{P1} = T_{PR1} + T_{PW1} = \sum_{j=1}^{n_P} \frac{l_j}{v} m_{Pj} + \sum_{j=1}^{n_P} \sum_{i=1}^{j} t_{Pi} m_{Pj} \tag{6}$$

The total consumption T_1 of project I is:

$$\begin{aligned} T_1 &= \lambda_X T_{X1} + \lambda_P T_{P1} + E_C \\ &= \sum_{j=1}^{n_P} \lambda_P m_{Pj} \left(\frac{l_j}{v} + \sum_{i=1}^{j} t_{Pi} \right) + \sum_{j=1}^{n_X} \lambda_X m_{Xj} \left(\frac{l_j}{v} + \sum_{i=1}^{j} t_{Pj} \right) + E_C \end{aligned} \tag{7}$$

where λ_X is the weight occupied by container per hour, λ_P is the weight occupied by wagon per hour, and E_C is the collecting & distributing train, usual container train energy consumption and operation cost.

Project II. In like manner, the total consumption T_2 of project II is:

$$T_2 = \sum_{j=1}^{k} [\sum_{i=1}^{j} (t_{Xi} + t_{Pi}) + \frac{l_j}{v}](\lambda_X m_{Xj} + \lambda_P m_{Pj}) + \lambda_X \sum_{j=k}^{n_X} (\sum_{i=1}^{j} t_{Xi} + \frac{l_j}{v}) m_{Xj} + \lambda_P T_{PY} + E_Y \tag{8}$$

where T_{PY} is the wagon consumed by usual wagon taking transfer operation in the junction station and feeding to the marshalling station by the terminal transfer train per hour, and E_Y is the collecting & distributing train, terminal transfer train energy consumption and operation cost.

Project III. In the same way, the total consumption T_3 of project III is:

$$T_3 = \sum_{j=1}^{k} [\sum_{i=1}^{j} (t_{Xi} + t_{Pi}) + \frac{l_j}{v}](\lambda_X m_{Xj} + \lambda_P m_{Pj}) + \lambda_P \sum_{j=k}^{n_P} (\sum_{i=1}^{j} t_{Pi} + \frac{l_j}{v}) m_{Pj} + \lambda_X T_{XY} + E_Y \tag{9}$$

where T_{XY} is the container consumed by container wagon taking transfer operation in the marshalling station and feeding to the junction station by the terminal transfer train per hour.

Comparison of each project

$$\text{Suppose}\sum_{j=1}^{k}[\sum_{i=1}^{j}(t_{Xi}+t_{Pi})+\frac{l_j}{v}](\lambda_X m_{Xj}+\lambda_P m_{Pj})=M \tag{10}$$

Then it can be obtained from (7), (8) and (9) that:

$$T_1=M+\lambda_X\sum_{j=k}^{n_X}(\sum_{i=1}^{j}t_{Xi}+\frac{l_j}{v})m_{Xj}+\lambda_P\sum_{j=k}^{n_P}(\sum_{i=1}^{j}t_{Pi}+\frac{l_j}{v})m_{Pj}+E_C \tag{11}$$

$$T_2=M+\lambda_X\sum_{j=k}^{n_X}(\sum_{i=1}^{j}t_{Xi}+\frac{l_j}{v})m_{Xj}+\lambda_P T_{PY}+E_Y \tag{12}$$

$$T_3=M+\lambda_P\sum_{j=k}^{n_P}(\sum_{i=1}^{j}t_{Pi}+\frac{l_j}{v})m_{Pj}+\lambda_X T_{XY}+E_Y \tag{13}$$

Between Project I and Project II, the following can be achieved from (11) and (12) that:

$$T_1-T_2=\lambda_P\sum_{j=k}^{n_P}(\sum_{i=1}^{j}t_{Pi}+\frac{l_j}{v})m_{Pj}-\lambda_P T_{PY}+E_C-E_Y \tag{14}$$

T_{PY} includes not only the consumption of the wagon of terminal transfer train feeding usual wagon to the marshalling station per hour, but also that of the transfer operation of usual wagon in the junction station. Generally speaking, n'_P within the terminal is smaller, close to the numbers of traveling stations by the terminal transfer train, and the difference between E_C and E_Y is not so large. However, the consumption of the wagon of the transfer operation of usual wagon in the junction station per hour is much more than that of the on-board operation of collecting train traveling through n'_P stations. Hence, T_1 is less than T_2. If the application efficiency of the container is the only factor considered, it is apparently seen that $T_1 \ll T_2$, and in the same way $T_1 \ll T_3$.

Between Project II and Project III the following can be achieved from (12) and (13) that:

$$T_2-T_3=\lambda_X\sum_{j=k}^{n_X}(\sum_{i=1}^{j}t_{Xi}+\frac{l_j}{v})m_{Xj}-\lambda_P\sum_{j=k}^{n_P}(\sum_{i=1}^{j}t_{Pi}+\frac{l_j}{v})m_{Pj}+\lambda_P T_{PY}-\lambda_X T_{XY} \tag{15}$$

Shown from formulation (15), it is hard to compare T_2 with T_3. Because the junction station is mainly concentrated on container operation, the conditions and organization taking the transfer of the terminal transfer train are more difficult than those of the marshalling station, so $T_{PY} < T_{XY}$. If precise comparison is required, the concrete terminal structure and weights between usual freight and container one needs knowing. But on the average, if $n'_X \ll n'_P$, then $T_2 < T_3$, and if the application efficiency of the container is the only factor considered, it is apparently seen that $T_2 < T_3$.

To summarize the analysis above, the consumption of each container per hour of Project I is the least and that of Project III is the largest. Among the three projects, Project I is the best for container transport due to its high container efficiency, but within the district the collecting & distributing train and usual container train would exist at the same time, whose implementation is difficult for the district or terminal with tight capacity [6].

Conclusions

According to the analysis mentioned above, the following conclusions are reached:

1. It is recommended that under permissible conditions the form of usual container train should be employed to feed the container flow between the junction station and transaction stations, and should be regarded as the developing orientation of container transportation organization, which is not only favorable to improve the container efficiency, but convenient to organize container transportation and manage the container;
2. Considered the practical situation of the railway in our country, during the current construction period of container junction station, Project II or III can be used as transition measures because of the ability limitation of the district or terminal.

References

The Second Railway Survey & Design Institute. (2003). The Planning Division under Policy Department of Ministry of Railway, 1-24.

The Second Railway Survey & Design Institute. (2002). The Third Railway Survey & Design Institute, 31-36.

Ministry of Railway, P.R.China. (2001). "10th Five-year" Development Plan of Railway Container Transportation.

The Second Railway Survey & Design Institute, Southwest Jiaotong University, Economics and Planning Research Institute of the Ministry of Railway. (2002). Railway Container Transportation Organization Monographic Study Research Report, Chengdu: Southwest Jiaotong University, 111-115.

YAN Yusong. (2004). "Railway Container Terminal Transportation Organization Research Report ", Chengdu: Southwest Jiaotong University, 31-39.

ZHOU Yong, PENG Qiyuan, and YAN Haifeng. (2003). "Ways of Organizing Container Traffic in a Node Station", China Railway Science, No.6, Vol.24, 124-128.

Evaluation on Route Guidance System and Its Application

Yun Mei-ping, Sun Jian, Yang Xiao-guang,

Yun Meiping, Ph.D Doctor, lector in Tongji University. Interested research field is traffic information and control engineering. Address: Siping road 1239, Shanghai, China, Traffic Engineering Department, Tongji University, 200092. Email: yunmp@mail.tongji.edu.cn, yunmp@126.com. Tel (Fax): +86-21-6598-5368.

Sun Jian, Ph.D Doctor candidate, Tongji University. Interested research field is traffic system analysis. Address: Siping road 1239, Shanghai, China, Traffic Engineering Department, Tongji University, 200092. Email: sunjian@126.com. Tel (Fax): +86-21-6598-1409.

Yang Xiaoguang, Ph.D Doctor, Professor, Tongji University. Main researches are concerned on transportation systemic engineering and theory of ITS. Address: Siping road 1239, Shanghai, China, Traffic Engineering Department, Tongji University, 200092. Email: yangxg@mail.tongji.edu.cn, Tel (Fax): +86-021-6598-8372.

Abstract

Route guidance system (RGS) is widely used as one of main measures of Intelligent Transportation System, which means to maintain an acceptable service level for mobility of persons and goods. But the traffic benefits brought by RGS vary a lot with the characteristics of road networks and traffic flow. Then how to evaluate the traffic benefits is of importance to its implementation. The paper analyzed traffic benefits of RGS. Evaluation Measures of Effectiveness (MOEs) of RGS and its definition models are put forward. An application of the research results is performed with TESS (Tongji transportation nEtwork Simulation System). Comparison of outputs from TESS and VISSIM is provided, which proved the validity of the research. The research is important for evaluation of RGS and integration of RGS and other traffic management and regulation measures.

Key words: Route Guidance System (RGS); Evaluation Measures of Effectiveness (MOEs); Travel time; Simulation

0 Introduction

With the rapid development of social economics, the relationship of supply of the infrastructure and travel demand is becoming unbalanced, which in turn deteriorate the service level of road transportation system. Thereupon, kinds of measures, based on making full use of road infrastructure, have been applied to improve transportation service, in which Route Guidance System (RGS) is widely used in America, Europe and Japan. And then, research on how to evaluate traffic benefits of RGS are of importance to optimizing implementation and extending market penetration, which is focus of this paper.

1. Literature review

Route guidance system integrating technologies such as computer and communication helps driver find one or more routes from a certain origin to a certain destination and provide real-time traffic information. It has been about forty years since route guidance system became the research focus. Most achievements can mainly be included in the following three kinds: optimization of guidance strategy, benefits analysis, and factor analysis that affects guidance benefits.

(1) Researches on guidance strategy

The main function of route guidance is to help traveler select reasonable routes for less travel cost by providing traffic information. Meanwhile, the simple strategies in the selection of information, which designate the shortest route from A to B, could have an adverse effect on traffic flows, which are over-saturated reaction, concentration reaction, and collective reaction (Ben-Akiva et al.1991). In order to solve these phenomenon, a number of approaches have been proposed which can be described as three types. The first type is called iterative (or predictive) and uses complex simulation models to predict the future consequences of the routing actions on network conditions (Ben-Akiva et al., 1997), which are designed to simulate the users reactions to the messages sent and to predict, in real time, the flows in the network according to the information supplied to users. However, it is influenced by the effectiveness of traffic models and requires long computation time. As an alternative, reactive routing strategies, applicable in the urban context, have been developed (Papageorgiou and Pavlis, 1999). They are based on the feedback concept and react to traffic conditions, observed in real time, in order to lead the system towards an objective state. Explicit prediction models of user behavior following control actions are therefore avoided. And decentralized (or distributed) structures have been adopted, so that each controller of the system can operate only on local variables (Gallager, 1977).

The third type is a particular strategy recently proposed which is a decentralized and feedback routing strategy, based on the concept of instantaneous travel time, calculated in real time by measuring traffic. No information on the predicted state of the network (traffic flow on the arcs) or on travel demand (O - D matrix) is required (Francesco Paolo Deflorio, 2003).

The first type is based on predictive traffic information, while the last two types are based on real-time traffic information.

(2) Analysis of guidance benefits

Research on benefits of route guidance system appeared with the application of RGS and its market penetration. Early research by TRRL(Transport and Road Research Laboratory) proved that it reduces vehicle mile by 8%~10% if drivers comply with the shortest route information. Levinson, David. (2003) found that travel time saving from guidance ranges from 2.7% to 55% according to relevant research, and proved that time saving is more extinct during incidents by using stochastic queuing model. Karl. E. Wunderlich et al. (2001), using experiments, also proved that the benefits are mainly on user's time schedule, which means that travel reliability is improved obviously. By using simulation, Yang (1998) proved that time saving from guidance strategy based on real-time information is more effective, while as to solving over-saturated reaction strategies based on predictive information is more useful.

(3) Factors that influence guidance benefits

Aldeek et al. (1991) revealed that during recurrent traffic congestion benefits from guidance is very limited if only a few vehicles are equipped with guidance system. Huang et al. (2002) believed that it is not certain that individual travel cost can be reduced by providing traffic information, which are dependent on the factual traffic situation on road networks and capacity of each route. And sensitivity analysis among market penetration, update frequency, flow prediction, and over-saturated reaction has been performed by Kaysi(1992), Yun (2004) discussed further about the influence of road networks capacity to guidance benefits. All these researches are helpful in understanding benefits of route guidance.

(4) Summary

From above, as to route guidance system, information providing and traveler behavior both are important, which in turn influence effect of route guidance. So as to evaluation of route guidance system, how driver react to information directly influence the system's effect. Moreover, research results from simulation in other countries can hardly describe the fact in most cities in China, which is mainly due to two kinds of reasons listed as following.

① Characteristics of traffic systems are different

Different with motorized cities, most cities in China are during a motoring process, in which travel behavior, traffic flow, and even road networks all present different characteristics.

② Travel behavior varies in different areas

Discretionary activities and commuting patterns differ in each city, so results from experiments and simulations in other cities abroad are not fit for the fact in China, which brings the necessity to evaluate the benefits of RGS.

③ Measures of evaluation are not fully considered

Questions such as what MOEs are reasonable and if the MOEs can completely show the whole benefits are not deeply studied yet.

2 Definition of evaluation measures

2.1 Selection of evaluation measures

Selection of MOE for RGS should consider the following rules (Yun, 2004):

(1) traffic benefits of RGS not only include travel time saving, but also time schedule reliability improvement that can reflect travelers' perception to the service of traffic system.

Karl E. Wunderlich analyzed real time reliability of RGS and drew the conclusion that traveler can get more benefits when reliability and prediction is improved.

(2) benefits from RGS include not only travel time saving from travelers, but also improvement of networks efficiency for administrators.

(3) As to traffic benefits, both guidance users and non-guidance users should be considered.

2.2 Determination of MOEs

With consideration of above three rules, MOEs of RGS benefits are listed in four kinds, efficiency, safety, comfort, and environment, which are shown in table 1. The MOEs are from the view of travellers and the whole networks.

Table 1. MOEs of RGS benefits

Types \ Subjects / MOEs	Travelers	Networks
Efficiency	(1) average travel time	(2) total travel time
	(3) average delay	(4) total delay
	(5) average velocity	(6) network velocity
	——	(7) network saturation
Safety	——	(8) death rate
	——	(9)incident delay (damage)
Comfort	(10) reliability	——
	(11) punctuality	——
	(12) information reliability	(12) information reliability
Environment	——	(13) fuel consume
	——	(14) pollution exhaust
	——	(15) noise

Take average travel time T as an example, T includes travel time of road link (T'_l) and intersection delay (d_k), then $T = T'_l + d_k$.

Average travel velocity (V) can be described as:

$V = \frac{V_f}{2} + \sqrt{\frac{V_f}{4} - \frac{V_f}{K_j}Q}$. Here V_f is velocity under free flow, K_j is maxim density of road links. Q is traffic flow. Then $T'_l = \frac{L_l}{V}$. Here L_l is length of road link l. As to intersection delay (d_k), it can be formulated as $d_k = \frac{r_k^2}{2c_k} \cdot \frac{S_l}{S_l - Q_l}$. Here d_k is intersection delay, r_k is red time, c_k is cycle time, S_l is saturated flow of road link, Q_l is factual flow of road link.

Then, travel time t_{ij} from i to j, in which there are m road links and n intersections included, is

$$t_{ij} = \sum_{l=1}^{m} T'_l + \sum_{k=1}^{n} d_k$$

Average travel time (T) is weighted average travel time of vehicle flow, formulated as:

$$T = \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{m} Q_{ij}} \times \sum_{i=1}^{n}\sum_{j=1}^{m} t_{ij} \times Q_{ij}$$

. Here Qij is traffic flow from i to j. Detailed determination of each MOE can refer to the relevant research (Yun, 2004).

3 Model validation based on TESS and VISSIM

In order to validate the applicability of determination of MOEs, TESS (Tongji transportation nEtwork Simulation System) is applied here which is a time space discrete micro-simulation that means to analyze urban arterial networks with consideration of information provision and above models are used in TESS. Further research on TESS is still going on in our research team (Sun, 2005). At the same time, micro-simulation VISSIM is applied, which is a popular microscopic simulation model produced by PTV in German. The main output of TESS and VISSIM selected here includes travel time, speed, queue, and delay. At the same time, Travel time of RGS users and non-users are selected here and comparison of outputs by TESS and VISSIM is performed. The test network is in Hangzhou shown in figure 1, and comparison by VISSIM and TESS is shown in figure2. The results from TESS and VISSIM are closely correlative, which reveals that TESS is also a useful tool in the analysis.

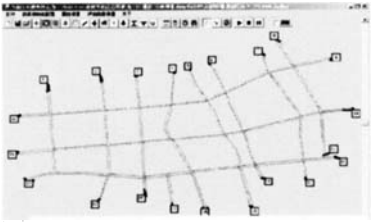


Figure1. Test network in the city of Hangzhou

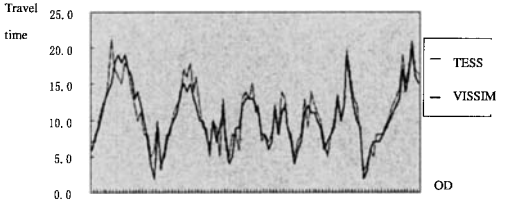


Figure 2. Travel time comparison of OD pairs

Car and bus are both considered in the test with 90% and 10% respectively, which both include RGS users and non-users (TRaditional, TR) accounting for 60% and 40 apart. RGS users are described as Pre-trip information Servicing (PS) and on-Board Routing (BR). The rates of TR, PS, and BR are 2:5:3. Average travel time of the above three users, in condition that 27 OD pairs are considered, is shown in figure 3. The outputs show that usually velocity of BR users is higher than PS users, and velocity of non-RGS-users is the lowest. However, velocity of some OD pairs turns out different conclusions, that probably due to the nonproficiency of simulation model.

4 Conclusions

To identify traffic benefits of RGS, measures of effectives and its determination models are presented in consideration of four types of guidance users. Then VISSIM and TESS are performed to validate the models, and comparison of outputs by VISSIM and TESS is provided that confirms the research results.

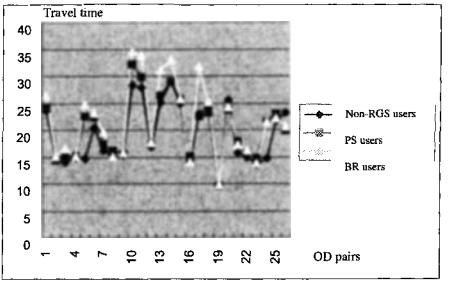


Figure 3. Average travel time of different kinds of users by VISSIM

References

A. Kanafani and H. Aldeek, 1991, A simple model for route guidance benefits, Transpn. Res. 25B, 191-201.

Ben-Akiva, M., A. De Palma, and I. Kaysi, 1991, Dynamic network models and driver information systems. Transpn Res. 25A, 251-266.

Francesco Paolo Deflorio, 2003, Evaluation of a dynamic route guidance atrategy, DITIC - Polytechnic of Turin, Italy.

Gallager R. G. 1977, A minimum delay routing algorithm using distributed computation. IEEE Transactions on Communications, vol. 25, No1, 73-84.

HUANG Hai-jun, WU Wen-xiang, 2002, Models for evaluating impacts by travel information systems on travel behavior. System engineering, 10, 81-83.

karl E. Wunderlich, Matthew H. Hardy, james J. Larkin, Vaishali P. Shah, 2001, On-time reliability impacts of Advanced Traveller Information Services (ATIS): Washington, Dc Case study.

Kaysi, I., 1992, Frameworks and models for the provision of real-time driver information, Ph.D. thesis, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge MA.

Levinson David, 2003, The value of advanced traveler information systems for route choice, Transpn Res. -C, 75-87.

Papageorgiou M. and Pavlis Y. 1999, Simple decentralised feedback strategies for route guidance in traffic networks. Transportation Science, Vol. 33, No. 3, 264-278.

Sun Jian, Yang Xiaoguang, 2005, A Simulation-based approach to evaluate dynamic multiple user route choice behavior under ATIS, the 5th international conference on traffic and transportation academy, Xi'an.

Yang, H. 1998, Multiple equilibrium behaviors and advanced traveler information systems with endogenous market penetration. Transpn Res. 32B, 205-218.

Yun Meiping, 2004, Study on evaluation methods of traffic benefits for Advanced Traffic Management Systems (ATMS), Ph.D. thesis, Tongji University, Shanghai, China.

Portable Non-Intrusive Traffic Data Collection in Variable Roadside Environments

Lloyd J. French, Ph.D., P.E.[1] and Millie S. French, P.E.[1]

[1]French Engineering, LLC, 114 Cooley Estate Road, Smithfield, PA 15478; PH (724) 564-2121; FAX (724) 564-0924; email: jfrench@frenchengr.com, mfrench@frenchengr.com

Abstract

Collecting traffic counts on high volume and high speed highways is particularly troublesome. Even in the early morning hours, placing road tubes or other devices in the road can be extremely hazardous to the field technicians and motoring public. To this end, Portable Non-Intrusive Traffic Data Collection Systems have been developed through extensive research in states including Minnesota, Virginia, New York, Wisconsin, and Kentucky. The purpose of this past research was to develop a traffic counting system that did not require entry into the traveled way. The solutions developed included mounting microwave or acoustic sensors on temporary poles attached to existing roadside signs or trailer-mounted telescoping poles. The testing grounds for these devices have typically been freeways, which had ample roadside space, favorable roadside topography, and were free of utilities. This permitted the devices to generally be mounted at the ideal location according to the manufacturers' specifications. However, the counting performance of the sensors is affected by the height and distance from traffic at which the sensors are placed. Placing the sensors at the recommended heights and offsets can be difficult in more restricted roadside environments. For example, the acoustic sensors are optimally placed high above the traffic (25-ft to 30-ft, 7.6-m to 9.1-m) and close to the traffic stream (10-ft to 20-ft, 3.0-m to 6.1-m), while microwave sensors are optimally mounted lower (15-ft to 25-ft, 4.6-m to 7.6-m) and further from the traffic stream (25-ft to 35-ft, 7.6-m to 10.7-m). In non-freeway environments, limited right-of-way, adjacent property development, and overhead utilities can complicate equipment set up. In addition, because the sensors will tend to be located closer to the lanes, roadside safety and capacity restriction become issues. Finally, in non-freeway environments, pedestrian / security issues also become important. The purpose of this research was to test the portable non-intrusive traffic data collection systems in a variety of environments, including some restrictive non-freeway environments. This research demonstrates the sensitivity of the equipment in matching manual traffic counts and vehicle classification with various roadside restrictions.

Introduction

The purpose of this research was to conduct a field test of four non-intrusive traffic data collection sensors in a temporary setup. As specified by the Pennsylvania Department of Transportation (PennDOT), the research sponsor, the four non-intrusive traffic counters tested included one acoustic sensor, two microwave sensors, and one infrared sensor. All sensors

were placed in the roadside and operated from a "sidefire" position. A brief description of each type follows:

Acoustic Sensor – The sensor used was a passive acoustic sensor, meaning that that it only "listened" and did not emit any sound. The sensor produced an acoustic image of the travel lanes with an array of microphones and related hardware on its compact, 96 sq in (620 sq cm) face. Vehicles were identified through the acoustic signals they emitted, and classified according to the calculated length. This sensor was mounted overhead, and needed to be located on only one side of the roadway.

Microwave Sensors – The sensors used emitted a microwave signal toward the travel lanes that was reflected back to the sensor. Based on the return time to the sensor, the distance from the sensor to the target could be determined. Signal reflections from background targets were identified and filtered out, leaving only reflected signals from the travel lanes. Vehicles were identified based on certain characteristics of the reflected signal, and classified according to the calculated length. These sensors were compact, measuring approximately 12-inches (32 cm) or less along their longest dimension, and weighing approximately 5 lbs (2.3 kg). These sensors were mounted overhead, and needed to be located on only one side of the roadway. Two microwave sensors were used in this research.

Infrared – This sensor consisted of two units, one of which was mounted on either side of the roadway. Unlike the acoustic and microwave sensors, which were mounted overhead and designed to identify vehicles, this sensor was mounted near the road surface and was designed to monitor axles. A series of parallel and intersecting infrared beams were projected across the travel way. As vehicles passed through the sensor area, the beams were broken and reconnected. An algorithm that monitored and processed the sequence and timing of the beam activity permitted the sensor to identify vehicles in the traffic stream and classify them according to the number and spacing of axles.

Field Data Collection

Four sites were selected in the Uniontown, Pennsylvania area, as follows:

Site #1:	S.R. 0119, Segment 0470 (Northbound Only): Freeway, Single Direction, Short-Term In-Pavement Site (STIP)
Site #2:	S.R. 0040, Segment 0160 (Both Directions): Two-Lane Highway, Automated Traffic Recorder (ATR) Site
Site #3:	S.R. 0119, Segment 0470 & 0471 (Both Directions): Freeway, Both Directions, STIP Site
Site #4	S.R. 0040, Segment 0260 & 0261 (Both Directions): Five-Lane Suburban Arterial

The primary objective of the site selection was to provide a cross-section of roadside environments for equipment setup. A secondary objective was to select sites near in-pavement traffic counting stations operated by PennDOT. PennDOT operates in-pavement counting stations that are both short-term (STIP), and permanent (ATR).

Note that Site #1 and Site #3 are in the same location, however, Site #1 testing included only the northbound lanes, while Site #3 testing included the monitoring of both directions with a single sensor.

The test was conducted over the course of two days, September 14 and 15, 2005. The equipment was set up at each site by vendor representatives. Vendors were given approximately two hours to set up, and data collection at each site was four hours in duration. Sunrise was at approximately 7:00 a.m., therefore it was dark during set up and during the first hour of data collection at Sites #1 and #3.

Figure 1 - Temporary Pole

A temporary pole similar to that used in the Minnesota Guidestar research (Kotzenmacher, et al, 2005) was constructed by the researchers for the purpose of mounting a sensor during the field testing. This pole was 26 feet (7.9 m) in height and was designed to be attached to an I-beam traffic sign support at Site #1. The pole used in the testing is shown attached to the sign post in Figure 1. The primary difference between this pole and that used in the Minnesota Guidestar research was that it was two feet (0.6 m) higher and much lighter in weight. In the photo in Figure 1, a five-pound weight is attached to the pole in lieu of a sensor.

Light towers with the lights removed were provided to the vendors of the acoustic and microwave sensors for mounting and elevating their sensors. The lights were removed as a matter of convenience so that ample room would be available to mount the sensors. These were Nighthawk LT 12 units manufactured by Multiquip. According to the manufacturer's specifications, the pole contained on these units telescoped to a height of approximately 30 feet (9.1 m). Each vendor was provided with their own light tower unit, which was placed and leveled with the assistance of the researcher. Each vendor then banded their sensor onto the light tower pole and raised the pole to the desired height. Vendors of the infrared sensor set up on tripods that they supplied.

The data were collected in 15-minute intervals and, depending on the sensor, classified according to vehicle length or the FHWA Scheme F classification scheme. The length-based classification scheme was as follows: 0 to 22-ft (passenger vehicles), 22 to 40-ft (single unit trucks), and longer than 40-ft (multi-unit trucks). The microwave and acoustic sensors used the length-based classification scheme. The infrared sensor collected data according to the FHWA scheme.

For comparison to the vendor data, the research team had two teams classifying and counting traffic in the field. The first team counted and classified traffic according to the length-based scheme. The second team counted and classified traffic according to the FHWA scheme.

Photographs of the equipment setups at Sites #2, #3 and #4 are shown in Figures 2, 3, and 4. As can be seen, utility conflicts and proximity to the travel lanes were issues at several of the sites.

Figure 2 - Site #2 S.R. 0040 Two-Lane

Figure 3 - Site #3 S.R. 0119 Freeway

Figure 4 - Site #4 S.R. 0040 Five-Lane

Data Analysis

The results of the comparison of the manual counts to the data from the non-intrusive traffic data collection equipment and PennDOT counters are provided in this section. For each counter, the Absolute Percent Difference (APD) between the counter and the manual count was computed for (1) the total volume (unclassified) for each 15-minute period (2) the total (unclassified) four-hour volume. The formula for APD is as shown in Eq. 1.

$$APD = [|Vs - Vm| / Vm] \times 100\% \quad (1)$$

where:

APD = Absolute Percent Difference (%)
Vs = Volume from the sensor or ATR / STIP (vehicles)
Vm = Volume from the manual count (vehicles)

Results

The following table provides a summary of the results for each sensor. APD was computed for the total volume for each 15-minute interval and averaged across the 16 time periods in each four-hour test. These data are presented in the columns labeled "APD %." It was also computed for the total volume in the four-hour test period. These data are presented in the columns labeled "4-Hr %." As can be seen, APD varied significantly from site-to-site for all of the sensors except Microwave Sensor #1. This highlights some of the sensitivity of these devices to their position relative the traffic stream. The threshold for acceptable performance was arbitrarily selected as matching the manual counts within 5%. As can be

seen, all of the sensors were within this limit at at least one of the sites, leading to the conclusion that all of the sensors are acceptable if used in the appropriate circumstances.

Table 1- Summary of Field Testing

Site # (Dir)	Microwave Sensor #1		Acoustic Sensor		Infrared Sensor		Microwave Sensor #2	
	APD %	4-Hr %	APD %	4-Hr %	APD %	4-Hr %	APD %	4-Hr %
1 (NB)	6%	5.8%	1%	1.0%	1%	0.2%	6%	4.1%
2 (EB)	4%	3.7%	2%	0.7%	1%	0.0%	25%	24.9%
2 (WB)	2%	0.5%	20%	20.4%	1%	0.2%	3%	1.3%
3 (NB)	2%	1.4%	1%	0.8%	19%	18.5%	2%	1.5%
3 (SB)	2%	0.9%	3%	1.0%	26%	26.0%	2%	0.1%
4 (EB)	2%	1.2%	3%	1.1%	25%	25.7%	3%	0.5%
4 (WB)	3%	1.1%	5%	5.0%	15%	14.9%	43%	40.6%
Average	**3%**	**2.1%**	**5%**	**4.3%**	**13%**	**12.2%**	**12%**	**10.4%**

Truck classification was assessed by comparing the percentage of trucks in each class predicted by the sensor against those from the manual counts. There are no concise numerical summaries that convey truck classification performance of the sensors. In general, the infrared sensor performed exceptional at identifying trucks in the traffic stream, even at the four- and five-lane sites where its total volume differed significantly from the manual counts. The overhead sensors in general did not match the manual classifications closely. Overall, Microwave Sensor #2 matched the manual truck classification the closest of the overhead sensors.

With respect to the PennDOT in-pavement counters, the ATR at S.R. 0040 and the STIP at S.R. 0119 provided counts that consistently matched the manual counts. The highest mean APD for either was 4.9%, which was encountered at S.R. 0040 two-lane section (Site #2), where some error was expected due to the distance between the ATR and the test site. Truck classification data were not provided for the STIP, however, at the ATR, only the infrared sensor was in closer agreement with the manually-determined traffic composition.

Other Key Issues

Roadside presence and safety is an important issue that must be considered during deployment of these devices. Both a temporary pole and light tower were used for the sensors, while the infrared sensor was set up on the manufacturer-supplied tripods. Each is discussed below.

The temporary pole had a minimal presence in the roadside, however, it could have reached the traveling lanes from where it was set up had it fallen towards the road. It was also quite heavy to lift into place, even though it was constructed of lighter materials than those developed in the Minnesota Guidestar research (Kotzenmacher, et al, 2005). The potential for one person to lift the pole into place while maintaining full control over it to minimize the

potential of it falling into traffic is questionable. In addition, the temporary pole installation is likely not suitable for smaller sign posts and those not securely planted in the ground because of its size and weight.

The light towers were much easier to control and their potential to fall into traffic was less, assuming they were properly leveled and the outriggers were extended. According to the manufacturer specifications, the particular light towers that were used are stable in gusty winds up to 65 mph. However, these took up significant space in the roadside. They could not be set up within the available right-of-way at the S.R. 0040 two-lane test site (Site #2) and consumed the entire 8-foot shoulder at the S.R. 0040 five-lane test site (Site #4). At a weight of over 0.75 tons, they also present a formidable roadside hazard if struck by a vehicle. However, they may be appropriate for the freeway environment if the roadside has sufficient cross-section to place them outside the clear zone, or they can be protected from traffic.

The tripods used to mount the infrared sensor are likely not appropriate for most installations because they leave the device open to vandalism and tampering. It was encouraging that the infrared sensor did not have to occupy the shoulder during the testing at the S.R. 0119 sites, however, the roadside was lower in elevation than the traveling lanes and shoulder. The vendor noted enclosures that they provide that might be used to secure and protect the sensor during temporary installations.

Conclusion

These technologies were assessed for PennDOT, and there appears to be good potential for using them in Pennsylvania. Any DOT considering non-intrusive traffic data collection should make use of all three technologies since each has circumstances in which it is likely to perform the best. Acoustic sensors might be most appropriate in areas where an overhead sensor is needed but the right-of-way is limited and overhead utilities are not an issue. Microwave sensors might be most appropriate in areas where the right-of-way is not an issue and it is desired to move the sensor as far as possible from the traveling lanes. This is particularly evident with the performance of Microwave Sensor #2 at Site #4, which had the least amount of roadside space. The sensor performed poorly in monitoring the nearside (WB) lanes, but quite well with the far (EB) lanes. The infrared sensors might be most appropriate in instances where truck classification is important and roadway geometry allows the device to be set up to manufacturer's specifications. This device, which detects axle configuration and spacing, matched the manual truck counts very closely—to within a vehicle or two at some of the two-lane sites. However, as can be seen at Sites #3 and Sites #4, wide cross-sections resulting in high crowns interfered with the sensor's operation.

References

Kotzenmacher, J., Minge, E., and Hao, B. (2005). *Evaluation of Portable Non-Intrusive Traffic Detection System*, Minnesota Department of Transportation, St. Paul, Minnesota.

Acknowledgement

The authors would like to acknowledge the Pennsylvania Department of Transportation for sponsoring this research.

Visualization of Bus Schedule Adherence Using GIS

Dan Yu[1], Santosh Mishra[2], Jie Lin[3]

[1] Department of Civil and Materials Engineering, University of Illinois at Chicago, 842 W. Taylor St. (MC246), Chicago, IL 60607, Dyu6@uic.edu
[2] TranSystems Corporation, One Cabot Road, Medford, MA 02155, 781-396-7775 ext 30266 (phone), 781-396-7757 (fax), skmishra@transystems.com
[3] Corresponding author. Department of Civil and Materials Engineering & Institute of Environmental Science and Policy, University of Illinois at Chicago, 842 W. Taylor St. (MC246), Chicago, IL 60607, 312-996-3068 (phone), 312-996-2426 (fax), janelin@uic.edu

ABSTRACT

Although public transportation has been using intelligent transportation infrastructure since the 1960s, recent advancements in technology and an increased agency focus to the Americans with Disability Act are benefiting the transit industry. Advanced Public Transportation System (APTS) includes Fleet Management Systems, Traveler Information Systems and Electronic Payment Systems. APTS utilizes advanced technologies ranging from Automatic Vehicle Location (AVL)/Computer Aided Dispatch (CAD) System, Automated Passenger Counter (APC) to Automated Fare Collection (AFC), Automated Vehicle Component Monitoring (AVM), on-board and wayside security systems, multi-modal coordination, and Transit Traveler Information (TTI).

The long-term goal of collecting AVL and APC data is to use these data to improve transit Level of Service (LOS). LOS includes service reliability measured by both on-time performance and regularity of headways. In this paper we present a computer visualization prototype, which maps static bus schedules and can be extended to integrate "real time" active APC/AVL data measure the on-time performance and headway adherence. This is accomplished using ArcObject in ArcGIS and VBA programming language. Using this visualization tool, the difference between active and static data of a trip, or between active data from two or more continuous trips, can be readily measured and compared on a dynamic map, which updates the active/schedule location of one or more buses given a specified frequency of a time frame (i.e., a snapshot in time). We demonstrate how the mapping is accomplished using the Chicago Transit Authority (CTA) bus route schedules.

Such a visualization tool is not only useful for transit planning and LOS measures but also beneficial to transit users if such a tool is made available online to the public. Travelers can gain real-time, accurate information about where the buses are and when they arrive at particular stops. Better utilization of APC/AVL data leads to better understanding of bus performance and thus results in better planning and operation management. The powerful functions of ArcGIS in visualizing spatial data and Oracle's database management capability will provides very strong IT support for transit data utilization. Lastly, future research directions with respect to full utilization of APC/AVL data are also identified in this paper.

INTRODUCTION

The public transportation industry has deployed technology in the past to improve transit operations and management. Latest advancements in the area of vehicle tracking, sensors, communication systems, database management and other relevant technologies have made transit systems more intelligent than earlier. The advanced public transportation system (APTS) utilizes advanced technology ranging from Automatic Vehicle Location (AVL)/Computer Aided Dispatch (CAD) System, Automatic Passenger Counter (APC) to

Automatic Fare Collection (AFC), Automatic Vehicle Component Monitoring (AVM), on-board and wayside security systems, multi-modal coordination, and Transit Traveler Information (TTI) (U.S. DOT, 2005)

Automatic Vehicle Location Technology helps in tracking vehicles in real time. Tracking media have changed from traditional radio systems to global positioning systems (GPS) based advanced systems along with development of new technologies. Older systems used signposts and odometer reading as location referencing methods. New GPS based systems make use of geostationary satellite based information for location determination. GPS could be combined with local positioning systems (LPS) to remove errors created due to building obstructions downtown (also known as "urban canyon" barrier). The primary purpose of Automatic Passenger Counting (APC) systems is to record boarding and alighting at different stops. Such information can be really useful for offline analysis e.g. demand estimation, load profile determination etc. Traditional APC systems include sensor and analyzer. Analyzer converts signals from sensor into passenger counts. New systems are integrated with AVL and data can also be sent as an event record through the radio channels. AVL/APC is combined with new computer aided dispatch information to communicate with control center.

The long-term goal of collecting AVL and APC data is to use these data to improve transit Level of Service (LOS). LOS includes service reliability measured by both on-time performance and regularity of headways, defined as the following:

On-time performance: is simply measured as the percent of trips that arrive and/or depart on (or off) schedule. In other words, it is to see how well the active data collected from an operational bus matches its schedule.

Headway regularity: for services operated with headways of ten minutes or less, the TCQSM (Kittelson & Associates et al., 2003) states that headway adherence should be used as the determinant of reliability. Headway adherence is to see the bus operation of scheduled sequential buses, which can be measured by the active data collected from two or more sequential operational buses.

Furth et al (2003) summarized transit ITS data usage including both static and active data into six categories: transit management, traveler information, safety and security, emergency management, maintenance, and reporting/analysis needs. Reporting and analysis are two of the most significant application areas of transit ITS data. Reporting or analysis can be performed in both graphical and non-graphical formats. Data visualization software need database to be organized properly to avoid any performance overhead in running queries. Existence of GIS route and stop inventory has established added needs to link all transit data with geographical information for spatial visualization. Spatial representation of schedule adherence gives a better picture of events on the road as compared to a mere space-time diagram representation of the same.

In this paper we present a GIS-based visualization prototype that maps static bus schedules and can be extended to integrate "real time" active APC/AVL data measure the on-time performance and headway adherence. Using this visualization tool, the difference between active and static data of a trip, or between active data from two or more continuous trips, can be readily measured and compared on a dynamic map, which updates the active/schedule location of one or more buses given a specified frequency of a time frame (i.e., a snapshot in time). We demonstrate how the mapping is accomplished using the Chicago Transit Authority (CTA) bus route schedules.

STATIC BUS SCHEDULE AND ACTIVE APC/AVL DATA

Figure 1 is a simple illustration of how the route/scheduling data (i.e., static data) and the

AVL/APC data (i.e., active data) flow within CTA. HASTUS and BusTool, the scheduling and inventory system respectively, generates schedule and route data. All CTA buses are performed based on the route and stop information in the distributed schedule. On the other hand, the AVL/APC equipped buses collect and store active field data on the on-board computers, which is then transferred to bus-link sever when buses are close to garage wireless network (during bus pull-in and pull-out). Data from bus-link server is eventually ingested into AVL database server and stored in ORACLE database format at the CTA data center. The active operation data together with the static schedule data provide feedbacks to static schedule data and examination of bus performance in the field.

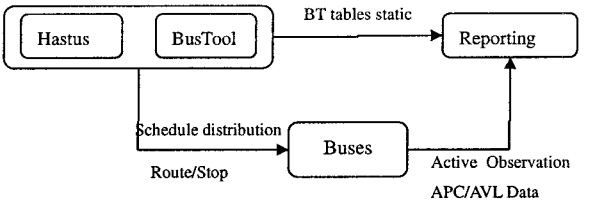


Figure 1: Simplified Data Flow in CTA bus operation

VISUALIZATION SYSTEM IMPLEMENTATION

The static dataset provides the time a bus passes a time point and a bus-stop. Time points are used as service management points while stops are the locations interactive with passengers. Thus, stop-based schedule visualization is practically more useful and needed for on-time performance and headway adherence measurement. CTA bus stops and routes information is available in digital geographic maps (in GIS shape-file format). Passing time at each stop (stored in Oracle Database Management System (DBMS)) can to be added into the map as a schedule-related attribute. Then the map is updated, through user-specified time frames, according to the recorded bus passing time. Only stops a bus passes at a particular time within a pre-defined time window will be selected and shown on the map. The map can be updated at a pre-specified or user requested frequency.

The prototype is built upon ArcObject in ArcGIS and Visual Basic for Application (VBA) programming language. ArcObject and VBA, the built-in ArcGIS Application Programming Interface (API) functions, allow direct connection to Oracle DBMS and are available to users in the standard ArcGIS package. It is also compatible with ArcIMS and ArcSDE server, which can be integrated with the GIS geodatabase and network server. End users are able to see a dynamic map visualizing the movement of a bus of interest through a specified time window at user's request. The prototype shows the schedule for up to any two trips from two different routes within a user-specified time period, specified through an input window. The program is flexible enough to be extended to multiple routes/trips. Figure 2 presents the detailed system structure of the prototype.

User Input Window

Figure 3 is the user input window to specify the trips and time period of interest. Users input the trip numbers in TripNo1 and/or TripNo2 textboxes. "Start time" and "End time" of the time period of interest is also specified. The time period should be within the trip operation time. For example, Figure 3 shows that a user wants to see the schedule for trips: 11132361 (Route 60 Weekday morning schedule) and 11543951 (Route 8 Weekday morning schedule), between 05:50am and 06:20am, when both trips are in operation.

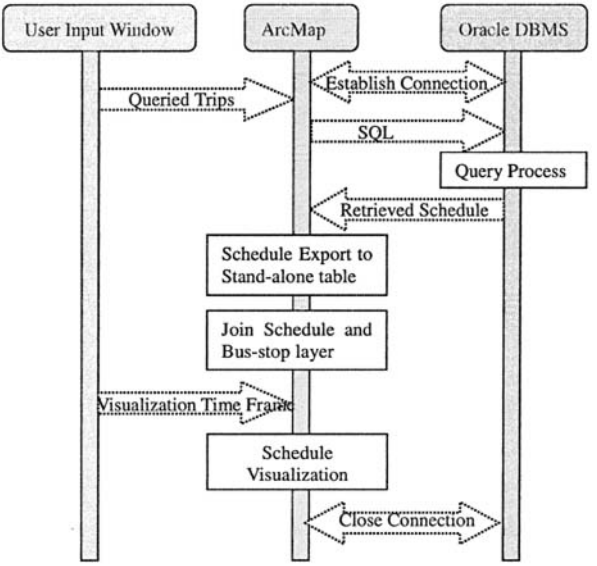


Figure 2: System structure of the visualization prototype

Figure 3: Example User Input Window

ArcMap Query of Schedule Data in Oracle Database

The stop-based bus schedule is stored in Oracle DBMS, containing the times the bus passes the stops along a trip, and trip pattern or route. The Oracle DBMS is connected through ArcMap application; user inputs (e.g., tripNO) are passed to Oracle; SQL queries are run in Oracle Database to retrieve the schedule of interested group and send back the result schedule as a stand-alone table to ArcMap application. The resulting table contains stop-based schedules for the requested trips. MicroSoft OLE Database reference package provides the API functions to connect with Oracle Database and run SQL queries. A user

simply clicks "Query from Oracle" button after entering the trip numbers of interest to send his/her query, and then "Export to Table" to get bus trip schedules.

Join Schedule Table to Bus-stop Feature layer

The resulting schedule table contains stop-based schedule for the trips of interest. The table is then joined to the bus-stop feature layer to link stop passing time to stop object (i.e., a bus) in the feature layer. After the "join" step, only those bus-stops that have corresponding schedule records in the retrieved schedule table are displayed on the map, and all other "uninterested" stops are invisible. Figure 4 show the bus stop maps before and after the "Join" function was performed. A click on button "Join to Layer" on the user input window shown in Figure 3 calls the join procedure and results in a joined feature layer and attribute table.

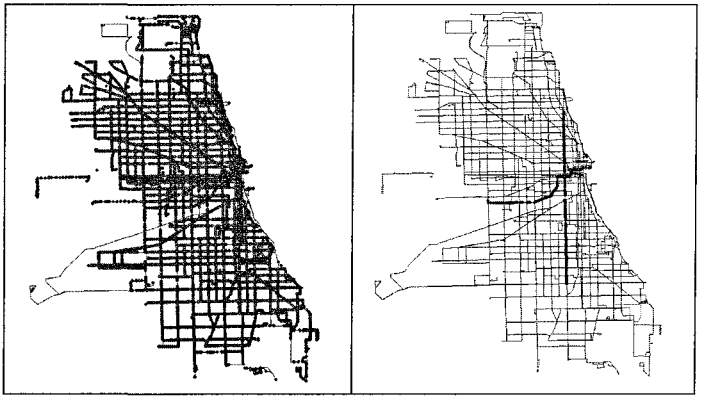


Figure 4: All and queried CTA bus stops in Chicago (before and after "Join to layer")

Visualization of Bus Schedule

After retrieving bus stops of interest from the feature class, user may specify a valid time window of interest within the bus schedule. In this example, trips 11132361 and 11543951 are the weekday early morning trips between 05:30 and 06:40 for route 60 and 8 respectively. So the valid start and end times should be within that time period.

With the user-specified study time period and the pre-defined map updating frequency (typically, 60 or 120 seconds), the schedules for the study trips can be divided into a sequence of time frames. In each time frame, the program selects the stops with scheduled stop-passing time closest to the time frame, shows them on the map, and increases the time frame by unit frequency (60 seconds in the demo program). The time frame is initialized as user-specified start time and terminated when reaching user-specified end time. The program automatically updates the map to visualize the bus schedule from frame to frame. Figure 5 demonstrates the bus locations of the two queried trips at two different time frames: 06:05am and 06:25am. Button "Schedule Visualization" in Figure 3 is designed for this process. It is worth pointing out that the pre-defined (or fixed) updating frequency can be easily modified to user-specified (or flexible) frequency from user input window in future version.

CONCLUSION AND FUTURE WORK

This paper presents a GIS-based prototype to map static bus schedules and can easily be extended to integrate active APC/AVL data to measure on-time performance and headway

adherence, since the system framework is independent of the database structure. Such a visualization tool is not only useful for transit planning and LOS measures but also beneficial to transit users if such a tool is made available online to the public. Travelers can gain real-time, accurate information about where the buses are and when they arrive at particular stops. Better utilization of APC/AVL data leads to better understanding of bus performance and thus results in better planning and operation management. The powerful functions of ArcGIS in visualizing spatial data and Oracle's database management capability will provides very strong IT support for transit data utilization.

There are several future research directions based on the preliminary result of the project. These include improving and modifying the prototype to incorporate bus schedules for multiple trips, routes or RUNs to satisfy various user requests, incorporating the active APC/AVL data, and studying the CTA bus distribution at a particular time to better manage the CTA bus cooperation.

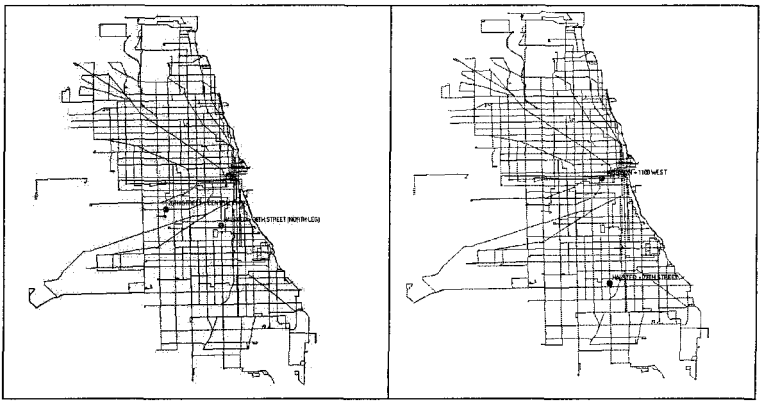


Figure 5: Schedule map at 06:05am and 06:25am.

ACKNOWLEDGEMENT

This research was funded under the MIT-UTC-CTA transit improvement project. We would like to thank CTA and Urban Transportation Center at University of Illinois at Chicago for their support.

REFERENCE

Furth P., Hemily B., Muller H., Strathman J. *Uses of Archived AVL-APC Data to Improve Transit Performance and Management: Review and Potential. Transit Cooperative Research Program, Project H-28*, Transportation Research Board, Washington DC, 2003, available at http://trb.org/publications/tcrp/tcrp_webdoc_23.pdf

Kittelson & Associates, Inc., KFH Group, Inc., Parsons Brinckerhoff Quade & Douglas, Inc., Katherine Hunter-Zaworski (2003) *Transit Capacity and Quality of Service Manual* (TCQSM), 2nd Ed. TCRP Report 100. Washington, D.C, Transportation Research Board.

U.S. Department of Transportation (U.S. DOT) *The National ITS Architecture: A Framework for Integrated Transportation in the 21st Century, Version 5.1*, available at http://itsarch.iteris.com/itsarch/, last accessed in December, 2005

Using Transit Vehicles to Measure Freeway Traffic Conditions

B. Coifman,[a,b,c] S. Kim[a,d]

[a] Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, Columbus, OH, 43210. USA

[b] Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH, 43210, USA

[c] Associate Professor, coifman.1@osu.edu

[d] Graduate Research Assistant, kim.1936@osu.edu

Abstract

Recently, many public transit systems have equipped their vehicles with automatic vehicle location (AVL) systems, which periodically provide position and speed for each transit vehicle in the fleet. Although AVL is deployed for transit operations, the vehicles also provide valuable information about the traffic stream throughout the road network. In this study we develop a methodology to mine the transit AVL data and find all trips that use any portion of a pre-specified portion of a freeway. These trips are then used to measure travel time and average speed over the freeway, thereby quantifying conditions on the facility. The results are validated against concurrent loop detector data from the corridor, though the greatest benefit would be expected in areas without fixed vehicle detection.

I. Introduction

Measuring and monitoring traffic conditions is important for freeway management. Performance measurements typically use various wayside based detection technologies for this purpose, e.g., inductive loop detectors. A floating car equipped with a global positioning system (GPS) receiver can capture the vehicle's trajectory and by extension, the traffic conditions that gave rise to it. A floating car can measure conditions in areas that are not otherwise instrumented. However, traditional floating car studies are usually limited in scope and costly to implement. As intelligent transportation systems (ITS) evolve, new information is becoming available for traffic monitoring. Automatic vehicle location (AVL) based on GPS is one such example, it is being deployed on various fleets to manage their operation. Like floating car trajectories, an AVL system can provide traffic conditions without wayside detectors.

For example, Moore et al (2001) studied the use of freeway service patrol (FSP) vehicles' AVL to monitor the freeway conditions in Los Angels, CA. In this research performance measurement from FSP was evaluated and compared against measurements from loop detectors. Bertini and Tantiyanugulchai (2003) looked at transit vehicle travel time and travel speed data in Portland, OR, to assess traffic conditions on arterials. While Cathey and Daley (2003, 2003) looked at many of the detailed issues involved in extracting meaningful information on various segments and corridors from transit AVL data.

II. Analysis

In Columbus, OH, USA, the Central Ohio Transit Authority (COTA) operates 58 fixed routes and additional paratransit service with a fleet of 276 transit vehicles (COTA, 2005). Almost all of these vehicles, as well as many non-revenue vehicles, are equipped with AVL. The AVL transmits position, heading and speed to the COTA operations center usually once every minute. There is no universal standard for AVL systems and this study addresses some of the

challenges that arise in monitoring traffic conditions from transit vehicles with such infrequently reported data. The scope of the present work is limited to freeways where the confounding problems arising from bus stops are not an issue. The study employs the instrumented portion of I71 for validation and uses the system-wide AVL data collected during the month of May, 2003.

The study area along I71 begins in the central business district at the interchange with SR315 and I70, extending to the northern suburbs, covering roughly 19.5km in the shape of a "J", as illustrated in Fig. 1A. It includes almost all of the instrumented freeway in Columbus, ending just south of I270 (another 3.5km of freeway is instrumented beyond the northern end of the study region but few transit vehicles use that portion of the freeway). About 100 transit vehicle trips are observed in each direction on some portion of this segment in a typical weekday. Each vehicle's trip through the corridor is considered a single sample and is used to measure travel time and average speed.

To find vehicles in the corridor a coarse filter area is defined around I71 based on the coordinates of the roadway and if more than two successive data points from a given vehicle are found in this region these points are considered a sample. Sorted by time, the data in a sample directly provide total travel time and indirectly distance. These measurements are calculated from the first data point entering the given area and the last data point before leaving the filter area. Fig. 1B shows an example of all of the observations from one sample, including the first and last points. Of course these boundary points can only be identified by observing an adjacent point falling outside of the filter area, denoted with "out of I71" in the figure.

Travel time it is simply difference in time between "first point" and "last point". However, calculating the straight distance between the two points will generally underestimate the distance traveled because vehicles do not travel along straight lines and the observations are infrequent. Therefore, for each sample another methodology is necessary to obtain a more accurate distance measurement. In this paper we project the data points on to the closest point on a reference run through the corridor (similar to snapping points to a map database). In this case the reference run comes from a dedicated prove vehicle passing through I71 in each direction, with position sampled every second, though it could also come from a map database or manually generated using empirical data. Fig. 1A shows the reference run with a solid line, coincident with a highly dense region of data points. The distance between the first and last points in a

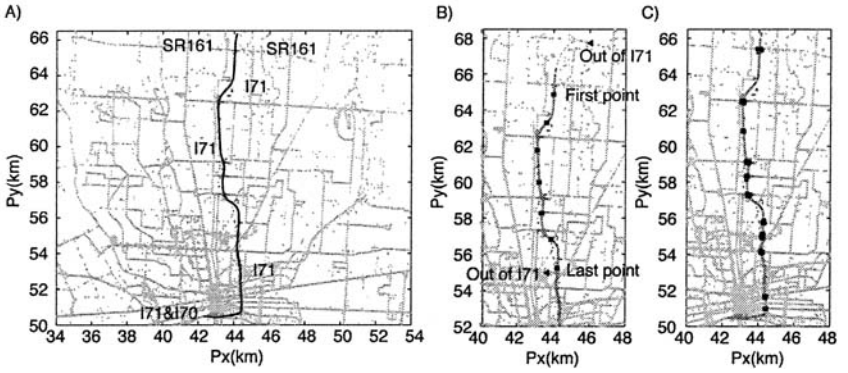


Figure 1, (A) Map of the study area on I71 (dark) and other routes (light), (B) a sample points from a southbound trip on I71, adjacent points just before and just after shown with triangles, (C) regions with uncertainty and errors due to interchanges.

sample is then taken to be the cumulative distance between their projections on to the reference run. Finally, speed is the quotient of travel time and distance. Figure 2 shows an example of the time series speed measured from the southbound trips from a single day.

Calculating link-based speed and travel time from loop data

To independently validate the transit measurements, this study employs the concurrent loop detector data from I71. There are 40 detector stations along the corridor at about 0.5km spacing collecting traffic data. This study uses five minute median speed from each station, measured across all lanes, representing traffic conditions at the given discrete point in space. To estimate link-based conditions from these point measurements we use the matrix of speeds indexed by detector station (distance) and sample (time) and assume that conditions remain constant on a given link for the entire 5 min sample and distance spanned by the link. Then we estimate vehicle trajectories assuming that vehicles traveled at these prevailing conditions. So trajectories change slope at the boundaries of the cells, either at the end of a link or at the end of a 5 min sample. From the trajectories one can estimate travel time directly, taking the difference in the times that a trajectory pass the chosen start and end points. In this case, every time a transit vehicle is observed in the corridor its first point is used as the start of a travel time estimate over the distance spanned to the corresponding last point.

Comparing transit measurements with loop measurements

After applying the algorithms to generate link-based measures independently from concurrent loop data and transit data, Fig. 3A and 4A respectively compare the resulting link-speed and travel times from the two methods. The three different symbols used for the points will be discussed shortly, for now consider all three groups as a single ensemble. The figures show the results from the entire month, both directions, with the x-axis presenting the loop estimates and the y-axis presenting the transit measurements. If the two independent measures for a given trip have the same value the point will fall on the line at 45-degrees, i.e., y=x, which is shown with a dashed line in the figures. Indeed, most of the measurements fall along this line, but many points are far away from it. Investigating the cause of the large deviations, most of them arose from errors in the data, specifically,

1. Extremely low (below 1mph) or high (e.g., infinite) reported in the loop speed matrix (45% of the errors),
2. Missing values in loop speed matrix causing erroneous travel time estimates (50% of the errors),
3. AVL appeared to report incorrect coordinates (2% of the errors),

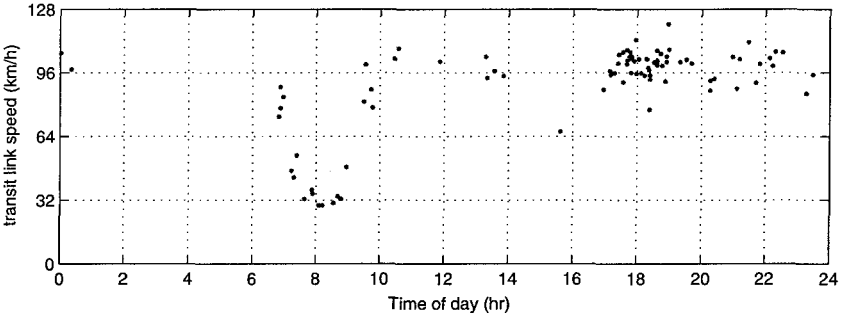


Figure 2, Time series of transit link velocity from southbound trips on I71 on Monday, May 5, 2003.

4. Ambiguity about the location of a transit vehicle that is near the freeway but potentially not on it (3% of the errors).

Addressing the errors

In the case of the first error, the erroneous speed measurements in the loop data were usually observed to last only short time periods and at only one station a time. If those erroneous measurements are reported from a station included within a given link, the travel time will be under-estimated or over-estimated and it will directly impact calculating average link speed. In this case, the travel time estimation algorithm was modified to either not generate any travel time measurement or to replace the infeasible measurement with the current speed from the loop detector station most immediately upstream of the problematic station. Using the upstream station, rather than previous time period at the given station, avoids the propagation of errors

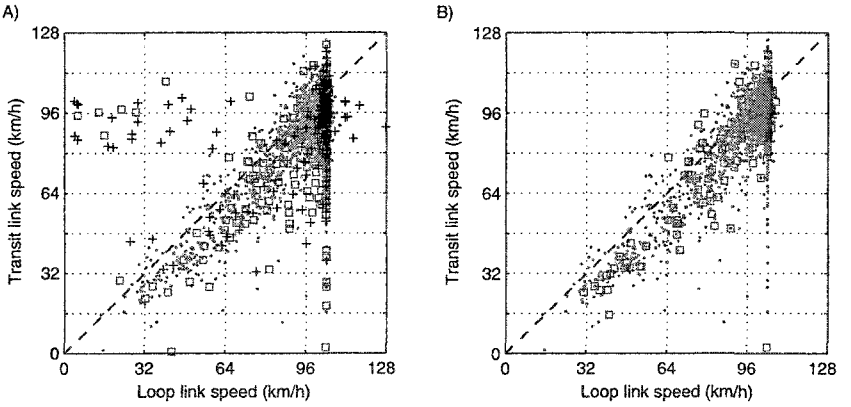


Figure 3, Using one month of data, measured transit link speed versus estimated loop link speed for the same trip (A) before filtering, (B) the same data after filtering.

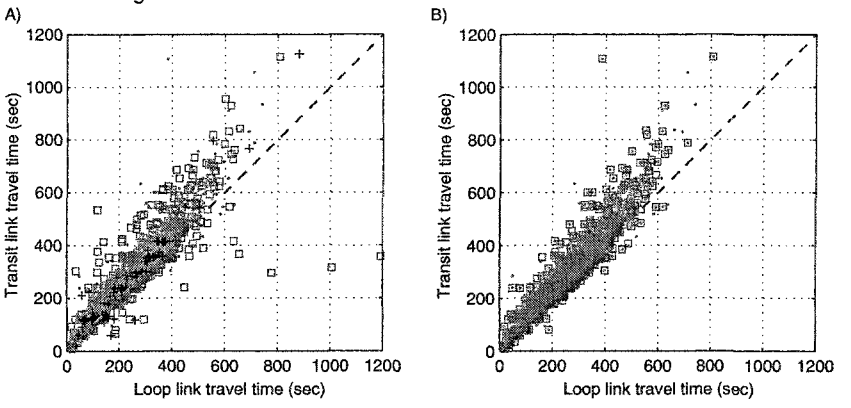


Figure 4, The travel times corresponding to the data in Fig. 3, (A) before filtering, (B) the same data after filtering.

when two successive samples at one station exhibit this problem (there were a few stations that were more prone to errors than others).

In the case of the second error, in the absence of any measurement at one station the loop based travel time estimate cannot be made across that cell of the matrix. Since these errors typically lasted for many samples and/or impacted several adjacent loop stations, it was decided to simply identify and exclude the trips impacted by such measurements from further analysis.

In the case of the third error, the AVL system would occasionally report the same coordinates for a vehicle in two or more successive samples. While it is possible for a vehicle to encounter stop and go traffic, it is not likely to remain completely stopped for one minute while on the freeway. So if two or more successive points in the AVL data from a transit vehicle have identical coordinates it is considered problematic in terms of calculating link travel time. If the last point and its predecessor point have the exact same coordinates, then the last point is reset to this predecessor point and the process is iterated if necessary. In all other cases, some of the points on the freeway are known to be wrong but the AVL system recovered before the vehicle left the freeway and thus the travel time should still be accurate.

In the case of the fourth error, the first or last point may be recorded off the freeway on a ramp, overpass, or underpass. Provided the point is close enough to the reference run and an interchange, it is impossible to tell from that single point whether it is on the freeway or off. So a vehicle that just exited via an off-ramp and was in the process of crossing over the freeway when it reported its position would lead to an erroneously large travel time since some of this time includes travel on the ramp and arterial. Or more formally, if the first or last point is close to an interchange, it is discarded and the next point on the freeway is used in its place (again, this process is iterated if necessary). If an intermediate point is close to an interchange it is retained, since the transit vehicles generally do not exit and immediately reenter the freeway. To this end, it is necessary to define several interchange areas along the corridor and filter out any end points in these areas, as highlighted by the squares in Fig.1C.

Re-comparing transit measurements with loop measurements, after filtering

After applying the filtering methods into current analysis, the link-based measurements were regenerated from both the loop and transit data. The combined results for both southbound and northbound directions from the entire month are presented in Fig. 3B and 4B, for link speed and travel time, respectively. Compare to Fig. 3A and 4A showing the unfiltered case. The majority of the points do not change going from the unfiltered to the filtered plots, as indicated with dots. A "+" in Fig. 3A and 4A indicates a point that was eliminated due to the filtering steps and is not included in Fig. 3B and 4B. Finally the square markers denote points that changed position as a result of the filtering, and in general they moved closer to the line at 45-degrees.

As evident in Fig. 3 and 4, many erroneous measurements have improved as a result of the filtering, addressing errors in both data sets. Some errors remain after filtering. Fig. 5 shows the cumulative distribution function (CDF) quantifying the difference between the before and after data from Fig. 3. About 90% of the measurements after filtering fall within 16km/h of one another. Still, some large errors remain in Fig. 3B, with roughly 10% of the differences exceeding the 16km/h threshold. Some of these arise due to the fact that the loop speeds are based on conditions in all lanes while the transit vehicle only traveled in a single lane, and the assumption that median conditions at the loop detectors represented an extended time and distance. These remaining errors also arise due to the fact that the AVL system was not deployed with the intention of monitoring the evolution of traffic conditions. In any event, the error rate may necessitate subsequent filtering, e.g., taking a median over a fixed time or fixed number of vehicles, to make the data practical for traffic monitoring.

III. Conclusions

Many conventional traffic management systems are limited in coverage due to the high cost to deploy the fixed detection infrastructure. Often times AVL data from vehicle fleets are available from roadways that are not otherwise instrumented, providing traffic conditions that would otherwise be difficult to collect. In short, AVL systems already deployed on vehicle fleets promise an inexpensive means to extend traffic monitoring to new roadways that may not otherwise be observed. To this end, this paper developed a methodology to monitor freeway traffic conditions from transit vehicle AVL with infrequently reported data. It used a rudimentary travel time estimation algorithm to estimate measures from loop detector data and validate the transit measurements. The scatter plots in Fig. 3 and 4 show that the link-based travel speeds from transit data are generally consist with the concurrent estimates from loop data. Several errors were found in both data sets and filters were developed to reduce their impact. Some errors remained, which may necessitate subsequent filtering, e.g., taking a median over a fixed time or fixed number of vehicles, to make the data practical for traffic monitoring.

IV. References

COTA (2005). "Central Ohio Transit Authority," http://www.cota.com/cota/cotaweb/main.xml, accessed on December 15, 2005.

F. Cathey, D. Dailey. (2002). "Transit Vehicles as Traffic Probe Sensors," *Transportation Research Record 1804*, TRB, pp 23-30.

F. Cathey, D. Dailey. (2003). "Estimating Corridor Travel Time by Using Transit Vehicles as Probes," *Transportation Research Record 1855*, TRB, pp 60-65.

J. Moore, II, S. Cho, A. Basu, D. Mezger (2001). "Use of Los Angeles Freeway Service Patrol Vehicles as Probe Vehicles," *California PATH Res. Program*, Berkeley, CA, Rep. UCB-ITS-PRR-2001-5.

R. Bertini and S. Tantiyanugulchai (2003). "Transit Buses as Traffic Probes: Use of Geolocation Data for Empirical Evaluation," *Transportation Research Record 1870*, TRB, pp 35-45.

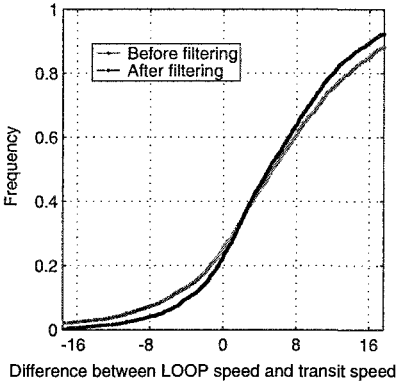


Figure 5, CDF of the differences between estimated loop link speed and measured transit link speeds for the data from Fig. 3.

Intensity of Lane Changing at a Freeway Ramp Weave Section

Ghulam H. Bham, Ph.D.[1]

[1]Civil, Architectural and Environmental Engineering, University of Missouri-Rolla, 135 Butler Carlton Hall, 1870 Miner Circle, Rolla, MO, 65409; email: ghbham@umr.edu

ABSTRACT

The study of lane-changing behavior of drivers is important for development of microscopic simulation models as they attempt to closely replicate real world phenomena. This paper presents the lane changing intensity at a ramp weave section observed from field data collected using aerial photography at 1-second intervals. Frequency and percentages of lane changing versus position of lane changes was observed for vehicles moving from the shoulder lane to the auxiliary lane and vice versa. The percentage of weaving vehicles from and to the auxiliary lane shows that the first 300 feet from the entrance gore represents a section of intense lane changing. The number of lane changes from the shoulder lane to the auxiliary lane was found to be double compared to the number of lane changes from the auxiliary lane to the shoulder lane. This paper also presents the frequency and percentages of lane changing from the shoulder to the median lane which occurs more heavily in the later part of the freeway section. This study may also be used to improve traffic operations by analyzing the adequacy of length of acceleration and auxiliary lanes.

INTRODUCTION

This paper analyzes lane change frequencies at a ramp weave section because of several reasons. First, weaving sections are usually accompanied by reduction in traffic flow. Second, in a weaving section the critical issue is the length of the auxiliary lane and the use of traffic control devices to reduce the number of lane change movements. Third, currently calibration and validation of microscopic traffic simulation models is performed by comparing traffic flow variables such as section density and average speed with field data (Bham and Benekohal, 2004). Comparison between lane-change patterns in simulation models and field data is not stressed, however. Fourth, safety in lane change movements is a concern as lane change crashes account for 4 to 10% of total crashes on freeways (Lee, Olsen and Wierwille, 2004). Lane changing frequency and driver behavior during a lane change plays a significant role in the above issues. In order to address some of these issues, this paper studies the lane change frequency and intensity at a ramp weave section of a multilane freeway (Figure 1). Lane changing mainly occurs due to weaving-in and weaving-out of the shoulder and the auxiliary lane. Vehicles also move from the shoulder lane to the median lane after they enter the highway. Vehicles exiting the highway make lane changes from the median lane to the shoulder lane and exit the highway using the auxiliary lane. Lane changing also occurs as drivers move from the shoulder to the median lane to increase speed and avoid vehicles entering the freeway. These movements create zones of intense lane changing where the highest percentage of lane change is observed. As lane changes are accompanied by intense acceleration and deceleration of vehicles, these zones may cause reduction in traffic flow. The effect of lane changing on traffic flow is out of the scope of this paper.

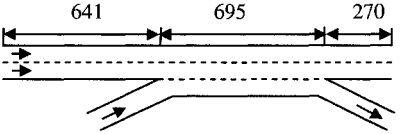


Figure 1. Ramp Weave Section, Baltimore –Washington Parkway

FIELD DATA

Field data (Smith, 1985) collected on Baltimore-Washington Parkway NB at I-95 (Capital Beltway) using aerial photography is used in the study (Figure 1). The ramp weave is a Type A weaving section on a cloverleaf interchange. The ramp inflow is from I-95 EB and the ramp outflow is to I-95 WB. The volume varies between 1200 and 2400 vehicles/hour (vph) on the freeway and between 300 and 1000 vph on the auxiliary lane. The data set provides lateral and longitudinal positions, speeds, and length of individual vehicles at 1-second intervals. The data was collected at p.m. peak period for an hour and is accurate to within 0.9 to 1.5 m. The data set was collected and has been analyzed for an hour.

DATA ANALYSIS

Lane change positions of every vehicle are analyzed in the field data. The starting positions of vehicles making a lane change are extracted from the field data and are preferred over ending positions of lane changers in the destination lane. As non-passenger vehicles were very small in number, they are not separated out from the main data.

DISCUSSION OF RESULTS

Frequency and Intensity of Lane Changing

Weaving-in and Weaving-out Movements

Figures 2 and 3 show the frequency of lane changes versus positions of lane changes for weaving movements. The frequency histograms are plotted for every 50 feet of the highway. Figure 2 shows the frequency of lane change positions for weaving-in vehicles from the auxiliary lane to the shoulder lane (right lane) and Figure 3 shows the lane change positions for weaving-out vehicles from the shoulder lane to the auxiliary lane. Figure 2 shows that the highest number of lane changes (112) occurs between 800 and 850 feet from the start of the section, about 160 feet from the entrance gore. This equals about two lane changes per minute in a 50 feet section. Figure 3 shows that the highest number of lane changes (251) occurs between 700 and 750 feet. This equals about four lane changes per minute in a 50 feet section. It can also be observed from Figures 2 and 3 that the frequency of weaving-out is almost double compared to weaving-in i.e. the total number of weaving-out movements is 1356 compared to 677 for weaving-in movements. Table 1 shows the number of lane changes for different lane movements. Figures 2 and 3 show a rapid increase in lane change frequency in reaching the maximum frequency and then a gradual decrease in lane change frequency. This clearly shows that the maximum frequency of lane changes occurs near the entrance gore. For weaving-in vehicles, the lane change frequency remains consistently higher for 200 feet (from 750 to 950 feet), and for weaving-out vehicles it remains consistently higher for 300 feet (from 600 to 900 feet).

Figure 4 complements Figures 2 and 3 by plotting the percent of weaving-in and weaving-out vehicles at every 100 feet of the highway. The figure shows bold rectangles on the shoulder and auxiliary lanes which indicate lane specific areas of intense lane changing. For weaving-in vehicles, the highest intensity is observed 100 feet from the entrance gore and the rectangular areas on the shoulder lane shows the area of highest intensity with 74% of lane changes for

weaving-in vehicles. The majority of weaving-in vehicles (82%) enter the freeway within 400 feet of the entrance gore. The intensity of lane changing for weaving-out vehicles is also concentrated near the entrance gore. Vehicles begin maneuvering even before the entrance gore as is evident from Figure 4 and continues for 400 feet (rectangular bold area). The main segment of intense lane changing is around the entrance gore around 300 feet of the weaving segment (indicated as "Intense Lane Changing" in Figure 4). In this segment, 73% of the total lane changing (weaving-in and out) takes place. The number of weaving-in and out movement (for the first 300 feet from the entrance gore) is 529, 545 and 401 per hundred feet during the pm peak hour data was collected i.e. a total of 1475 (73%) lane changes. This equals around 25 lane changes per minute per three hundred feet section during the peak hour.

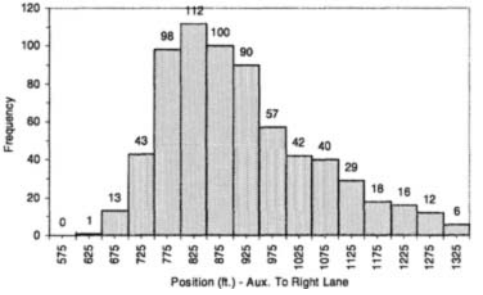


Figure 2. Frequency of lane changes – Auxiliary to Shoulder lane

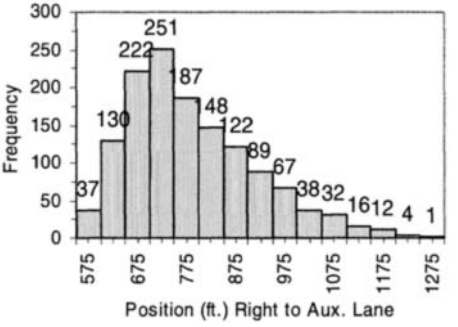
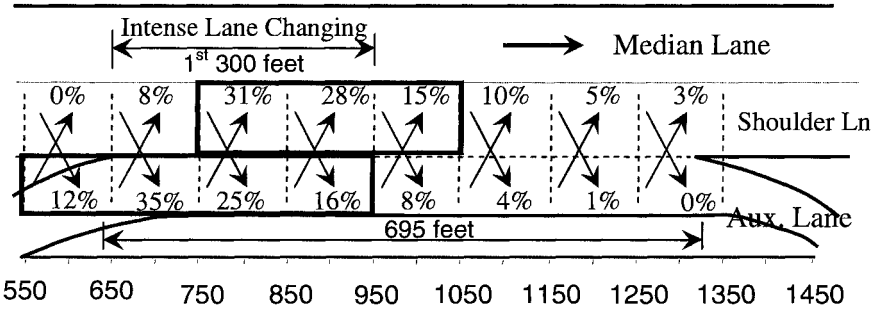


Figure 3. Frequency of lane changes – Shoulder to Auxiliary lane

Median and Shoulder Lane Movements

Figure 5 shows the lane change movements (364) from the median lane to the shoulder lane. From the figure it can be observed that lane change movements increase to more than double at 1000 feet, 335 feet before the exit gore. The exit gore is around 1336 feet from the start of the section. Lane change movements to the shoulder lane start increasing from around 1000 feet and almost half of total movements (176) occur after 1000 feet from the start of the section. A total of 186 lane changes occur between the start of the section and 1000 feet. Drivers move from the median lane to the shoulder lane to exit downstream of the section or to return to the

shoulder lane after moving to the median lane initially to avoid weaving-in vehicles from the auxiliary lane.



Note: Drawing Not to Scale

Figure 4. Percentage of lane change – Auxiliary to Shoulder and Shoulder to Auxiliary lane change movements

Figure 6 complements Figure 5 which shows the intensity of lane changing from the median to the shoulder lane. From 1000 to 1400 feet, lane change intensity is observed to be the maximum in the shoulder lane. A total of 148 (87 plus 61, representing 24 and 17%, respectively) lane changes take place. From Figure 4 it can be observed that the shoulder lane also has high lane change intensity for weaving-in movements. Thus the shoulder lane from 650 to around 1400 feet has the highest lane change intensity; 818 lane changes to the shoulder lane from adjacent lanes, and 1356 lane change movements (weaving-out) from the shoulder lane. This segment of the freeway is, therefore, most vulnerable to traffic flow breakdown.

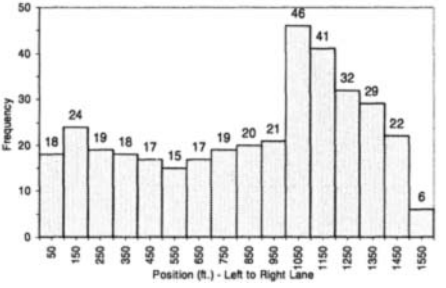
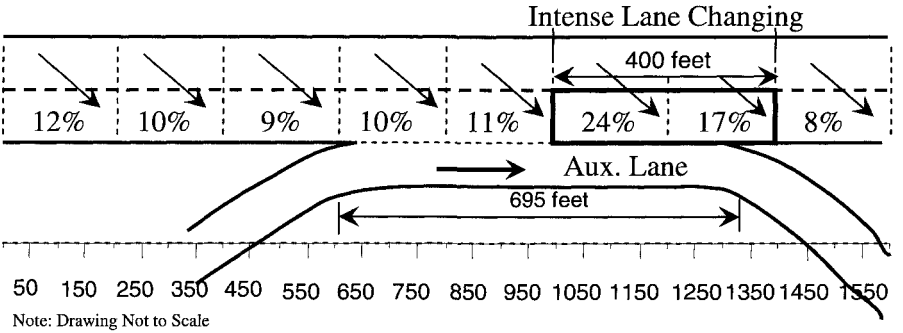


Figure 5. Frequency of lane changes – Median to Shoulder lane

Figure 7 shows a consistent movement of vehicles from the shoulder to the median lane. Vehicles move to the median lane to avoid weaving-in vehicles or to increase speed in the median lane which is the faster lane. Almost 45% (94 lane changes) of the movements occur between the start of the section and start of the entrance gore. A total of 205 lane change movements to the median lane are made during the peak hour.



Note: Drawing Not to Scale

Figure 6. Percentage of lane change – Median to Shoulder lane movements

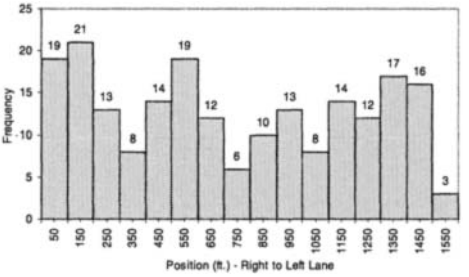


Figure 7. Frequency of lane changes – Shoulder to Median lane

Auxiliary to Median and Median to Auxiliary Lane Movements

Figures 8 and 9 show the cumulative frequency distribution of lane change movements from the auxiliary to the median lane and from the median to the auxiliary lane, respectively. These vehicles made two lane changes in the section and the figures show the cumulative percent and position of lane changes from the different lanes. A total of 128 vehicles made lane changes from the median lane to the auxiliary lane and 71 vehicles made lane changes from the auxiliary lane to the median lane.

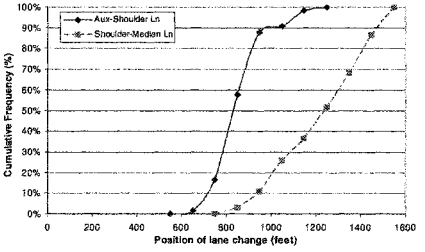


Figure 8. Cumulative Frequency (% age) of lane change movements – Auxiliary to Shoulder and Shoulder to Median lane movements

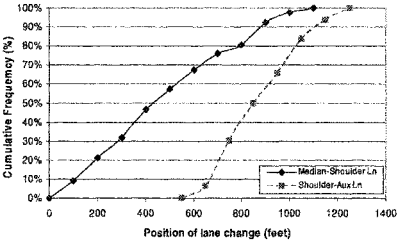


Figure 9. Cumulative Frequency (% age) of lane changes – Median to Shoulder and Shoulder to Auxiliary lane movements

Table 1. Number of Lane Changes (one hour data)

From (lane)	To (lane)			
	Median	Shoulder	Auxiliary	**# Total Lane Changes**
Median	-	364	128*	**492**
Shoulder	205	-	1356	**1561**
Auxiliary	71*	677	-	**748**
~ Total Lane Changes	**276**	**1041**	**1484**	**2801**

* indicates subset of total movements (e.g. 71 is a subset of 677 lane changes)
# number of vehicles leaving the lane
~ number of vehicles entering the lane

Table 1 shows in addition to the lane change movements from the origin to the destination lane, the total number of lane changes from and to a lane. The maximum number of lane changes made is to the auxiliary lane and made from is the shoulder lane. Since the entering volume in the shoulder lane is much higher (not shown in the paper) than the auxiliary lane, shoulder lane is most vulnerable to traffic flow break down.

CONCLUSIONS

It is observed that intense lane changing; both weaving-in and out, takes place within 300 feet of the entrance gore. Median to shoulder lane changing intensifies after 350 feet of the entrance gore. The shoulder lane, therefore have intense lane changing from near the entrance gore to the exit gore. It is also observed that 60% of weaving-out movements takes place in the first 200 feet and 76% in the first 300 feet of the entrance gore. The first 300 feet from the entrance gore on both the shoulder and auxiliary lanes are most vulnerable to collisions as they have the maximum conflicting points because of weaving-in and out movements. The weaving-out movement was found to be double compared to weaving-in movements; 1356 compared to 677 lane changes.

RECOMMENDATIONS FOR FUTURE RESEARCH

Lane changing behavior is important to study so that it can be replicated in microscopic traffic simulation models to represent more realistic traffic conditions. Current models mainly concentrate on traffic flow characteristics such as flow rate, density and average speed, however, very little has been done to show that the lane change behavior in simulation models is in agreement with real world data. To maximize traffic flow, the effect of lane change movements on traffic flow should be studied. Additionally, the use of traffic control devices such as pavement markings should be explored to reduce the effect of lane changes on traffic flow. As a result of a more detailed study on lane change movements, the geometric design of highways may be improved such as length of acceleration and deceleration lanes, length of weaving sections, etc.

REFERENCES

Bham, G. H. and Benekohal, R. F. (2004). "A High Fidelity Traffic Simulation Model based on Car-Following and Cellular Automata Concepts", *Trans. Res. C*, Vol. 12, No. 1, 1-32.

Lee, S. E., Olsen, E. C. B., and Wierwille, W. W. (2004). "A Comprehensive Examination of Naturalistic Lane-Changes", *DOT HS 809 702*.

Smith, S. A. (1985). "Freeway Data Collection for Studying Vehicle Interaction", Technical Report, *FHWA/RD-85/108*, U.S., Department of Transportation.

Construction Quality Effects on Pavement Asset Preservation and Valuation: A Mechanistic-Empirical Performance-Based LCCA Approach

Neville A. Parker[1], P.E., M.A.S.C.E., Sajjad Hussain[2], Kaan Ozbay[3] and Dima Jawad[4]

[1]CUNY Institute for Transportation Systems, The City College of the City University of New York, Convent Avenue & 138th Street, New York, NY 10031; Tel (212) 650-8054; Fax (212) 650-8374; email: parker@ti-mail.engr.ccny.cuny.edu
[2] Communication and Works Department, Government of Balochistan, Civil Secretariat, Block #7, Quetta, Pakistan; Tel (92) 81-9202374; email: beebarg@earthlink.net
[3]Department of Civil & Environmental Engineering, Rutgers, The State University of New Jersey, Piscataway, NJ 08854; Tel (732) 445-2792; Fax (732) 445-0577; email: kaan@rci.rutgers.edu
[4]Cultural Heritage and Urban Development-World Bank Project, Council for Development and Reconstruction, Beirut, Lebanon; Tel (+961)33742745; Fax (+961)7729528; e-mail: djawad@chud.cdr.gov.lb

Abstract

Direct computation of life cycle cost consequences of as-constructed – as opposed to as-designed -- quality measures, is presented as a basis for developing the pro-forma financial statements necessary for managing highway pavements as assets. The current and next-generation mechanistic-empirical pavement design and analysis models, combined with the development of easily deployable non-destructive testing technologies, increase the feasibility of estimating the combined effects of variations in the in-situ material properties that influence pavement quality and performance. These models can be deployed to develop the inputs to financial statements. In the process, one can not only develop a framework for monitoring financial health over any given performance period, but also estimate the difference between the life cycle costs associated with the as-designed and as-built pavement, which can, in turn, provide a basis for determining contractor penalties or bonuses, suggested here as being an important component of pavement asset management.

Introduction

Asset Management Challenge. Pavements are assets. Central to the concept of asset management is life cycle cost analysis and the business principle of financial reporting – income, financial position, and cash flows (FHWA, 2002; GASB, 1999). To be sure, that is a difficult practical proposition for public managers of public infrastructure assets, but nevertheless is an imperative for public accountability. While "asset value" and "cash flows" have been well understood and largely

reported, "income" and its financial implications, on the other hand, have been largely ignored.

As-Built vs. As-Designed Construction Quality. During the design-development phase of an infrastructure asset, proforma financial projections are made under the assumption of as-designed quality. However, as-built quality variables can literally vary from point to point. The initial financial projections are therefore subject to "adjustment" following the quality assurance check of the finished product.

Financial Proformas

Life cycle performance cost approach. Original SHRP wet-freeze distress performance models for roughness, rutting and thermal cracking (SHRP, 1994), calibrated for New York State – Table 1 -- and NYSDOT cost models (Prancl, 2000) – Table 2 -- were used to illustrate the development of the financial proformas. A typical life cycle output, for these variables and distress trigger values, is shown in Table 4, for a 45-year period, showing the maintenance, repair and rehabilitation (MR&R) quantities and their associated agency costs, for the pavement characteristics and input data in Table 3. A discount rate of 4% was assumed.

Table 1. SHRP distress prediction models used in the illustration

Roughness: $IRI = N^{0.25}\ 10^{C}$

$$C = 0.0403 + 0.00014AC_{visc} + 0.0704\ AC_{void} + 0.314\ \log(AC_{thick}) - 0.00162\ B_{thick} - 0.00165\ DGT + 0.00001628\ (FI*AC_{void})$$

Rutting: $RD = N^{0.183}\ 10^{C}$

$$C = 0.0289 - 0.189\ \log\ (AC_{void}) - 0.181\ \log\ (AC_{thick}) - 0.592\ \log\ (AGG) + 0.0000180\ AC_{visc} - 0.0436\ \log\ (B_{thick}) + 0.0000323\ (PRE * FI)$$

Thermal Cracking: $CS = N^{B}\ 10^{C}$

$$B = -0.106 - 0.0074\ AC_{thick} - 0.0504\ DGT$$

$$C = -0.0201 - 0.0131\ AGG + 1.84\ PRE - 0.0159\ (AC_{thick} * \log\ (PRE)) + 0.0024\ (AC_{thick} * DGT) + 0.00408\ (SUB * \log(PRE))$$

where: IRI = Roughness (International Roughness Index) (inch/mile)
RD = Rut Depth (inch)
CS = Crack Spacing (feet)
N = Cumulative KESAL (1000 ESAL)
AC_{visc} = Asphalt Viscosity (Poise)
AC_{void} = Asphalt Air Voids (% Volume)
AC_{thick} = HMAC Layer Thickness (inch)
B_{thick} = Base Layer Thickness (inch)
SUB = Subbase Passing Number 200 Sieve (% Weight)
AGG = HMAC Aggregate passing Number 4 Sieve (% Weight)
DGT = Annual Days Greater Than 90 Degrees (days)
FI = Freeze Index.(degree-days)
PRE = Annual Precipitation (inch)

Table 2. NYSDOT maintenance and rehabilitation costs for flexible pavements

Maintenance Treatment	Variable Cost ($/Lane-mile)	Variable Cost w/OH ($/lane-mile)
Crack Filling	1.63 * LFC	2.18 * LFC
Patching & Shimming	13006 + (80.15*Tac)	17428 + (107.40*Tac)
1.5" HMAC Overlay	34644 + (80.15*Tac)	46423 + (107.40*Tac)
2.0" HMAC Overlay	41223 + (80.15*Tac)	55238 + (107.40*Tac)
2.5" HMAC Overlay	47837 + (80.15*Tac)	64101 + (107.40*Tac)

Notes:

1. LFC = Linear feet of thermal cracking
2. Tac = Tons of asphalt shim course required

Source: Prancl, 2000

Table 3. Characteristics and input data for NYSDOT design alternative #8

a. pavement cross-section

Description	Input	Units
HMAC Layer	10	inches
Permeable AC Base	4	inches
Subbase	12	inches
Select Granular Fill	12	inches

b. traffic data

Description	Input	Units
Number of Lanes	4	each
DT	62500	veh/day
Heavy Vehicles	10	%
Truck Factor	1.35	ESAL/T
Directional Distribution	50	%
Lane Distribution	85	%
Volume Growth Rate	2.00	%
Weight Growth Rate	0.00	%

c. pavement quality characteristics

Description	Input	Units
Asphalt Viscosity	1600	poise
Asphalt Air Voids	4	% vol.
HMAC agg. passing #4 sieve	65	% wt.
Subbase passing #200 sieve	10	% wt.

d. environmental data

Description	Input	Units
Resilient Modulus (Mr)	5000	psi
Annual Days >90 deg	9	days
Freeze Index	925	deg-days
Annual Precipitation	38.00	inch

Table 4. MR&R life cycle costs (per mile) for the example

Age		Prediction Models			MR&R Strategies			Agency Cost	
Year	Cum KESAL's (1000)	IRI (in/mile)	Rutting (inch)	Crack Spacing (ft)	Maint. Strategy	Req Crack Fill (LF)	Req Shim (TON)	Est Agency Cost	Discount Agency Cost (@ 4%)
0.0	0	26.00	0.00	0.00	Initial Construction			1,433,656	1,433,656
1.0	1,309	36.32	0.22	103.56					
2.0	2,644	43.30	0.25	66.49					
3.0	4,006	48.04	0.27	51.30					
4.0	5,395	51.75	0.28	42.68					
5.0	6,811	54.86	0.29	37.01	Crack Fill	6,848	-	14,929	12,271
10.0	14,332	66.07	0.34	23.76	Shim & Fill	10,668	33.33	79,119	53,450
15.0	22,635	74.07	0.37	18.33	Crack Fill	13,825	-	30,138	16,735
20.0	31,802	80.64	0.39	15.25	Shim & Fill	16,617	38.57	92,650	42,284
25.0	41,923	86.41	0.41	13.22	Crack Fill	19,165	-	41,779	15,672
30.0	**53,098**	**91.67**	**0.43**	**11.77**	**Overlay 1.5"**	21,535	42.36	190,764	58,816
35.0	65,436	96.58	0.44	10.66	Crack Fill	23,765	-	51,808	13,129
40.0	**79,058**	**101.26**	**0.46**	**9.79**	**Overlay 1.5"**	25,883	45.56	200,588	41,780
45.0	94,098	105.77	0.48	9.08	End Life		-	**Total**	**1,687,793**

Proforma financials compressed. Table 5 is a compressed financial projection for the first 20 years, based on the LCCA in Table 4. Revenue assumptions are based on a multiple of projected traffic intensity, either through direct toll charges or "shadow pricing" reflecting value to the user. Zero depreciation reflects the assumption that MR&R expenditures "preserve" the structural and functional sufficiency of the asset. The 4% valuation reflects an assumed rate of return on the investment.

Table 5. Example proforma financials

Year	Capital Costs	Income (assumed)	MR&R	Depre-ciation	Asset Value	Net Returns	Total Cash Returns
2005	1,433,656	0	0	0	1,433,656	0	0
2006	0	130,900		0	1,433,656	130,900	130,900
2007	0	264,400		0	1,433,656	264,400	264,400
2008	0	400,600		0	1,433,656	400,600	400,600
2009	0	539,500		0	1,433,656	539,500	539,500
2010	0	681,100	14,929	0	1,433,656	666,171	666,171
2011	0	825,600		0	1,433,656	825,600	825,600
2012	0	973,000		0	1,433,656	973,000	973,000
2013	0	1,123,400		0	1,433,656	1,123,400	1,123,400
2014	0	1,276,800		0	1,433,656	1,276,800	1,276,800
2015	0	1,433,200	79,119	0	1,433,656	1,354,081	1,354,081
2016	0	1,592,700		0	1,433,656	1,592,700	1,592,700
2017	0	1,755,500		0	1,433,656	1,755,500	1,755,500
2018	0	1,921,500		0	1,433,656	1,921,500	1,921,500
2019	0	2,090,800		0	1,433,656	2,090,800	2,090,800
2020	0	2,263,500	30,138	0	1,433,656	2,233,362	2,233,362
Year 2020 Valuation @			**4**	**% =**		**11,050,284**	

Contractor Pay Adjustment

Assumed as-built vs. as-designed quality variables are shown in Table 6a, and the combined effects of these variations are shown in Table 6b. Note that in Table 6b, the impacts of quality variations on the life of the pavement, in terms of ESAL's, for the different failure criteria, are also shown. A combined run with the SHRP models – similar to Table 4 -- computed a present value increase of $9,213, per mile, in agency cost. By this methodology, the contractor pay adjustment should be in the form of a penalty, computed as a fraction of this increase, up to, but not exceeding 1.0. The penalty should then be added to the agency revenue base and the proforma financials updated.

Table 6. Multiple quality effects of as-built variations on life cycle cost

a. as-designed vs. as-constructed quality variables

Quality Variable	As-Designed	As-Built
Air-void contents(%)	4	5
HMAC aggregate passing # 4 sieve (% weight)	65	70
Subbase passing #200 sieve (%weight)	10	15
Asphalt concrete thickness (inch)	10	11
Base layer thickness (inch)	24	22

b. life cycle cost impact of as-built variations

Failure Criteria / Scenario	Roughness Trigger @ 100 inch/mile	Rutting Trigger @ 0.38 inch	Thermal Cracking @ 15 ft spacing	Life-Cycle Agency Cost ($) PV	Variation in Life Cycle Agency Cost ($)*
Life with as-designed values	34.73 Million ESAL's	35.73 Million ESAL's	32.77 Million ESAL's	1,697,853	-
Life with as-built values	13.55 Million ESAL's	60.35 Million ESAL's	28.03 Million ESAL's	1,707,066	-9,213

*NOTE: Lower agency cost is taken as positive value (benefit to the agency) and higher agency cost as negative value (loss to the agency).

Conclusions

The degree of sensitivity of preservation costs to an agency, due to variation in mix and laydown properties, as shown in Table 6b, can also assist an agency in estimating the impact on the valuation of the asset. For example, it can be shown that, for the example presented in this paper, if the added MR&R costs were assumed to be apportioned to the three time periods shown in Table 5, the valuation, for the

assumed rate of return – 4% -- would be reduced from $11,050,284 to $11,044,438, or by $5,846 per mile.

The byproduct of a mechanistic-empirical basis for assessing penalties or rewards to contractors would also be a move by industry away from the traditional concept of "pay adjustment schedules", to a state-of-the-art concept of a "pay adjustment system" over time (TRB, 2002). If it is accepted that the mechanistic response of pavements to load and environment will indeed become state-of-the-practice in pavement design and analysis, then most certainly quality assurance will have to move in the direction of measuring the full spectrum of in-situ quality variables. NDT technologies make this a reality, with random sampling eventually giving way to dense-grid acquisition of this data.

References

FHWA (2002). "Asset Management Primer." Office of Asset Management, U.S. Department of Transportation, Federal Highway Administration, Washington, D.C.

GASB (1999). "Statement 34 Resource Center." www.gasb.org/repmodel/index.html

NYSDOT (1994). "New York State Thickness Design Manual for New and Reconstructed Pavements, Rev. 1." *Technical Services Division, New York State Department of Transportation.*

Prancl, P.J. (2000). "An LCCA Procedure for Selecting and Evaluating Durable Pavement Structures Using SHRP-LTPP Mechanistic-Empirical Relationships." Ph.D. Dissertation, City University of New York.

SHRP (1994). "Evaluation of the AASHTO Design Equations and Recommended Improvements." *Strategic Highway Research Program. Report SHRP-P-394.* Strategic Highway Research Program, National Highway Council, Washington, D.C.

TRB (2002). "Glossary of Highway Quality Assurance Terms." Transportation Research Circular, Number E-C037. Transportation Research Board, Washington, D.C.

Vibration-based System for Pavement Condition Evaluation

Bill X. Yu[1] and Xinbao Yu[2]

[1] Assistant Professor, Department of Civil Engineering, Case Western Reserve University, 10900 Euclid Avenue, Bingham 210, Cleveland, OH 44106-7201, xiong.yu@case.edu
[2] Graduate Assistant, Department of Civil Engineering, Case Western Reserve University, 10900 Euclid Avenue, Bingham 210, Cleveland, OH 44106-7201, xxy23@po.cwru.edu

Abstract

Proper evaluation of pavement conditions provides important decision-support to implement preventative rehabilitation. Traditional method for pavement inspection relies on human observation that has low-efficiency and is subjective to errors. More recent tool for pavement distress inspection utilizes digital video and image analyses to record and identify the pavement surface conditions. This significantly increased the inspection efficiency and reliability. However, the video based system requires large storage space and extensive computation for image processing. It is also difficult to automate the pavement rating. This paper describes the use of recent data acquisition hardware to develop a vibration-based system for preliminary evaluation of pavement conditions. In analogy to the video record which "looks" at the surface of the pavement, the vibration-based system "feels" the ground conditions based on mechanical responses of the testing vehicle. The interactions of the ground and vehicle can be described by a model where vehicle is under random force excitations. The distresses of the pavement, including the cracks and surface rutting, impose impacting forces on the vehicle. The frequency and magnitude of the forces are dependent on the extent and magnitude of pavement distresses. On the other hand, the pavement surface conditions can be estimated from the recorded responses of the testing vehicle when driving on the pavement. A testing system was set up and data were collected to validate the measurement concept. Analyses of the testing data indicate that there is a good correlation between vibration responses and the pavement conditions. Factors such as the driving speed and the transition of vehicle motion were investigated. The vibration-based pavement evaluation system described in this paper has the advantage of small storage requirement, cost-effective and amenable for automatic real-time data processing. While this system does not provide the complete details of distress characteristics as by video-based system, it can be an inexpensive tool for routine inspections to provide preliminary evaluation of pavement conditions.

Introduction

The major roadway construction practice in the United States is switching from building new pavement to innovative management and renovation of existing ones. Preventative rehabilitation is an economic way of maintaining the existing pavement system. Appearance and extent of distresses on the pavement surface is an important indicator that justifies the necessity of pavement rehabilitation. Information of the pavement conditions is an important part of the pavement management system. The commonly used approach for inspecting the surface distress of highway pavements is based on visual observation. The method has low efficiency. The results by this approach are subjective to human preference. There have been prolonged research interests in developing automated distress detection and recognition system using digital image analysis. Further research is needed to reduce the implementation cost, increase the processing speed and accuracy (Wang 2000, Wang and Gong 2005).

For assessing the distress of a pavement, a number of crack attributes may be used for their detection as listed by Mendelsohn (1987). While the most obvious method is to detect the

cracks visually, the appearance of cracks can also be alternatively detected using high resolution pavement profiling. Both methods essentially "look" at the surface of the pavement. There are also indirect methods based on analyzing the interactions between vehicles and the pavement. From this, it is possible to estimate the conditions of the existing pavement. Study of the interactions between vehicles and road has recently gain considerable research interest. The interactions can be used, for example, to study the influence of pavement condition on vehicle driving cost. A variety of information on the pavement conditions can also be estimated from analyzing the mutual interactions between pavement and vehicles. Most drivers can "feel" or "hear" the pavement conditions while driving through the roads. The existence of cracks and rutting causes pronounced vibrations of the passing vehicles. Extensive surface cracks also generate significant slapping sound due to pavement-tire thread interactions. All these information can be potentially utilized. Nevertheless, research efforts are needed to summarize these daily experience into rugged methods for pavement conditions rating.

This paper introduces the preliminary results to explore the possibility of using vibration-based system to estimate the pavement conditions. With the progress in instrumentation technology, vibration-based system can be constructed with low cost. The recorded signal requires significantly less data storage compared with that based on imaging the ground surface. The reduced amount of data also makes it possible to process the information in real time.

Rationale of Vibration-based Pavement Evaluation

Roads undergo deterioration under traffic loads and various environmental loads. The deteriorating of a road is typically first noticed by drivers as they feel jarred by the rutting surface. Surface distress is defined as "Any indication of poor or unfavorable pavement performance or signs of impending failure; any unsatisfactory performance of a pavement short of failure" (Highway Research Board, 1970). Broadly speaking, surface distress modes include fracture, distortion and disintegration (Washington Asphalt Pavement Association). The surface distress can act as indicator for the structural integrity. It is a sign of impending or current structural problems. The existence of surface distresses also poses additional dynamic loads on the vehicles.

The cracks along the pavement surface generate random force on the surveying vehicle. Vibration measurement can thus provide indirect estimation of the pavement conditions. Figure 1 shows the schematic of evaluating the pavement conditions based on measured vertical acceleration. The surveying vehicle, which is an ordinary passenger car, can be reasonably described by a mechanical system with two degree of freedoms. The response of this system is equivalent to a two degree of freedom system under random excitations. For a linear time invariant system, the input of force following Gaussian distribution generates output which also follows Gaussian distribution. The time domain characteristics of a random signal can be described by its statistical values. The vertical vibration of the suspension can be measured by an accelerometer.

It needs to be pointed out that acceleration is part of the instrument for pavement roughness indices such as the International Ride Indices (IRI) (Gillespie 1992). The information of acceleration is typically used in conjunction with the laser profiler to measurement ground profile. From another viewpoint, the acceleration record alone might present certain useful information on pavement conditions. This is the prospect this paper tries to explore.

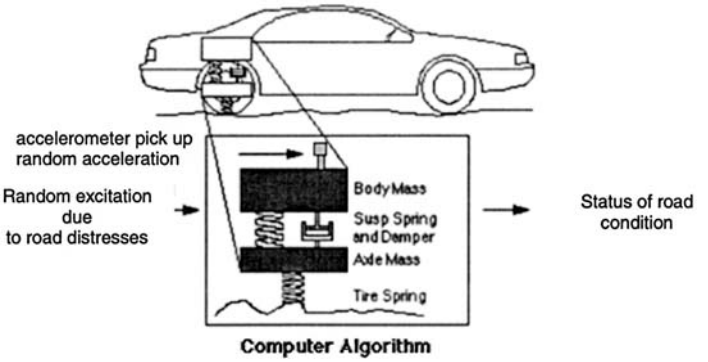


Figure 1. Schematic of the principle for evaluating the road surface conditions

Measurement System

Accurate measurement of acceleration can be realized with an inertia system, which is expensive. For this study, the vertical acceleration is measured by an ICP (Integrated Circuit Piezoelectric) accelerometer with sensitivity of 100mV/g. The advantages of ICP accelerometer compared with an ordinary piezoelectric accelerometer include that the sensor signal is not strongly influenced by the vibration and tapping of the connection wires. Thus the signal conditioning unit or charge amplifier does not need to be placed adjacent to the accelerometer. Long connection wire can also be used. A batter powered signal conditioner is used to excite the ICP accelerometer. If necessary, the signal conditioner can be replaced by an ICP signal conditioner powered by a USB port. This will reduce the reliance on battery power. The output from signal conditioner is connected to a PC-oscilloscope using an ordinary coaxial cable. The oscilloscope has two channel inputs and is then connected to the laptop via the USB port. The acceleration signal can be displayed in real time on the laptop. The laptop can be powered using the power adapter. The accelerometer can be mounted at any convenient locations. In this investigation, it is mounted on the frame under of the passenger seat via a magnetic base. The whole measurement package is very compact with total cost around $2,600.

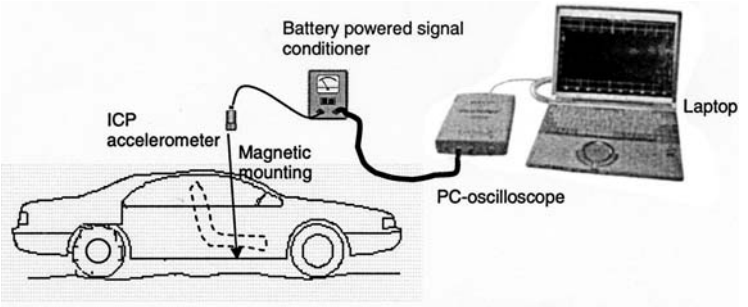


Figure 2. Set-up of the vertical acceleration measurement system

Figure 3 shows examples of recorded signals using this system. Figure 3a is the vibration due to idle of the engine. Figure 3b shows the measured vertical acceleration during on a highway section visually rated as good road condition. Figure 3c shows the measured vertical acceleration while driving on a highway section visually rated as poor road condition. A few sharp peaks appear in the recorded signals. This might be due to noises in the electronics. The exact source of this noise is unknown at this time and is still under investigation. The peaks were first removed in the subsequent analyses. The figure indicates that the vibration due to idling of the engine is relatively small. Bad road condition causes significant vertical vibration of the vehicle.

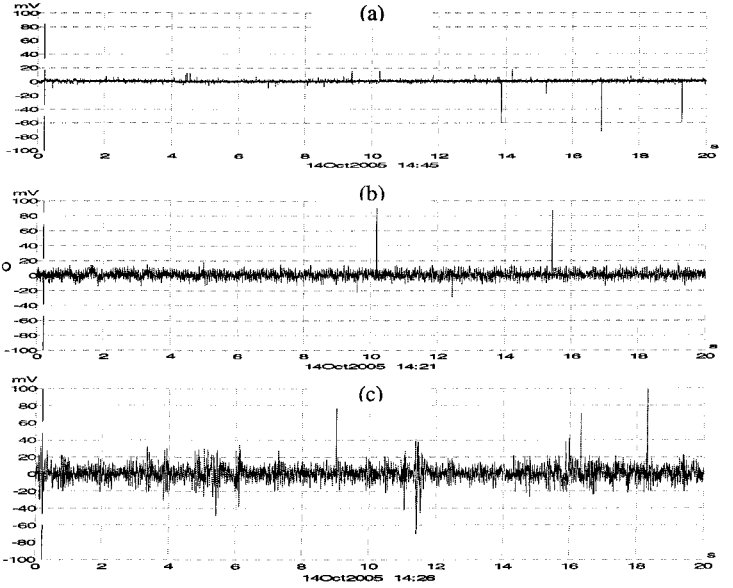


Figure 3. Example of acceleration records: a) Idle of engineering; b)pavement rated as good; c)pavement rated as poor (accelerometer sensitivity: 100mV/g)

It was observed during the testing that the results of a vibration-based system are consistent for rehabilitated sections. This is believed to be an advantage of this system. For the case where the surface cracks of the pavement are sealed using Bituminous Crack Sealants. The new sealant has significant contrast in color to the existing pavement surface. This however can be misinterpreted as cracks by the image processing algorithm.

Results and Analysis

Statistics of Recorded Acceleration

From the acceleration record, the statistics of the magnitude of acceleration can be obtained. Figure 4 shows an example of the statistics of the vertical acceleration record measured in 20s. The car drives at a speed of 50 MPH. Figure 4a, 4b are examples of the statistical distribution of the vertical acceleration when driving on pavement sections visually rated as good

or bad road conditions respectively. The statistical distribution of the vertical acceleration follows Gaussian distribution, which is typical of a random signal.

The values of average acceleration and the standard deviation by fitting the Gaussian distribution to the measured data are also shown in Figure 4. For both cases, the mean values of the acceleration are close to 0.01g. The standard deviation is around 0.08g for road section rated as in bad condition. The standard deviation is around 0.049g for section rated as in very good condition.

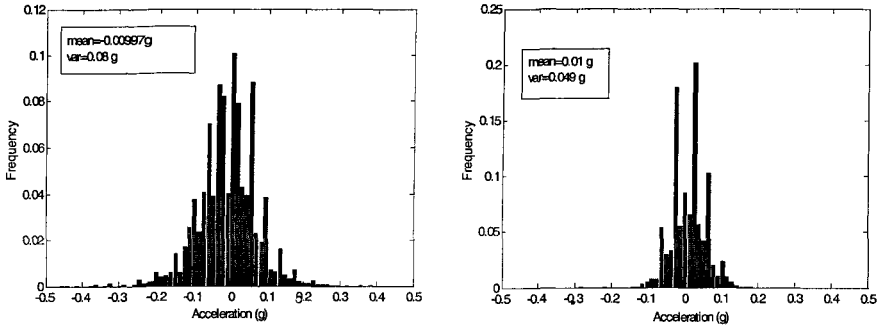


Figure 4. Statistics of acceleration record: a) Interstate 90 East around Exit 200, visually rated as bad road condition with extensive horizontal cracks; b) Interstate 90 East, around exit 190, visually rated as very good road condition

The fact that the mean value of acceleration is not exactly zero might be due to small drift of data acquisition electronics. The standard deviation of acceleration is a strong indicator of the road conditions. The standard deviation of the acceleration for riding on a very good road is around 0.05g. The vertical acceleration is related to the dynamic load applied to the road. Thus for the road rated as very good, the dynamic load applied to the ground is around 0.05 of the weight of the vehicle. This is consistent with the observations by Ervin et al. (1983) and Gillespie (1992). Gillespie (1983) found that for perfectly smooth road, the dynamic load index (defined as the standard deviation of the load normalized by the static load) is around 0.05. This corresponds to dynamic load generated not by the road, but by non-uniformities (imbalances and runouts) in the tires of the vehicles. Although the dynamic load index was obtained by Ervin et al. (1983) on trucks, this research indicates the dynamic load index might apply to a passenger car as well.

Influence of bridge abutment

The existence of bridge abutment causes significantly higher level of vibration. Figure 5 shows an example of vertical acceleration when the car approaches a highway bridge from a composite road visually rated as in very good condition. Figure 5a shows that there is a big bump when the vehicle approaches the bridge. After this, the vehicle continues to drive on asphalt surfaced abutment section, where the magnitude of vibration starts to decrease slightly. The surface of the bridge, which is made of concrete, causes much larger vertical acceleration compared with composite surface. The trend continues until the vehicle drove off the bridge. Figure 6b plots the change of the standard deviation of acceleration with time. The standard deviation is calculated in 0.2s interval. The average standard deviation of the vertical acceleration on the approaching road section is 0.0478g. That on the bridge is around 0.094g. The ratio of the standard deviation implies that the dynamic load acting on the bridge is around 1.97 times that normally acting on the road.

The observation suggests it is necessary to account for the influence of bridges on the overall rating using vibration-based system. The large vertical acceleration also implies larger dynamic load acting on the bridge. This potentially affects the service life of a bridge. Besides, the risk of riding a motor vehicle is related to the acceleration experienced by the driver. The observation from this measurement indicates measures might be necessary to reduce the dynamic load on the bridge as well as to increase the safety of the driver. A few proposed measures include 1) improving the bridge abutment design specification to reduce the vertical acceleration when approaching the bridge; 2) improving the bridge surfacing to reduce the magnitude of vertical acceleration.

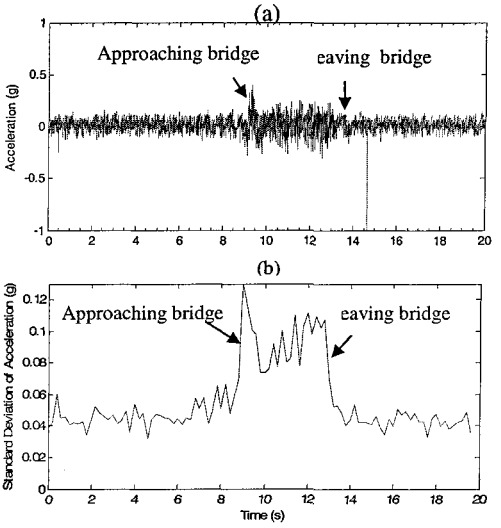


Figure 5. Responses of vehicle when passing a highway bridge: a) Vertical acceleration; b) Standard deviation of acceleration

Discussions and Conclusions

The paper introduces the pavement evaluation based on recording the vertical acceleration of a driving vehicle. The standard deviation of the acceleration record is found to be strongly related to the pavement conditions. Plotting the standard deviation with time provides a simple approach to identify the changes in the driving conditions. For implementation of this method, visual rating can be used as calibration to establish the rating criteria. The major advantage of the vibration-based system includes the low cost, small storage, and capability for real time data processing. The shortcomings with the vibration-based rating system include that the recorded response of the vehicle is that modulated by the response of the surveying vehicle. Thus the results might not be comparable unless the service condition of the vehicle such as tire pressure is calibrated. Alternatively, the idea employed by IRI could be employed, i.e., the measured road profile can be used to calculate the acceleration of a standard vehicle model. The variation of standard deviation of the acceleration record can then be plotted with time. The pavement rating can be obtained from this plot by referring to the standard deviation-based rating criteria. This will require additional computational efforts but the results will be independent of the measurement equipment.

References

Ervin, R. D., et al. (1983). "Influence of Truck Size and Weight Variables on Stability and Control Properties of Heavy Trucks." University of Michigan Transportation

Gillespie, T. D. (1992). "Everything You Always Wanted to Know about the IRI, But Were Afraid to Ask!" Presented at the Road Profile Users Group Meeting, Lincoln, Nebraska, September 22-24

Highway Research Board. (1970). Special Report No. 113: "Standard Nomenclature and Definitions for Pavement Components and Deficiencies", Highway Research Board, National Academy of Sciences, Washington, D.C., *http://www.asphaltwa.com/wapa_web/modules/08_evaluation/08_categories.htm*

Research Insitute UMTRI-83-10/2, Federal Highway Administration FHWA-RD-83-030, 179 p.

Wang, K.C.P. (2000), "Design and Implementation of Automated Systems for Pavement Surface Distress Survey", ASCE Journal of Infrastructure Systems, Vol.6, No1, March, pp. 24-32.

Wang, K.C.P. and Gong, W.G. (2005). "Real-Time Automated Survey System of Pavement Cracking in Parallel Environment", Journal of Infrastructure Systems, Vol. 11, No. 3, pp. 154-164

Definition of Yield Zones on Concrete Barrier Structures Under a Transverse Impact Load

Huali Geng[1], Nien-Yin Chang[2], and Trever Wang[3]

[1]Huali Geng, P.E., M.S., JR Engineering, LLC., 6020 Greenwood Plaza Blvd., Greenwood Village, CO. 80111, Tel: (303)740-9393, Fax: (303)721-9019, Email: hgeng@jrengineering.com
[2]Nien-Yin Chang, P.E., Ph. D., Prof., Department of Civil Engineering, University of Colorado at Denver and Health Science Center, Denver, Colorado, Tel: (303) 556-2810, Fax: (303) 556-2368, Email: Nien.Chang@cudenver.edu
[3]Trever Wang, P.E., Ph.D., Senior Bridge Engineer, Colorado Department of Transportation, Denver, Colorado, Tel: (303) 512-4072, Fax: (303) 758-7152, Email: ShingChun.Wang@dot.state.co.us

Abstract

The American Association of State Highway & Transportation Officials (AASHTO) LRFD (Load and Resistance Factor Design) Bridge Design Specifications[1], the nationally-accepted specifications for bridge designs, stipulate the design method for concrete barrier on concrete bridge deck. This study observes stress distributions on the concrete barrier and its adjacent structures when they are under a transverse impact load. Three research models are established for this study. An area of potential maximum tensile or compressive stress on the concrete barrier, which is called a "yield zone", can be considered the location of failure under the extreme transverse impact load. The results of this research show that the shapes and locations of yield zone are affected by the support structures and their rigidity. This study also discusses the limitations of the analysis method used by the AASHTO Specifications.

Yield Line Analysis Method

AASHTO has adopted the analysis method of yield line as a strength design method for concrete barriers under transverse load.[1] An assumed yield line pattern is consistent with the geometry and boundary conditions of the barrier for the design method. A solution is obtained by "equating the external work due to the applied loads to the internal work done by the resisting moments along the yield lines". Based on the concept of this method, the yield zone (the range of maximum tensile stress in X-direction) decreases from top to bottom on outside face of barrier and the two branches of the yield zone (the range of maximum tensile stress in X-direction) meet at the bottom on inside face of barrier (see Figure 1).[2]

Research Details

These research models are established and analyzed using SAP2000 finite element program, widely used in structural analysis.[6] The basis of these three

research models is the Colorado Type 7 Concrete Barrier as shown in the Colorado Transportation of Department (CDOT) design specifications.[5]

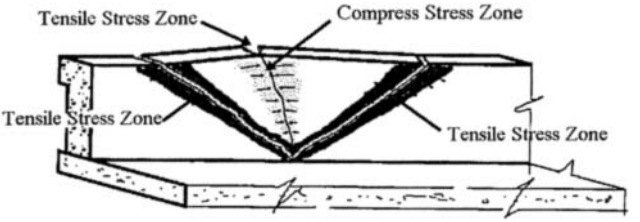


Figure 1. Assumption of Yield Zones on Concrete Barrier of AASHTO

The transverse impact load is adopted from AASHTO[1] Table A13.2-1 (p. A13-5). In order to more easily observe the stress distribution and yield zone, the transverse (Y-direction) load component of TL-6 is exaggerated and the longitudinal and vertical load components ignored. Figure 2 presents the load component directions of the impact load system. To simplify the analysis, the uniform transverse load component is converted into a concentrate load and is applied on the models. A linear time-history analysis method is used for these research models.[3,4] The impact loads on the research model are assumed to have a rise and decay time of 0.05 seconds, respectively, as shown in Figure 3.

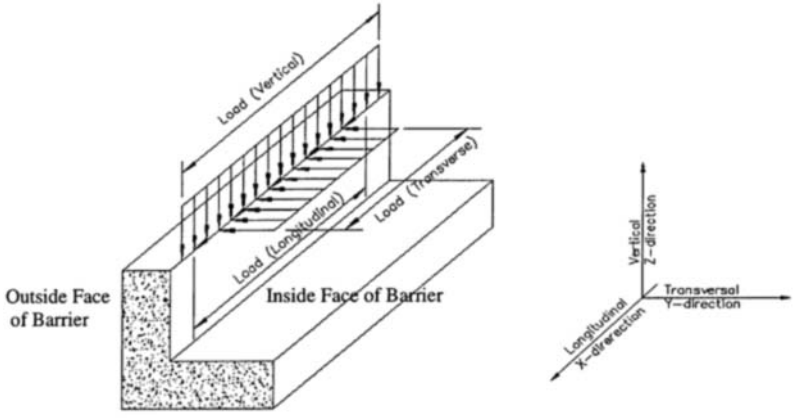


Figure 2. Three Load Components (inside face view)

Concrete is strong in compression and weak in tension. The research focuses on tensile failure on the research models. Contours of maximum tensile stress in the X (longitudinal) direction are presented in the figures.

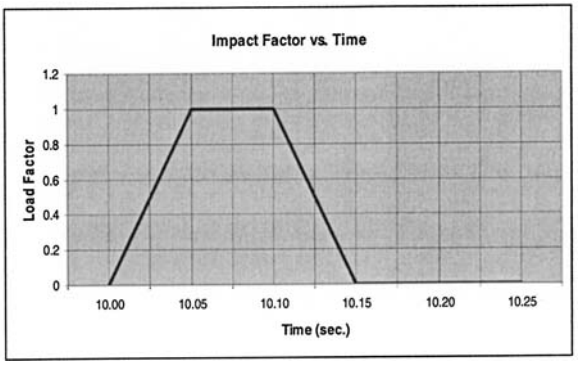


Figure 3. Impact Load Factor vs. Time, (Chang and Fatih, CDOT-DTD-R-2003-2) [3]

Research Model A

The research model A simulates a simplified classical structural analysis model. The barrier is fixed on the ground. The length of the model is 80 feet. The cross section of the barrier and the research model is shown in Figure 4 and Figure 5.

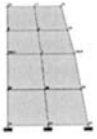


Figure 4. Cross Section of Concrete Barrier

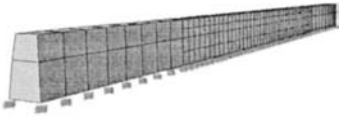


Figure 5. 3-D View of the Model A

Due to the fixed support condition at the bottom of the barrier, the only possible location of deformation is at the top of the barrier. The region of deflection of the barrier is a relative wider at the top, tapering to zero at the bottom. Therefore, the tensile yield zone shows an upside down triangle on the outside of the barrier (Figure 6) and "V" shape on the inside of the barrier (Figure 7).

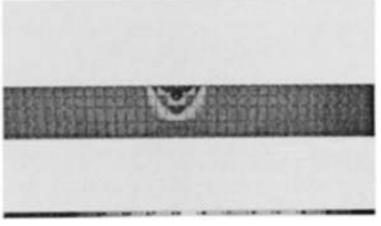


Figure 6. Yield Zone on the Outside Face of Concrete Barrier

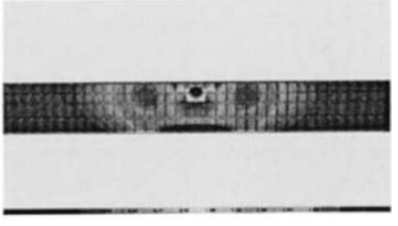


Figure 7. Yield Zone on the Inside Face of Concrete Barrier

It is concluded that the yield zone pattern of the Model A under impact load matches the AASHTO Yield line pattern and is similar to that of the yield line on the concrete plate.

Research Model B

The research Model B is a concrete barrier with an anchor slab on top of the retaining wall system. The thickness of the concrete anchor slab is 8 inches, while the length of the model is 120 feet (Figure 8). The movement of this barrier is only somewhat restricted by the soil (MSE). This model allows for the vertical displacement (sink), horizontal movement (slide), and rotation (about X-axis) at the junction of barrier and anchor slab.

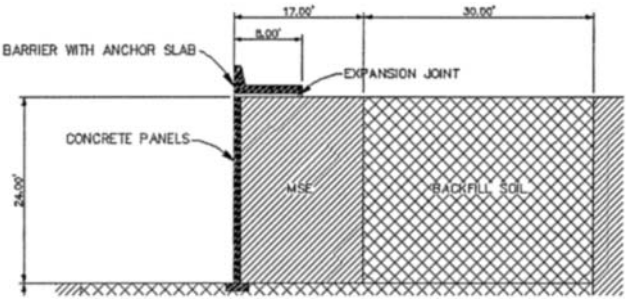


Figure 8. Model B---Concrete Barrier with Concrete Anchor Slab

Figure 9 presents a triangular tensile yield zone on the outer surface of the barrier and Figure 10 presents a roughly "V" shape on the inner surface of the barrier. In this case, the barrier has less restriction on the boundary since it connects to the anchor slab. The yield zone on the inside face of the barrier still retains the character similar to that of the plate's yield zones, such as the V-shape of the yield zone in model A. The opening of the yield zones is wider on the top of the inner face of the barrier and turns into the anchor slab when it reaches the junction of barrier and slab. Previous research shown that the impact load can affect the concrete barrier more than 200ft on each side. [3] The deformation of the model is not just bending but also sliding between the anchor slab and soil (MSE). The deformation of the barrier is a deformation of an

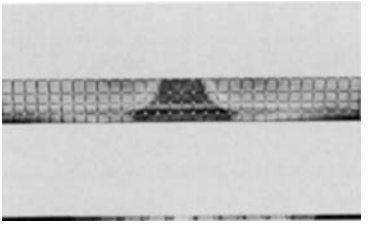


Figure 9. Yield Zone on Outside Face of the Concrete Barrier

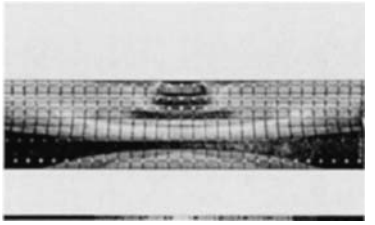


Figure 10. Yield Zone on Inside Face of the Concrete Barrier

assembly of the barrier and the anchor slab. Since this assembly of structures functions like a single structure, the tensile yield zone does not show up significantly in these figures.

Research Model C

The research model C simulates a practical case that is commonly encountered by bridge engineers and presented as a typical design case in AASHTO. The model includes a concrete barrier seated on the edge of 8-inch thick concrete deck with a 7-ft overhang (Figure 11). The span of the model is 8.5 ft, measured as the space between the exterior girder and the interior girder. The length of Model C is 80 feet in the longitudinal direction.

Figure 11. Model C---Concrete Barrier on Concrete Deck

The horizontal movement of this barrier is restricted by the concrete deck. The rigidity of the barrier is greater than the deck. When the rotation (about X-axis) of the barrier occurs, the concrete barrier and deck will rotate together to maintain the continuity of the junction.

Figure 12 shows the tensile yield zone on the outer face of the barrier. It appears that the yield zone pattern is actually the opposite of the one assumed by AASHTO. Figure 13 shows that the yield zone spreads continuously from the bottom of the barrier to the first support of the deck (exterior girder). The stresses decrease significantly after passing the first support. It appears that the V-shape nature of the plate yield zone is observed on both the outer and inner faces of the barrier. Observing the yield zones of this model, it appears that to apply the yield line design method as specified in the AASHTO Specifications is improper. However, when the

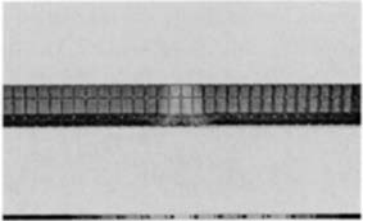


Figure 12. Yield Zone on Outer Face of the Concrete Barrier

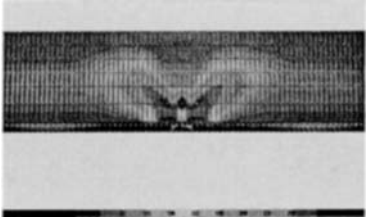


Figure 13. Yield Zone on Top Face of Concrete Deck (View from Top of Outside of Barrier)

yield zone extends into the concrete deck, the pattern of the plate yield line will appear on the top of the deck when the yield zone extends into the concrete deck. The above results show that the barrier and deck act together as an integral structure. The deformation of concrete deck structure may match the theory of the classical thin-plate structural analysis. However, the yield zone pattern of the barrier does not match the assumed failure pattern of the yield line as shown by AASHTO. This barrier does not have the character of the plate yield line.

Conclusions

The research shows that the theory of the yield line and failure pattern in the AASHTO Specifications has significant limitations and its closer examination is warranted. The yield zone is dictated by the boundary support structures.

For Model A, where the barrier is fixed on the elastic half space and its base movement is not permitted, the AASHTO specifications on yield line appear to be applicable. However, such ideal grounding is not realistic in the real world.

When the support structures of the barrier are flexible, like in Models B and C, the yield zone extends into adjacent supporting structures. The yield zone on the barrier no longer matches the assumption of the yield line theory of AASHTO. The shapes and locations of yield zone vary per the rigidity of supporting structures.

It is proposed to use a rigidity ratio, the ratio between the rigidity of a barrier and its adjacent supporting structure, as a factor in defining the yield zone. The greater the ratio of barrier rigidity to the support structures, like in Models B and C, the farther the yield zone extends into the support structure. On the contrary, as the ratio tends toward zero, that is the supporting structure is much more rigid than the barrier, such as in Model A, the pattern of the yield zone asymptotically approaches the shape assumed in the AASHTO specifications.

References

[1] American Association of State Highway & Transportation Officials (AASHTO). (1997). *LRFD Bridge Specifications ---U.S. Units*. 1997 Interim Revisions. Washington, D.C.

[2] Barker, Richard M., and Puckett, Jay A. (1997). *Design of Highway Bridges*. John Wiley & Sons, Inc., New York.

[3] Chang, Nien-Yin, and Oncul, Fatih. (2003). "Three-Dimensional Load Transfer of Colorado Type 7 and Type 10 Rails on Independent Moment Slab under Test Level Impact Load." *Colorado Department of Transportation Research Branch Report No. CDOT-DTD-R-2003-2*. Colorado Department of Transportation, Denver, Colorado.

[4] Chang, Nien-Yin, Oncul, Fatih, and Wang, Trever (2004) "Colorado Type 7 and Type 10 Rails under High Test Level Loads" ASCE GeoTransportation, UCLA.

[5] Colorado Department of Transportation. (2000). *Construction Standard Plans, M&S Standards.* Colorado Department of Transportation, Denver, Colorado.

[6] Computers and Structures Inc. (2002). *Analysis Reference Manual and Getting Started Basic Analysis Reference Introductory Tutorial (SAP2000 Version 8).* Computers & Structures Inc., Berkeley, California.

Paper withdrawn immediately prior to publication.
Next paper begins on page 202.

Paper withdrawn immediately prior to publication.
Next paper begins on page 202.

Paper withdrawn immediately prior to publication.
Next paper begins on page 202.

Paper withdrawn immediately prior to publication.
Next paper begins on page 202.

Paper withdrawn immediately prior to publication.
Next paper begins on page 202.

Paper withdrawn immediately prior to publication.
Next paper begins on page 202.

The Role of Advanced Technology in Asset Management: International Experiences

D. Mizusawa[1] and S. McNeil[2]

[1]Graduate Research Assistant, College of Urban Planning and Public Affairs, University of Illinois at Chicago, 412 S. Peoria St, Suite 340 CUPPA Hall, Chicago, IL 60607-7036; PH (312) 996-9818; FAX (312) 413-0006; email: dmizus1@uic.edu

[2]Ph.D., P.E., Professor of Engineering, Department of Civil and Environmental Engineering, University of Delaware, 360D DuPont Hall, Newark, DE 19716-3120; PH (302) 831-6578; FAX (302) 831-3640; email: mcneil@ce.udel.edu

Abstract

This paper reviews experiences with advanced technologies and asset management in Canada, New Zealand, Australia and England in national, state and local agencies visited on the Spring 2005 Asset Management International Scanning Tour supported by the Federal Highway Administration and American Association of State Highway and Transportation Officials. The paper describes both the technology and the context in which it is applied to Asset Management (AM) in terms of the institutional setting and the AM process used in order to identify how those countries implemented Asset Management Systems (AMS) using advanced technology in various areas. The paper concludes with lessons that can be translated to agencies in the United States in terms of technology adoption and implementation.

Introduction

AM is defined as a systematic process of maintaining, upgrading, and operating physical assets cost-effectively using an inventory of assets, a method of assessing current condition or performance, a process for determining needs, tools to evaluate and select appropriate strategies to address the needs and methods to evaluate the effectiveness of each strategy (U.S. DOT et al., 1998). AMS provide a holistic approach to strategic decision-making, which combines different management elements such as pavement management systems (PMS) and bridge management systems (BMS), based on a consistent evaluation method, to support trade offs in investments across the different elements. Also, AMS help agencies to understand the implications of different investment options (Cambridge Systematics Inc., 2004).

Scanning Tour

Federal Highway Administration (FHWA), American Association of State Highway and Transportation Officials (AASHTO) and the National Cooperative Research Program (NCHRP) jointly sponsored the international scanning tour "Transportation Asset Management". The team consisting of representatives from FHWA, state Departments of Transportation, a unit of local government and academic representatives visited Australia, Canada, England and New Zealand. The purpose of the scan was to investigate best-case examples of asset management techniques and processes in the world and identify lessons and applications for the U.S. The scanning tour took place in April 2005. The final report is available from FHWA or on the web (International Scan Study Team, 2005).

The countries and agencies visited were selected based on a "desk scan" of published material and interviews with experts. Once agencies indicated that they were willing to participate in the scan, they were sent a set of amplifying questions (developed by the scan team) intended to identify the areas of interest to the scan. Key questions related to technology addressed in this paper included issues such as inspection, data, analysis methods, and technology itself (International Scan Study Team, 2005).

Different levels of government are also represented. England and New Zealand serve as examples of national experiences. Alberta, New South Wales, Queensland and Victoria represent state or provincial experiences. Brisbane, Edmonton and London represent local government experiences.

Based on this framework the host agencies provided written materials, presentations and web references to support their responses to these questions. This paper attempts to organize these materials in terms of the role of advanced technology to support AM in the agencies visited as part of the international scan. The following section describes the sources for this paper, and the methodology used. The subsequent section presents the findings. A final section presents some conclusions.

Data and Methodology

The final report documenting the findings of the scan team, as well as the original source materials and personal notes from the scanning tour were reviewed to identify uses of advanced technology. The review focused on the role of advanced technologies in 1) data collection; 2) information technology including databases and geographic information systems (GIS); and 3) systems including decision support systems and integrated AMS. Each of these technologies is highly dependent on the widespread availability of inexpensive storage capacity and faster computing. The following section documents these applications and the subsequent section summarizes the benefits derived from the application of these technologies as identified by the agencies visited.

Findings: Applications of Advanced Technologies

Although the drivers for implementing AM varied in the agencies visited, there were several common themes. Specifically, AM becomes critical as resources become more and more constrained, the asset themselves age and deteriorate and continued growth (in terms of both the volume of traffic and increased truck traffic) exerts significant pressures to expand existing infrastructure. In some cases, such as England, laws require AM; in other cases, such as New Zealand, AM is an important tool for contract monitoring and management as the asset provision, maintenance and operation are outsourced. In all cases, AM is a data-driven process that requires significant technological support and consistently benefits from increased automation and improved systems to collect, manage, and access the data to support the process.

Data collection. All agencies reported that data collection, such as inventory and condition data, is time consuming and expensive. While the focus was on visual data collection, most agencies sought high speed and accurate data collection. Examples of advanced technology included the use of a high speed deflectograph and SCANNER surveys to obtain geometry, profile, texture and cracking data in the Highways Agency in England; the use of the Falling Weight Deflectometer (FWD) in New Zealand; smart sensors for applications such as slope stability in New South Wales and Victoria; and videolog in Queensland and Victoria. The SCRIM (Sideways-Force-Coefficient Routine Investigation Machine) was used to collect skid resistance, rutting, roughness, texture and videolog data in New Zealand, and Victoria.

Information technology. All agencies visited use databases and GIS to organize and reference their data. The applications continue to evolve as the technology changes and improves. Like experiences in agencies in the United States, legacy systems, and maintaining the data present significant challenges. Given that AM is a data-driven process, the use of information technology is the foundation for successful AM.

Asset management systems. The AMS found in the various agencies demonstrate the full range of approaches to asset management. The systems range from massive integrated systems to specialized tools to support decision-making. All are continually changing. The following subsections describe the AMS in the agencies visited.

England, the United Kingdom The Highways Agency in England has developed a series of AMS that use data from a network data repository and provide public information, and input to budget and program development tools. Specifically, these tools are the Highways Agency Pavement Management Systems (HAPMS) and Software for Whole-of-Life Economic Evaluation of Pavements (SWEEP); Structures Management Information System (SMIS); Highways Agency Geotechnical Database Management System (HAGDMS); Highways Agency Traffic Information System (HATRIS); Highways Agency Environmental System (HA-ES); and Technology Equipment Database (NOMAD).

At present, the decision support system resides in spreadsheets. However, it does integrate the principles of Whole-of-life Costing (WLC) and risk in assigning project priorities. Systems for integrating the process are currently being considered.

New Zealand The AMS used by Transit New Zealand have evolved over the past two decades. The Road Asset Maintenance Management (RAMM) system contains basic inventory and condition data, as well as maintenance history. This system also supports a program development tool, the National Optimization and Maintenance Allocation by Decade (NOMAD), and interfaces to the pavement management system dTIMS. In addition, two other systems, the Traffic Monitoring System (TMS) and a bridge information system including inventory and structural condition, are the core systems to support asset management.

The institutional structure and legislation in New Zealand requires that Land Transport NZ approve project requests. Land Transport NZ have a well defined set of priorities that require supporting cost benefit analysis and data that permits the consideration of tradeoffs between different types of investments. The AMS used by Transit New Zealand support both the project proposal phase of AM as well as monitoring the project delivery that is accomplished through outsourcing and continuous performance monitoring to ensure progress towards overall goals.

Alberta, Canada Alberta Infrastructure and Transportation (AIT) uses a Transportation Infrastructure Management System (TIMS) built around 20 software applications. Modules include:

- Network Expansion Support System (NESS) – Rates each highway section from the point of view of engineering, safety and socioeconomic policies and determines the class of work activities – data collection, engineering analysis, and rehabilitation/capital improvement.
- Bridge Expert Analysis and Decision Support (BEADS)
- Highway Pavement Management Application (HPMA)

TIMS is nearing the end of a 10-year development period. The use of a consistent scoring method for ranking projects in terms of both health and safety, and economic factors for all types of assets is an interesting approach. Also, TIMS is the only application seen on the scan that explicitly addresses cross-asset tradeoffs.

New South Wales, Australia The Road Asset Management System (RAMS) and the Traffic Asset Information Management System (TAIMS) are commercial software systems that inventory the road and traffic assets including condition data and work history for Road and Traffic Authority (RTA) of New South Wales (NSW). Separate systems manage the bridge assets and track contract maintenance activities.

These various systems communicate with each other to provide a map-based representation of condition along with performance measures and proposed work. Work is triggered using a deficiency analysis and projects are prioritized on the basis of risk.

Queensland, Australia The Road Asset Management System (RAMS) serves as the organizing principle for data, and the basic platform for decision-making tools, reporting tools and investment studies Queensland Main Roads. The Bridge Asset Management System (BAMS) and A Road Management Information System (ARMIS) address asset specific data.

Queensland Main Roads has particularly interesting tools for scenario analysis. *SCENARIO Millennium* is a rule based decision support tool for programming maintenance treatments. A similar tool, "Whichbridge" prioritizes bridge maintenance on the basis of risk.

Victoria, Australia VicRoads uses a Road Infrastructure Management System (RIMS) to structure their AM process. RIMS has five phases: 1) developing standards and guidelines, 2) developing the maintenance program, 3) implementing the maintenance program, 4) auditing, and 5) reviewing.

The Road Asset System (RAS) provides basic pavement and bridge data that is supplemented with data from the Road Crash Information System (RCIS) and the financial management system (PARMS) through a State Road Referencing System (SRRS). Within this framework formulae have been developed for various classes of assets that score projects and the score is then used to prioritize the projects.

Brisbane, Queensland, Australia The City of Brisbane, Queensland has developed the Total Asset Management System (TAM) for 16 classes of assets to support both short- and long-term decisions related to capital investment and maintenance. TAM also serves as an important communication tool for project justification and monitoring the condition of the assets. Decisions are based on whole-of-life costing and risk assessment. Specialized modules focus on "winks and blinks" assets and pavement management.

Edmonton, Alberta, Canada The City of Edmonton manages roads, sidewalks, bridges, streetlights, traffic signals, parking meters and traffic signs. Edmonton initiated the Municipal Pavement Management Application (MPMA) in 1986. This application has grown into an integrated approach to AM across all asset classes including life cycle costing and risk analysis. Common criteria – condition, functionality and demand/capacity – are used to prioritize project and compare different classes of assets.

London, England, the United Kingdom Transport for London's (TfL) Street Management Division uses the Asset Inventory and Management System (AIMS) to integrate historic and current data and sets project priorities using condition ratings. TfL uses the GIS capabilities of AIMS as a communication tool for presenting various scenarios based on deterioration and cost models.

Findings: Benefits Derived from the Application of Advanced Technologies

All the agencies visited were enthusiastic about the implementation and use of new technologies. However, very few agencies were able to articulate, much less quantify, the benefits of implementing these technologies. In Alberta, "AIT officials estimate that even if only 20 percent of AIT staff use the finished system, TIMS, it will have paid for itself." Specific benefits from using TIMS include best return on investment, documentation of decision-making, access to data and information, and support of the learning process. In Queensland Main Roads, a realistic asset value is used. This value has been used to assess the benefits of the pavement management system. For example, the PMS is credited with increasing asset value by 2.5% by accurately determining remaining life.

Conclusion

In summary, technology is used in the agencies visited to support a data driven AM process. In general the data collection technologies were very similar to those used by US agencies. However, several agencies demonstrated more sophisticated techniques for the management of data, data analysis and decision support. In all cases the tradeoffs between maintaining the existing assets and expansion of the network to support and encourage growth were recognized. Only Edmonton explicitly addressed the issue in an integrated way. All systems used GIS and life cycle costing (also referred to as whole-of-life costing) to support the analysis, representation and display of information. Several agencies also integrated risk into their decision support systems. More detailed exploration of these techniques is of value to US agencies.

Acknowledgements

The authors gratefully acknowledge the work of Michael Meyer of Georgia Institute Technology who served as the reporter for the scanning tour, and the staff of the agencies who participated in the scan. This research was partially supported by the Midwest Regional University Transportation Center and the Metropolitan Transportation Support Initiative (METSI) at the University of Illinois at Chicago.

References

U.S. DOT et al. (1998). *Pavement Management Systems: Participant's Workbook*, First Edition, U.S. DOT, FHWA, and NHI.

Cambridge Systematics Inc. (2004). Analytical Tools for Asset Management, Draft Final Report for NCHRP Project 20-57.

International Scanning Study Team (2005). Transportation Asset Management in Australia, Canada, England, and New Zealand, Federal Highway Administration, December 2005. http://international.fhwa.dot.gov/assetmanagement/index.htm. Accessed December 23, 2005.

Coordination of Maintenance and Rehabilitation Policies for Transportation Infrastructure

Pablo L. Durango-Cohen[1] and Pattharin Sarutipand[2]

[1]Assistant Professor, Department of Civil and Environmental Engineering, Northwestern University, 2145 Sheridan Road, A335, Evanston, IL 60208-3109, Phone: (847) 491-4008, Fax: (847) 491-4011, Email: pdc@northwestern.edu

[2]Ph.D. Candidate, Department of Civil and Environmental Engineering, Northwestern University, 2145 Sheridan Road, A321, Evanston, IL 60208-3109, Email: p-sarutipand@northwestern.edu

Abstract

This paper formulates the maintenance and repair (M&R) decision-making process as a quadratic program. The functional interdependencies in transportation infrastructure are captured in the model, and the conditions under which it is optimal to synchronize maintenance schedules for groups of components are found. This result suggests that an effective management process depends on attending to the interdependencies that link a system's facilities.

Introduction

To determine when to schedule maintenance or repairs to facilities, managers must weigh various factors including user costs and agency expenditures. In making their decisions, managers have had to rely on existing optimization models. However, many of these models focus primarily on computational tractability and, therefore, do not incorporate important functional characteristics. This type of approach creates obstacles to providing effective decision support because it cannot incorporate the significant benefits or additional expenditures created by the interdependencies that link various facilities in a system.

This current paper attempts to rectify this oversight by reformulating the decision making process as a constrained quadratic program (i.e., a QP problem). A QP problem is comprised of a set of linear constraints and a second-degree polynomial of the decision variables in an objective function. Using this approach, we design and test a model that captures the economic interdependence between a system's various facilities. Results generated by applying this model provide insight into transportation

infrastructure management and demonstrate the importance of incorporating interdependencies into M&R models.

Literature Review

Existing optimization models for the problem of finding maintenance and repair (M&R) policies for transportation systems are formulated as constrained Markov Decision Processes (MDPs) (Golabi et al., 1982; Golabi and Shepard, 1997). This type of model omits information about the individual components making it impossible to capture functional relationships. In contrast, several other studies in the field have argued that economic interdependencies are important for M&R decisions.

Recently Durango-Cohen and Tadepalli (2006) presented the discrete-time optimal control framework that effectively overcomes the computational difficulties associated with the multi-dimensional MDP approach of Parallel Machine/Equipment Replacement Problems (Vander Veen, 1985) and the continuous-time optimal control approach (Friesz and Fernandez, 1979; Li and Madanat, 2002). This motivates and justifies the extension of the framework to address the problem of a system comprised of multiple facilities.

Notation and Parameter Specification

We consider the problem of finding an optimal M&R policy for a transportation system that consists of n facilities over a planning horizon consisting of T periods/stages. The state and decision vectors for the problem are defined as follows:

$X_t \in \mathbb{R}^n$: Vector that represents the system's condition at the start of stage t.

$X_t \equiv [x_t^1, x_t^2, \ldots, x_t^n]'$ Each vector component represents the condition of a given facility, where larger values of a vector component correspond to worse condition.

$Y_t \in \mathbb{R}^n$: Decision vector representing the set of M&R decisions for period t.

$Y_t \equiv [y_t^1, y_t^2, \ldots, y_t^n]'$ Each vector component corresponds to the investment level/maintenance rate for the corresponding facility.

The fundamental assumptions that we use to specify our model are as follows:

A1: The period cost function, $p(\cdot)$, and the salvage/residual value function, $s(\cdot)$, can be represented (or approximated) by general second-order polynomials. Using linear algebra notation, we may write the polynomials as follows:

$$p(X_t, Y_t) = X_t' A X_t + X_t' B Y_t + Y_t' C Y_t + d' X_t + e' Y_t + f \quad (1)$$

$$s(X_{T+1}) = X_{T+1}' L X_{T+1} + w' X_{T+1} + z \quad (2)$$

A2: Facility deterioration is represented as a deterministic linear system of the form:

$$X_{t+1} = g' X_t + h' Y_t + k \quad (3)$$

Condition-Dependent User Costs: These costs correspond to the sum of user costs that are directly attributable to a facility's condition. In terms of specifying the period cost function, Equation (1), we assume that these costs are independent for each of the facilities that comprise the system. Thus, in stage t the costs may be written as:

$$CDUC = \sum_{i=1}^{n} \left[a_{ii} \left(x_t^i \right)^2 + d_i x_t^i + f_i^{cduc} \right] \tag{4}$$

Intuitively, one expects the condition-dependent costs for each facility to be convex and increase as condition worsens. Thus, $a_{ii} > 0$ and $d_i > 0$, $\forall i$. f_i^{cduc} is the portion of condition-dependent user costs attributed to facility i that is fixed and thus included in the constant term f.

Independent User and Agency Costs Associated with M&R Decisions: These costs correspond to the sum of costs associated with M&R of the facilities that comprise the system and that are independent of all other facilities. Thus, in stage t the costs may be written as

$$IM\&RC = \sum_{i=1}^{n} \left[c_{ii} \left(y_t^i \right)^2 + b_{ii} x_t^i y_t^i + e_i y_t^i + f_i^{IM\&RC} \right] \tag{5}$$

Intuitively, one expects that M&R costs for each facility would increase with the level of intensity of M&R actions. Thus, $c_{ii} > 0$ and $e_i > 0$, $\forall i$. One might also expect that the application of an action would be either independent or increasing with worsening conditions. Thus, $b_{ii} \geq 0$. $f_i^{IM\&RC}$ is the portion of the costs attributed to i that is fixed, i.e., included in the constant term f.

Dependent User and Agency Costs Associated with M&R Decisions: These costs correspond to the sum of costs associated with M&R decisions that link or that are dependent on the state or on the decisions for other facilities in the system. In stage t these costs may be written as

$$DC = \sum_{i=1}^{n} \sum_{j \neq i} \left[b_{ij} x_t^i y_t^j + c_{ij} y_t^i y_t^j + f_{ij}^{DC} \right] \tag{6}$$

The first set of terms reflects how the costs of investing y_t^j on j depend on the condition of facility i, x_t^i. One expects these set of terms to be nondecreasing with both x_t^i and y_t^j. Therefore, $b_{ij} \geq 0$. The second set of terms captures the link between the decisions for the different facilities in the system. When $c_{ij} > 0$ the terms might reflect costs associated with loss of functionality/throughput when facilities i and j are disrupted simultaneously. When $c_{ij} < 0$ the terms might reflect savings derived from personnel and equipment delivery costs when adjacent facilities are maintained in the

same period. Finally, the last set of terms accounts for fixed costs. Without loss of generality, we also assume that all parameters are stationary.

Salvage Value: We assume that the salvage/residual value of a facility is independent of other facilities, and thus, we assume that matrix L is diagonal.

Mathematical Model

The problem of obtaining an optimal M&R policy can be written as follows:

$$\text{Minimize:} \quad \sum_{t=1}^{T} \delta^{t-1}\left[X_t' A X_t + X_t' B Y_t + Y_t' C Y_t + d' X_t + e' Y_t + f\right] + \delta^{T}\left[X_{T+1}' L X_{T+1} + w' X_{T+1} + z\right] \tag{7}$$

$$\text{Subject to:} \quad X_{t+1} = g' X_t + h' Y_t + k \tag{8}$$

$$X_1 = \hat{X}_1 \tag{9}$$

$$Y_t \geq 0 \tag{10}$$

Equation (7) represents the minimum discounted cost over T stages. Equation sets (8) and (9) represent the deterioration of the facilities that comprise the system and their initial condition, respectively. Equation (10) constrains the investment level in each period to be nonnegative. We assume that the inequality applies componentwise to the vector Y_t.

Numerical Examples

In this section, we introduce examples that compare optimal policies obtained with the proposed model to policies obtained with an existing model. The QPs are solved with a nonlinear optimization solver, LOQO, which is available online through http://www-neos.mcs.anl.gov/neos/solvers/nco:LOQO/AMPL.html.

There are two simple types of networks to consider: substitutable and complementary, as presented in Figures 1(a) and 1(b), respectively.

a) Substitutable Network b) Complementary Network

Figure 1: Types of Network

Condition-Dependent Policy: Base Case

As a baseline for the more complicated case, we examine the results of the case when interdependencies are ignored. This means that the parameters b_{12} and c_{12} are set to zero. Thus, the results are independent of types of network. Table 1 shows the parameters of Equations (7) and (8) that are used in the example.

Table 1: Costs Parameter for Basic Network

Link	a_{ii}	b_{ij}	c_{ij}	d_i	e_i	f	g_i	h_i	k_i
1	1	b_{12}	1 c_{12}	-50	0	1049.5	1	-1	10
2	1	b_{21}	1	-50	0		1	-1	10

In this example, an infinite horizon version of the problem was considered. As a result, the optimal policy converges to a steady-state. The cost parameters in the model were also chosen so that the infinite horizon sum of discounted costs of managing each facility independently would be \$0. The discount rate was set to 5%. The deterioration was chosen so that it would take 10 periods to deteriorate through a hypothetical scale with range 100 - intended to mimic the PCI scale.

The results for the base case show that the optimal policy is to apply the same amount of investment on both links, as illustrated in Figure 2. This shows that the QP model to obtain the condition-dependent policy is intuitively similar to existing models.

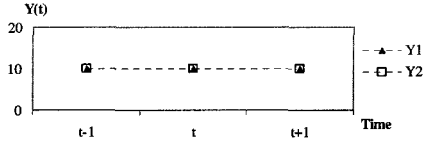


Figure2: Optimal Investment level of the Base Case

Capturing Interdependencies of the Substitutable Network

In this next example, a parametric study is conducted. In this example, the effects of parameters b_{12} and c_{12} are isolated and their effects on the ensuing optimal M&R policies are considered.

By varying a positive c_{12}, we observe that: When c_{12} is small, the optimal policies are similar to the base case. When c_{12} is large, the policy specifies alternating investments between links, as shown in Figure 3, suggesting that the traffic disruption dominates the decision-making process.

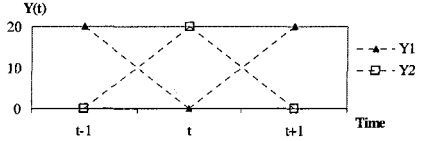


Figure 3: Optimal Investment level of the Substitutable Network with large c_{12}

Next, the user costs associated with rerouting must be considered. By varying parameter b_{12}, we observe that: When b_{12} is small, the optimal policies are similar to

the base case. When b_{12} is large, the optimal steady-state policies now are to alternate the investment between link 1 and 2, also as is shown in Figure 3. This is because one would expect an alternative route to be in very good condition during the original link is maintained. To achieve this situation, the investment must be alternated between links.

Capturing Functional Interdependencies of the Complementary Network

The negative parameter c_{12} is varied in this section. The results show that the optimal policy does not change from the base case. This arises because when one of the complementary links is maintained, a bottleneck is already created along the path. Therefore, the impact of maintaining the complementary links together would be the same as doing it on either one of them.

Conclusions

In this paper, we propose a QP formulation for maintenance optimization model of multiple facilities in transportation systems. This proposal led to a model that captures the economic interdependencies that reflect costs associated with (i) traffic disruptions caused by the presence of multiple M&R activities on the network, (ii) rerouting users, and (iii) the reduction in an agency's expenditures if there are benefits in synchronizing M&R actions on link pairs. By trading off user-costs with M&R agency-costs, our model identifies the condition under which it is optimal to synchronize maintenance schedules for groups of components.

Results generated by our model suggest that the optimal policies are to coordinate all complementary links while alternate between all substitutable links. This result indicates that an effective management process depends on attending to the interdependencies that link a system's various facilities. Hence, these interdependencies should be incorporated in the maintenance optimization model.

References

Durango-Cohen, P. and Tadepalli, N. (2006). "Using advanced inspection technologies to support investments in maintenance and repair of transportation infrastructure facilities." *ASCE Journal of Transportation Engineering*, 132(1).

Friesz, T. and Fernandez, J. E. (1979). "A model of optimal transport maintenance with demand responsiveness." *Transportation Research Part B*, 13(4), 317–339.

Golabi, K., Kulkarni, R., and Way, G. (1982). "A statewide pavement management system." *Interfaces*, 12(6), 5–21.

Golabi, K. and Shepard, R. (1997). "Pontis: A system for maintenance optimization and improvement of US bridge networks." *Interfaces*, 27(1).

Li, Y. and Madanat, S. (2002). "A steady-state solution for the optimal pavement resurfacing problem." *Transportation Research Part A*, 36(6).

Vander Veen, D. (1985). "Parallel replacement under nonstationary deterministic demand," PhD thesis, University of Michigan.

Impact of Higher Truck Loads on Remaining Safe Life of Louisiana Bridge Decks

Aziz Saber[1], Ph.D., P.E.; Freddy Roberts[2], Ph.D., P.E.;
Xiang Zhou[3]; and Walid R. Alaywan[4], P.E.

[1]Associate Professor, Louisiana Tech University, Department of Civil Engineering, PO Box 10348, Ruston, LA 71272; PH(318)257-4410; Fax (318) 257-2306; e-mail: saber@latech.edu
[2]Professor, Louisiana Tech University, Department of Civil Engineering, PO Box 10348, Ruston, LA 71272; PH(318)257-4611; Fax (318) 257-2306; e-mail: froberts@latech.edu
[3]Ph. D. Candidate, Louisiana Tech University, Department of Civil Engineering, PO Box 10348, Ruston, LA 71272; PH(318)257-2000; Fax (318) 257-2306; e-mail: xzh002@latech.edu
[4]Senior Structures Research Engineer, Louisiana Transportation Research Center (LTRC) 4101 Gourrier Avenue, Baton Rouge, LA 70808; PH(225) 767-9106; Fax (225) 767-9108; e-mail: WalidAlaywan@dotd.louisiana.gov

Abstract

Evaluation of the load capacity of existing bridge decks is a major concern for the Louisiana Department of Transportation and Development. Available funds must be expended in a rational strategy, which would allow for inspection, maintenance or rehabilitation. This paper presents the results of the study on the proposed higher truck loads and their economic impact to the state and to the industry. The impact of adapting such regulations on the existing Louisiana bridge decks is determined based on a deterministic load capacity evaluation as well as a reliability assessment. The target reliability level is derived from bridge deck structures designed to satisfy AASHTO standard design specifications and to satisfy safe and adequate performance levels. Fatigue analyses are performed using finite element modeling with detailed load capacity evaluations and reliability analyses on a representative sample of bridge decks to provide specific examples of expected changes in rating and safety levels.

Introduction

The Louisiana Department of Transportation and Development (LA-DOTD) conducted a review for the laws that govern the operation of vehicles that haul Louisiana products in excess of the standard limitations, as a result of The Senate Concurrent Resolution 123 that was passed in July 2004, by the Louisiana state senate. Under current state law, truck operators hauling specific agricultural and natural resource commodities can purchase overweight permits and haul at a gross vehicle weight (GVW) that exceeds the legislated limit of 80,000 lb. (36,365 kg.). One of these agricultural and natural resource

commodities is timber, which is harvested in all but two of Louisiana's parishes. In 2003, the forestry products accounted for almost 22 percent of the total agricultural production in Louisiana. The permit fees for trucks hauling timber at a GVW of 86,600 lb. (39,365 kg.) is $10.00 for one year. Because forestry is a very important part of the economic base of Louisiana, any changes in the legal weight or overweight permit structure for Louisiana must consider the potential impact on the forestry product industry as well as the cost to maintain and rehabilitate the roads and bridges used by vehicles hauling forest products. Figure 1 shows the vehicles (AASHTO 3S2; FHWA class 9) that are used to haul timber products along with the proposed new loads.

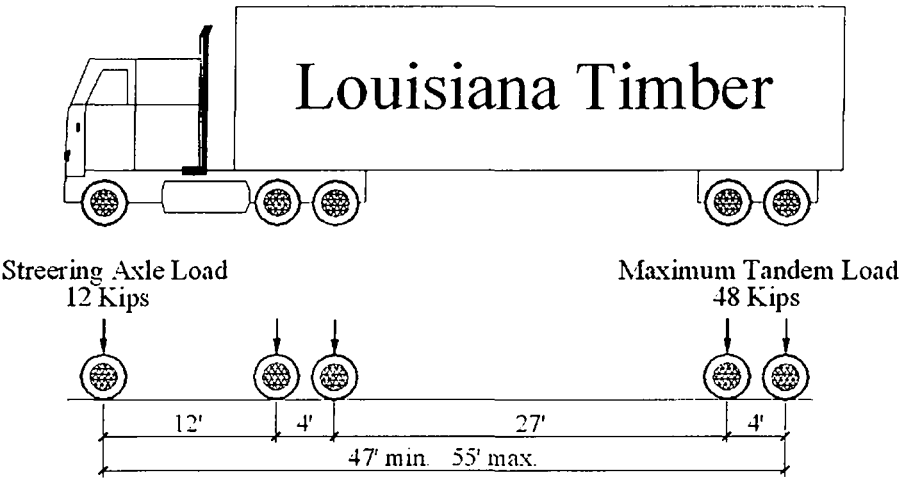


Figure 1 – Truck Configuration with Proposed New Loads (1 kip = 4.45 kN)

Objectives

The main objectives of this study are to : (1) Assess the impact of vehicles hauling timber products on the maintenance and rehabilitation of Louisiana state bridges under current state laws that set forth gross vehicle weights; (2) Assess the permit structure that describes the conditions under which legal overweight permits may be purchased; (3) Provide the legislature with proposals to modify current laws by providing new weight restrictions to reduce damage to Louisiana state bridges, while keeping these Louisiana industries economically viable. This paper presents only the results of the evaluation on the concrete bridge decks. The impact on bridge girders is presented in (Saber, Roberts 2006), and the highway cost consequences are presented in (Roberts, Saber 2005).

Scope

The primary thrust of this paper is to assess the magnitude of bridge deck rehabilitation costs resulting from vehicles hauling timber in excess of the 80,000 lb. (36,365 kg.) gross vehicle limit on Louisiana bridges. Some trucks hauling timber start out on parish roads adjacent to the land where the timber is harvested. However, this paper

concentrates on determining costs for concrete bridge decks on state bridges that the LA-DOTD is responsible for constructing, rehabilitating, and maintaining. For impact on off-system, the reader is referred to (Roberts, Saber 2005).

State Bridges Crossed By Timber Trucks

The bridges located on the Louisiana State Highways, U.S. Numbered Roads and most frequently traveled by timber trucks are considered critical in this study. The review and selection processes are based on the amount of timber harvest each parish produces. The control section numbers for these roads were identified and the bridge inventory database was used to identify the critical state bridges for this study.

Previous Studies

The truck industry can get more carrying capacity by increasing truck weight, while bridge owners need to control the loading on the bridges to keep the structures in a safe condition and to limit the deterioration of the existing bridge infrastructure. To achieve these goals, regulations are constituted to allow the truck weights to increase to a certain range while guaranteeing the safety and serviceability of the bridge system. In 1982, the Federal-aid Highway Act introduced a program regulating truck weights. This legislation prescribes the gross vehicle weight at 80,000 lb. (36,365 kg.) and the axle group weights are regulated based on the truck weight formula. This keeps the bridges designed with HS20-44 loading from being overstressed by more than 5 percent and the bridges designed with H15 loading from being overstressed by more than 30 percent, as presented in (Noel et. Al., 1985). Consequently, many states have increased their legal loads above the standard. In 1990, the Transportation Research Broad completed research Report 227 on a modification of the truck weight formula. The Bridge Formula reduced the limits on axle loads and allowed the higher gross weights.

Methodology

There are many studies done and methods used to manage and evaluate fatigue of bridge structures, (Ghosn et. Al., 1995) and (Moses, 1992). The site-specific information available at the time of this study on timber trucks was very limited and statistically insufficient for use with the NCHRP 495 approach or the other methodologies referenced.

The effects of hauling timber on Louisiana bridge decks are determined by comparing the strength and serviceability conditions of the concrete bridge decks under their design load to the conditions produced by a loaded 3S2 truck. The short and long term effects of the timber truck loads are determined based on the ratio of the strength and serviceability criteria for each concrete bridge deck in the sample.

Loads

The truck loads for hauling timber are based on the 3S2 truck configuration as shown in Figure 1. The standard truck configuration HS20-44 with load conditions listed in LRFD Bridge Design Specifications are used. In addition, a wearing surface loading of 12psf (575Pa) according to LA-DOTD Bridge Manual, is placed on the deck.

Bridge Decks

All bridges in this study had concrete decks. According to the LADOTD Bridge Manual, concrete bridge decks are designed as a continuous span over the girders. The bridge deck analyses for this study are performed using finite element models and GTSTRUDL software. The finite element models for typical bridge decks are generated with a typical width of 30-ft. (9.15 m.) and thickness of 8-inches (200 mm.) supported by 5 girders. The design load for the bridges included in this study, and the loads from 3S2 truck configuration are applied to the deck. The fatigue load combination, as presented in LRFD, is performed for these bridge deck models, with typical reinforcement details provided in the LADOTD Bridge Manual. To determine the corresponding stresses; truck loads for HS20-44 and 3S2 are applied at critical locations for maximum positive and negative moment in the bridge deck. The maximum value of longitudinal, transverse, and shear stresses in the bridge deck are grouped as the tensile stress and compressive stress at the top and bottom surfaces and presented in Tables 1 and 2.

Table 1 - Top Surface of Continuous Bridge Deck for 3S2 Truck Loads

3S2						
Span Length	Max Value of Stress (Ksi)					
(ft.)	Max Tensile Stress			Max Compressive Stress		
	Longitudinal	Transverse	Shear	Longitudinal	Transverse	Shear
20	0.055	0.156	0.060	-0.222	-0.317	-0.085
30	0.114	0.137	0.057	-0.204	-0.380	-0.065
45	0.222	0.121	0.066	-0.223	-0.435	-0.060
60	0.295	0.148	0.073	-0.240	-0.471	-0.064
75	0.317	0.187	0.084	-0.263	-0.548	-0.099
90	0.493	0.253	0.101	-0.396	-0.589	-0.112
105	0.660	0.308	0.118	-0.304	-0.622	-0.118
120	0.844	0.366	0.140	-0.382	-0.621	-0.149

1-ksi = 6.89 MPa, 1-ft = 0.305 m.

Table 2 - Bottom Surface of Continuous Bridge Deck for 3S2 Truck Loads

3S2						
Span Length	Max Value of Stress (Ksi)					
(ft.)	Max Tensile Stress			Max Compressive Stress		
	Longitudinal	Transverse	Shear	Longitudinal	Transverse	Shear
20	0.222	0.317	0.085	-0.055	-0.156	-0.060
30	0.204	0.380	0.065	-0.114	-0.137	-0.057
45	0.223	0.435	0.060	-0.222	-0.121	-0.066
60	0.240	0.471	0.064	-0.295	-0.148	-0.073
75	0.263	0.548	0.099	-0.317	-0.187	-0.084
90	0.396	0.589	0.112	-0.493	-0.253	-0.101
105	0.304	0.622	0.118	-0.660	-0.308	-0.118
120	0.382	0.621	0.149	-0.844	-0.366	-0.140

1-ksi = 6.89 MPa, 1-ft = 0.305 m.

The results from the bridge deck analyses indicate that the bridge deck is under a stable stress state, no matter whether the stresses are in the tension zone or the compression zone. Moreover, the decks of bridges with spans larger than 30 ft. (9.15 m.) may experience cracks in the longitudinal direction under 3S3 trucks. Such cracks will require additional inspections along with early and frequent maintenance.

Cost Based on Remaining Safe Life of Bridge Decks

Bridge cost is a combination of new design, rehabilitation, and fatigue costs. The focus of this study is only on the fatigue cost resulting from the increase in the load permits for timber trucks.

The long term effects of heavy trucks play an important role in the bridge life evaluation. The selected bridge decks for this study are designed under standard H15 or HS20-44 truck loads. Overloaded trucks traveling across these bridge decks will increase the cost of maintenance and rehabilitation. An accurate estimate for the cost of the damage is hard to obtain since fatigue damage may lead to repairs, rehabilitations, or replacements. Most of the bridges in Louisiana are designed for a fatigue life of 50 years. Overloaded trucks will definitely shorten the life of the bridges and their decks. The bridge decks in this study are evaluated for fatigue cost based on the results from the strength analyses presented earlier in this paper. Based on a review of the bridges considered in this study, the truck ADT value of 2,500 is used. The concrete bridge deck costs used in this study are based on projects completed by LA-DOTD during 2004. The average cost to replace a concrete bridge is approximately $90 per square foot. The fatigue damage due to the new proposed truck loads is determined based on fatigue damage law in which the strength criteria is raised to the third power. The percentage of life of the bridge deck used when a truck with the proposed loads crosses it, is the ratio of the fatigue damage and the span life design criteria. The estimated cost per trip across the bridge is obtained by multiplying the percentage of the life of the bridge by the average cost to replace a concrete bridge deck. The effect of the trucks hauling timber on the fatigue life of the bridge deck is ignored when the ratio from analysis is equal to or less than one. Since the trucks are operating on a broad route structure, the total fatigue cost is estimated on a per bridge basis. This applies to cases with no defined route for the vehicle. The estimated average cost across the bridge deck is obtained by multiplying the percentage of the life of the bridge by the cost to replace the bridge by the length of the bridge, then dividing the sum by the total length of the bridges. The results of this study indicated that the fatigue cost for bridges is controlled by the fatigue in bridge girders as presented in (Saber, Roberts 2006).

Conclusions

The impact of vehicles hauling forestry products on the maintenance and rehabilitation of Louisiana state bridge decks under current and proposed loads was evaluated. Forestry products accounted for almost 22 percent of the total agricultural production in Louisiana in 2003. The results of this study indicate that the current $10/truck/year permit fee on a FHWA class 9 vehicle 3S2 will not cover the additional maintenance and repair costs for bridges due to the new proposed loads. Since forestry is such an important part of Louisiana's economic base, any changes in the legal weight or overweight permit structure for Louisiana must consider the additional costs for bridge fatigue, as reported in the study, and the amount the state will provide to subsidize the forestry product industry.

Recommendations

Based on the results of this study and in order to assist the forestry product industry in Louisiana and reduce the bridge fatigue damage on the state system, the new proposed truck loads should be supplemented with modifications to the timber truck trailer (FHWA class 9 vehicle 3S2). The axle configuration should be modified from a tandem axle to a triple axle, and the gross vehicle weight should be at 86,600 lb. (39,365 kg.) uniformly distributed among these axles.

Acknowledgments

Support of this work was provided by Louisiana Transportation Research Center under research project number [05- 2P] and state project number [736-99-1299]. The support provided by Harold "Skip" Paul, Masood Rasoulian and Mark Morvant of LTRC, is gratefully acknowledged and appreciated. The assistance provided by personnel from LaDOTD, the Project Review Committee, Savage Industries and the Louisiana Forestry Association is also acknowledged and appreciated.

The contents of this study reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Louisiana Department of Transportation or the Louisiana Transportation Research Center. This paper does not constitute a standard, specification, or regulation.

References

"AASHTO LRFD Bridge Design Specifications," American Association of State Highway and Transportation Officials.

"Bridge Design Manual," LA-DOTD Fourth Edition.

Ghosn, M.; Schilling, C.; Moses, F.; and Runco, G. (1995); Bridge Overstress Criteria, Publication Number FHWA-RD-92-082, Federal Highway Administration.

Moses, F. (1992); Truck Weight Effects on Bridge Costs, Publication Number FHWA/OH-93/001, The Ohio Department of Transportation.

National Cooperative Highway Research Program Report 495.

"New Trucks for Greater Productivity and Less Road Wear." Special Report 227, Transportation Research Board, National Research Council.

Noel, J., James, R., Furr, H., Bonilla, F., (1985) "Bridge Formula Development." Report No. FHWA/RD-85-088.

Roberts, F., Saber, A., Ranadhir, A., Zhou, X., (2005) "Effects of Hauling Timber, Lignite Coal, Coke Fuel on Louisiana Highways and Bridges". Louisiana Transportation Research Center Report Number 398.

Saber, A., Roberts, F., (2006) "Economic Impact of Higher Truck Loads on Remaining Safe Life of Louisiana Bridges", Transportation Research Board 85th Annual Meeting, Washington, D.C.

Tests of Automated Incident Detection with Video Image Processors in Attica Tollway Tunnels

Bill M. Halkias, MSCE, P.E.,[1] Kostas Papandreou, MSCE,[1] Pantelis Kopelias, PhD,[1]
Vily Vegiri, BSCE[1] and Panos D. Prevedouros, PhD,[2]

1 Attikes Diadromes, S.A., 41.9 km Attiki Odos, 190 02 Peania, Athens, Greece
2 Department of Civil and Environmental Engineering, University of Hawaii at Manoa, 2540 Dole Street, 383, Honolulu, HI 96822 (pdp@hawaii.edu)

ABSTRACT

European regulations for the management of roadway tunnels became tricter after the 1999 Mount Blanc inferno. Several video image processing (VIP) devices for traffic surveillance have automated incident detection (AID) capabilities. Autoscope, Citilog and Traficon were tested with eight pre-existing cameras in Attica Tollway tunnels. The suppliers of the competing devices were responsible for setting and calibrating the devices for best results. No literature was found in which VIPs developed in year 2000 or later were evaluated for incident detection. The results show promise for the future, but the evaluated performance was poor. This is due to three reasons (1) low maturity of the technology, (2) complex algorithms due to the provision of extensive functionality, (3) suboptimal camera location and height for image processing, but VIP devices must adapt to tunnel limitations.

INTRODUCTION

Large scale video-based traffic surveillance equipment installations started in the early 1990s and they accelerated in development roughly in parallel with Intelligent Transportation Systems (ITS) and the Internet. Substantial efforts on extracting traffic parameters from analog video were made in the 1980s, and by the early 1990s there were several devices that could produce volume, occupancy and speed measurements from traffic video. Progressively, devices started implementing incident detection technology to their image processing technology.

Partly due to the 1999 accident in Mount Blanc which claimed 39 lives, European regulations for the management of roadway tunnels became stricter. In response to these requirements, the manufacturers of video image processing (VIP) devices for traffic surveillance accelerated the development of automated incident detection (AID) algorithms and produced units with such capabilities.

The 65.2 centerline km long Attica Tollway (40.5 miles) was developed between 1997 and 2004 and was the first urban freeway in Greece. The 15.4 km (9.6 mile) spur of Attica Tollway leading to the center of Athens through mountainous terrain has 56 tunnels and cut-and-cover sections which comprise 12% of its length.

The ITS infrastructure of Attica Tollway is similar to that of world-class freeways with integrated management systems: There is continuous monitoring of traffic at the Traffic Management Center (TMC) with 197 CCTV cameras placed every 1000 m (0.6 miles) in open sections and every 125 m (415 ft.) in tunnels and covered sections. Integrated management at the TMC is accomplished with a SCADA system (Supervisory Control And Data Acquisition.) The SCADA system of Attica Tollway uses interactive consoles to assist in controlling the ITS infrastructure and tunnel systems. Tunnel systems include fire stations, fire detection, fire extinguishers, photometers for visibility assessment (opacity), five levels

of lighting, CO and NO_x concentrations, wind speed sensors, fans, emergency phones, CCTV, inductive loops, lane control signs and variable speed limit signs.

To take advantage of the existing CCTV surveillance infrastructure, competing devices for incident detection in tunnels were selected based on their experience and deployments in highway tunnels. Three providers, Autoscope, Citilog and Traficon made installations for testing. The same cameras provided video to all three devices, and the suppliers of the competing devices were responsible for setting and calibrating the equipment for best results. Attica Tollway staff did periodic analyses of data and caused some artificial incidents for testing, but made no changes to the detection parameters.

The objective of this project was to produce best results in anticipation of a possible decision to adopt several such devices for supplementary incident detection on Attica Tollway. This paper describes these tests and results and is among the first to report on the performance of VIP AID. The analysis is presented in two parts: One with incidents recorded by the devices over a period of three months, and another with incidents that were artificially generated by Attica Tollway engineers and staff.

Katrinaki et al., 2003, Prevedouros et al., 2006 provide comprehensive descriptions of the theory about video-based incident detection, the features of video detection devices that are currently in the market, and summaries of past studies on the evaluation of video detection of traffic flow. Literature on evaluating the capability of VID is limited (Martin et al., 2001). Most evaluations focused on the accuracy of basic traffic characteristics including volume, speed, presence, and occupancy. No literature was found in which video-based traffic detection devices developed in year 2000 or later were evaluated for incident detection.

Fully-featured video-based incident detection systems currently in the market include the *Autoscope (Pro Solo or RackVision)* of Image Sensing Systems, Inc., *VisioPaD* of Citilog, S.A., Traficon *VPI/I* of Traficon, N.V., *Vantage Express* of Iteris, Inc., and *VideoTrak 905* of Peek Traffic Systems, Inc.

DEVICES AND DATA COLLECTION

The systems of ISS (Autoscope *RackVision*), Citilog (*VisioPaD*) and Traficon (*VIP/I*) were selected because they had the most applications in tunnels. All are 3U systems (U is a rack sizing measure; 1 U=1.75 inches or 4.4 cm). The Autoscope connects to the supervising server with a serial connection and the other two with an IP connection. Autoscope required the installation of an external DVR (recorder) in order to record images once incidents have been detected by it. All three systems can connect to the existing SCADA of the TMC, but were not connected to the SCADA as part of these tests.

Eight cameras were fixed to an appropriate field of view for the needs of this project. Although it was possible to orient the cameras in ways that minimized the mix of natural and artificial light (e.g., sunlight coming from the end of the tunnel), this was avoided so that interference from sunlight could be evaluated. The field of view selected was appropriate for virtual detection zone setting in all three VIPs and was satisfactory for manual surveillance.

Although all video-based detectors recommend that the camera is over the middle of the cross-section being monitored, this is rarely the case, and the tunnels of Attica Tollway are no exception. All eight cameras are mounted on the right side of tunnels and view departing traffic.

The installations of the incident detection VIP devices concluded with the installation of exclusive workstations for the operation of each of the three systems, in mid-February 2005. Debugging and training of Attica Tollway staff occurred for a few weeks, and the evaluation period started on March 21, 2005. The suppliers were asked to make calibrations remotely or on-site in order to achieve good results. These changes always made a difference and the results of calibration changes are presented herein.

Routine data collection involved the manual verification of all incident alarms generated by the three devices. Recall that with each alarm the devices also recorded the camera images that correspond to it. Generating the 3,013-case database over 96 days of observation was a long and tedious effort. The record for each case includes case number, date, time of day, CCTV number, the analyst's determination on whether the incident was real or not, and the type of incident. Then, the following information was recorded, separately for each device: Whether the incident was detected, the type of incident detected, and the time of day. The type of incident was categorized as stopped vehicle, pedestrian, traffic queue, smoke, opposite direction, object on the road, low speed (congestion), spillage, animal, bicyclist, obstacle or debris on the lane.

ANALYSIS OF SYSTEM RECORDED DATA

In the incident database, the number of incidents is defined as the total sum of those detected by the three systems. Some incidents were detected by all three systems, and others by only one or two of them. It is likely that some true incidents were not detected by any of the three incidents, thus, the total number of true incidents is not known. It is known that 3,013 of all incidents in 96 days of evaluation were recorded by at least one of the VIPs. The 3,013 cases also include false alarms. Once false alarms were visually verified, they were recorded as such, and the remaining incidents are true and form the Database Sum of Incidents (DSI) in contrast to the true Total Sum of Incidents (TSI). Subsequent statistics are based on the DSI and are, necessarily, a best case outcome because DSI≤TSI. Three main indices of performance, detection rate (DR), false alarm rate (FAR) and value ratio (VR) were defined:

$$\mathbf{DR_j} = (TI_j) / (DSI) \qquad \mathbf{FAR_j} = [(AI_j) - (TI_j)] / (AI_j) \qquad \mathbf{VR_j} = (FAR_j) / (DR_j) \qquad (1)$$

where, TI_j = true incidents detected by system *j* and AI_j = all incidents detected by system *j*.

It was a-priori determined that a system is outright unacceptable if VR≥1, and a system would be characterized as good if VR≤0.1. In this database, only one system, and only at one camera produced VR=0.10. The data presented herein present a poor-to-barely-acceptable performance by these advanced VIP detectors and their performance varied widely. The devices were randomly labeled A, B and C. The results were fully disclosed to vendors (each one was revealed the letter corresponding to their device).

The outcomes of the analysis of the 3,013-case database included (1) the derivation of the detection rate (DR), false alarm rate (FAR) and value ratio (VR) for each device for the whole database; (2) the derivation of DR, FAR and VR separately for each CCTV camera to check the consistency of performance in the various locales, (3) results by type of incident, and (4) assessment of the change in performance after each manufacturer intervention and recalibration.

Table 1 summarizes the DR, FAR and VR estimates for the whole database and separately for each tunnel, by CCTV camera. Highlighted cells reflect best and worst performance by each detector. Detector A scored the best VR of 0.1 in CCTV camera No. 139 by combining a detection rate of 66.7% which is good, and a FAR of 9.1% which is high (not good). The overall VR was somewhat similar for systems A and B, but C had a clearly inferior overall performance.

Table 2 gives a breakdown of incidents that occurred but were not detected by each of the three devices. These statistics reveal potential weaknesses in each system. Device A missed a combined 75% of cases of either stopped vehicle or slow speed vehicle. The same sum was 53% for device B, which makes it relatively less sensitive to A in this type of error. However, device B had the most difficulty with objects on the road (28%). The aforemen-

tioned sum was 39% for device C making it the least sensitive to stopped or slow vehicles. On the other hand, device C had a notable weakness in the detection of pedestrians with a share of 38%.

The intervention and re-calibration by the suppliers produced variable results. There was only one intervention for device A. It reduced its high FAR, but the detection rate also decreased. However, VR improved from 2.5 to 1.3. Device B started with a very high FAR which was reduced by two subsequent interventions with no ill effect on the detection rate. As a result its VR improved from 4.6 to 1.9 and then to 1.8. Device C was exhibiting a low detection rate which improved substantially after the first intervention. VR improved from 2.4 to 1.3. Towards the end of the testing period, a major upgrade of the software was installed, resulting in a low detection rate, a high FAR, and a VR of 4.5.

Various factors affected the performance of these systems. In terms of detection rate, devices A and C were significantly affected by the level of volume. A and C also were significantly affected by lighting changes. Device A was affected by weekend drivers whose driving behavior under similar volume levels may be different. All three devices produced significantly different FAR under low, medium and high volume conditions, but weekend/weekday differences were not significant. Natural and artificial lighting had significant effects in the DR of devices A and C, and strongly significant effects in the FAR of all three devices.

ANALYSIS OF ARTIFICIALLY GENERATED INCIDENTS

In the historical recording of incidents analyzed above, the number of true incidents is not known, and the lack of a precise coordination (or drifting) of server clocks precluded analyses of the "mean time to detect" or MTTD variable, which could show which system is quicker in detecting a given incident.

To sidestep these limitations, a number of artificial incidents were orchestrated during low volume periods by having staff perform irregular maneuvers such as stopping or reversing a vehicle, or walking on the pavement. Due to safety concerns, the artificial incidents were conducted during off-peak hours in weekends. Upstream of the incident both a police car and a van with lighted arrows indicating lane closure were deployed, in addition to activating the LED lane closed signs in the tunnel.

The characteristics and results of 34 artificial incidents are detailed in Table 3. Rank represents whether system A, B or C was 1st, 2nd or 3rd in detecting the incident. The results of system B are clearly superior to those of systems A and C. Only systems B and C detected incidents within 20 seconds, on the average. System A took over 50 seconds on the average, and in one case of a stopped vehicle, it took two minutes to identify it as an incident.

CONCLUSIONS

The performance of the devices cannot be characterized as robust. The detection and false alarm rates were significantly different for low, medium and high levels of traffic flow, and for artificial or natural lighting, for all devices tested. There were differences in the performance of the three devices with B being best and C being worst. However, there were cameras for which B was outperformed by either A or C. Therefore, specific devices in specific environments can perform adequately (which also depends on the roadway authority's expectations) if conditions of traffic volumes and speeds, lighting, shape of the tunnel and placement of the camera are more suited to a particular device and its settings. Highway agencies interested in this technology should conduct tests before making a decision to deploy a VIP system for AID. The results show promise for the future, but the current performance is poor. The highly variable and overall poor outcomes in the Attica Tollway study are due to three reasons: (1) low maturity of the technology, (2) ambition to provide an extensive

functionality resulting in complex algorithms, and (3) suboptimal camera location and height for image processing.

Appropriate mounting heights for cameras used with VIPs for the extraction of traffic measurements and for incident detection exceed 6 m (20 ft.) However, CCTV mounting heights above 5 m (15 ft.) are rarely available in tunnels. The mounting heights for the cameras involved in this evaluation were between 4.0 and 4.8 m (13 and 16 ft.) The developers of these devices must take height limitations in tunnels into account and develop suitable algorithms for limited height installations.

REFERENCES

Kastrinaki, V., M. Zervakis, K. Kalaitzakis. (2003) "A Survey of Video Processing Techniques for Traffic Applications." *Image and Vision Computing* 21, pp. 359-381.

Martin, P. T., J. Perrin, B. Hansen, R. Kump and D. Moore. (2001) "Incident Detection Algorithm Evaluation. http://www.ndsu.nodak.edu/ndsu/ugpti/MPC_Pubs/html/MPC01-122/index.html.

Prevedouros, P. D., J. Ji, K. Papandreou, P. Kopelias, V. Vegiri. (2006) "Video Incident Detection Tests in Freeway Tunnels." Paper No. 06-0798. Forthcoming in *Transportation Research Record*, TRB, Washington, D.C.

TABLE 1. Detection Metrics by Each CCTV Camera

CCTV	No. 24	No. 30	No. 48	No. 51	No. 63	No. 67	No. 134	No. 139	Total	
NTI	49	52	117	83	77	35	33	45	491	
A	61.2%	75.0%	28.2%	69.9%	39.0%	**85.7%**	75.8%	66.7%	56.0%	Detection Rate
B	63.3%	65.4%	72.7%	74.7%	59.7%	45.7%	54.6%	**77.8%**	66.6%	
C	53.1%	42.3%	**69.2%**	60.2%	46.8%	14.3%	36.4%	35.6%	50.5%	
A	77.1%	48.0%	52.2%	33.3%	51.6%	54.6%	55.4%	**9.1%**	52.6%	False Alarm Rate
B	76.3%	74.4%	55.0%	41.0%	69.9%	80.3%	**14.3%**	72.2%	65.2%	
C	83.9%	84.2%	59.9%	**54.1%**	85.5%	85.7%	84.2%	64.4%	75.6%	
A	3.37	0.92	1.09	0.50	1.07	1.20	1.24	**0.10**	1.11	Value Ratio (FAR/DR)
B	3.23	2.91	1.22	0.69	2.33	4.06	**0.17**	2.60	1.87	
C	5.19	5.32	1.49	**1.18**	5.92	6.00	5.33	1.81	3.10	

NTI = Number of True Incidents. Extreme rates are highlighed: best in bold and worst in gray cells, per device and metric.

TABLE 2. Incidents That Occurred but Were not Detected, by Incident Type

		Incident Type								
VID Systems	Un-detected *Incidents*	Stopped Vehicles	Pedestrian	Traffic Queue	Smoke	Opposite Direction	Object on the Road	Low Speed	Spillage (liquid, sand)	Animal
A	216	47.7%	11.1%	4.6%	0.5%	3.7%	4.2%	26.9%	0.5%	0.0%
B	164	28.7%	16.5%	0.6%	0.0%	0.0%	28.0%	23.8%	0.6%	0.6%
C	243	32.1%	37.9%	0.4%	0.4%	3.3%	17.3%	7.0%	0.0%	1.6%
Incident Total	**491**	**44%**	**22%**	**5%**	**0%**	**2%**	**10%**	**14%**	**0%**	**1%**

TABLE 3. Artificially Generated Incidents for Testing the Performance of Video AID Devices

	System A			System B			System C		
Type of Incident Created	**Rank**	**Dura-tion**	**Type of Incident Detected**	**Rank**	**Dura-tion**	**Type of Incident Detected**	**Rank**	**Dura-tion**	**Type of Incident Detected**
slow speed									
stopped vehicle	3	40"	stopped vehicle	2	8"	stopped vehicle	1	5"	stopped vehicle
stopped vehicle									
pedestrian	2	32"	pedestrian	1	30"	pedestrian			
pedestrian				1	15"	pedestrian			
stopped vehicle	2	32"	stopped vehicle	1	15"	stopped vehicle			
opposite direction				1	10"	pedestrian			
stopped vehicle				1	10"	stopped vehicle			
stopped vehicle				2	11"	stopped vehicle	1	10"	stopped vehicle
opposite direction				1	14"	opposite direction			
opposite direction				1	30"	opposite direction			
stopped vehicle									
stopped vehicle				1	10"	stopped vehicle			
opposite direction									
stopped vehicle				1	10"	stopped vehicle			
stopped vehicle				1	25"	stopped vehicle	2	35"	stopped vehicle
opposite direction							1	10"	stopped vehicle
stopped vehicle	3	120"	stopped vehicle	1	10"	stopped vehicle	2	15"	stopped vehicle
opposite direction				1	3"	opposite direction	2	10"	stopped vehicle
stopped vehicle	2	60"	stopped vehicle	1	10"	stopped vehicle			
opposite direction				1	40"	opposite direction			
stopped vehicle	3	60"	slow speed	1	10"	stopped vehicle	2	20"	stopped vehicle
opposite direction				1	5"	opposite direction			
stopped vehicle				1	10"	stopped vehicle	2	15"	stopped vehicle
opposite direction									
slow speed									
opposite direction									
slow speed									
opposite direction									
stopped vehicle				1	7"	stopped vehicle	2	20"	stopped vehicle
opposite direction				1	5"	opposite direction			
pedestrian				1	5"	pedestrian			
pedestrian	2	10"	pedestrian	1	5"	pedestrian			
pedestrian				1	5"	pedestrian			
% Detected			**21%**			**71%**			**26%**
Time to Detection		51 s			12 s			16 s	

A New Algorithm of Traffic Monitoring Based on Real-time Video

X.J. TAN, Z. YU, J. LI, Z. C. HE

Engineering School, Sun Yat-sen University, 135 Xin'gangxi Road, Guangzhou, 510275, China
PH (+8620) 84111938; FAX (+8620) 84113689; email: tanxj@mail.sysu.edu.cn

Abstract

During the past 10 years, different approaches of traffic monitoring have been reported. Most of these methods either rely on a large knowledge-base or are subject to many restrictions. That means they are time-consuming and difficult be carried out in real time. This paper tries to present a new algorithm to solve this problem. We divide the process into two main procedures. Firstly, an improved method is put forward to perform background subtraction, which segments slow moving objects from an illumination-invariant background. Secondly, object tracking is performed with an improved particle filter, which can avoid the difficult problem of matching. The proposed approach has been tested on several highways and satisfactory results have been obtained.

Keywords

Traffic Monitoring, Motion Segmentation, Vehicle tracking

1 INTRODUCTION

Real-time traffic information, such as traffic volume, is important for traffic control or transportation management systems. For the past several decades, the inductive loops and microwave detectors have been put into use. Meanwhile, the image sensors are widely used in recent years and many efforts have been made to develop practical algorithms to process the image information.

Motion segmentation and object tracking are the most important procedures in the traffic monitoring algorithm. In the past decade, many schemes have been proposed. There are two main methods to detect a vehicle's appearance and motion, which are detecting brightness deviation at a line or at dispersed sampling points [1][2] and extracting a vehicle's spatial outline feature and tracking it [3][4]. For segmentation of the moving vehicles, many methods have been proposed, such as differentiating the consecutive video frames [5] and analyzing the optical flow [6]. To avoid the error caused by repeatedly counting the same car, a tracking procedure is often performed after segmentation. The basic idea of tracking is matching the same object among the frame series. This is usually a hypothesis-testing process. Classically, the prediction is made with Kalman filters [7]. The disadvantage of this

lies in two aspects. For one thing, Kalman filters can only be applied only to Gaussian densities. For the other, each object to be tracked requires a Kalman filter, which might lead to intolerable computational complex.

In this paper, we try to present a new algorithm of traffic monitoring. At first, we segment moving vehicles with a motion segmentation process which is improved from the conventional technique called background differencing [8]. Then, a tracking procedure is performed. We track vehicles with the particle filter, which overcomes the disadvantages of the Kalman filter.

The rest paper will be arranged as follows. In Section 2, a new algorithm of motion segmentation will be presented. Then object tracking is discussed in Section 3. Section 4 describes the experiments and results.

2 MOTION SEGMENTATION

In this section, an algorithm is presented for motion segmentation, which is based on the considerations below. Firstly, the cameras capturing the video frames are stationary and motion segmentation can be done by the difference between the current frame and the reference background. Secondly, the illuminant of the scene changes from time to time and the reference background must be updated in real-time. Thirdly, in the traffic scene, the pixel difference caused by the vehicles is much larger than that caused by noise.

2.1 Extraction of the reference background

Let I_t be the luminance of the pixel (i, j) and B_t is the value of the reference background with the same coordinate. The simplest form of reference background can be the average of the gray-values of the previous N frames, as

$$B_t = \frac{1}{N}\sum_{\tau=t-1}^{t-N} I_\tau \tag{1}$$

The disadvantage of Eq.1 lies in two aspects. For one thing, the objects moving slowly in the video scene might be mistakenly extracted as the reference background, which will cause the results of the segmentation severely invalid. For the other, it requires lots of memory buffers to save the previous N frames of the video. To overcome these deficiencies, a modified scheme can be

$$B_{t+1} = B_t + (\alpha_1(1 - M_t) + \alpha_2 M_t)D_t \tag{2}$$

where M_t is a binary value determined by the motion property of the pixel. If pixel (x, y) is moving at time t, then $M_t = 1$, or else $M_t = 0$. The difference of the current frame and the reference background is denoted as $D_t = I_t - B_t$. α_1, α_2 are two constants controlling the speed of the background extraction. However, Eq.2 depends heavily on the accuracy of the motion segmentation. If a moving object is judged as *still* by mistake, then a false-negative error happens and the reference background will be greatly blotted.

In this paper, we try to calculate the background with a method improved from the form of Eq.2. Our method considers the color channels of the video frames. Denoting the RGB values of the pixel (x, y) as $I_R(t), I_G(t), I_B(t)$ and the

reference background at the same coordinate as $B_R(t), B_G(t), B_B(t)$, we have the instantaneous difference at the coordinate

$$D(t) = \sqrt{[I_R(t) - B_R(t)]^2 + [I_G(t) - B_G(t)]^2 + [I_B(t) - B_B(t)]^2} \tag{3}$$

Then the reference background of the next frame can be computed with $D(t)$. Among the three-color channels, the reference background in red is

$$B_R(t+1) = (1 - \frac{1}{D(t)})B_R(t) + \frac{1}{D(t)} I_R(t) \tag{4}$$

The reference background of the other two (Green and Blue) channels will have the same form as the equation above. In Eq.4 $D(t)$ characterizes the intension of the change. The more the color is changed, the less the current frame is extracted as the reference background.

2.2 Motion segmentation with a dynamic threshold

In most of the traditional methods, motion segmentation is performed with a fixed threshold, which means if the absolute value of the difference exceeds a threshold, the pixel should be considered with a motion. Here we adopt a dynamically computed threshold. The motion property can be judged with the inequations below. Color channels are considered respectively.

$$\left| \frac{I_{R,t}(i,j)}{k_t(i,j)} - B_{R,t}(i,j) \right| > [4 - n_t(i,j)] \cdot \alpha + \beta \cdot \sigma_{R,t}(i,j) \tag{5}$$

$$\left| \frac{I_{G,t}(i,j)}{k_t(i,j)} - B_{G,t}(i,j) \right| > [4 - n_t(i,j)] \cdot \alpha + \beta \cdot \sigma_{G,t}(i,j) \tag{6}$$

$$\left| \frac{I_{B,t}(i,j)}{k_t(i,j)} - B_{B,t}(i,j) \right| > [4 - n_t(i,j)] \cdot \alpha + \beta \cdot \sigma_{B,t}(i,j) \tag{7}$$

The inequations of (5) ~ (7) can be used for motion segmentation. If one of the three inequations is satisfied, the pixel of (i, j) can be segmented as a moving point, or else the pixel is considered to be static. The three inequations can be interpreted as below.

(1) α, β are both constants that can be determined empirically.

(2) $n_t(i,j)$ is a connection factor whose value is determined by the motion properties of the 8 neighbor pixels. As the pixel (i, j) is concerned, the more pixels around it is moving, the less the threshold of the segmentation will be, hence the easier (i, j) is judged as a moving pixel. Further discussion can be referred to [9].

(3) $\sigma_t(i,j)$ is a standard variance, as

$$\sigma_t(i,j) = \sqrt{\sum_{\tau=1}^{t} [I_\tau(i,j) - \mu(i,j)]^2} \tag{8}$$

where $\mu(i,j)$ is the mathematical expectation of $I_\tau(i,j)$.

(4) $k_t(i, j)$ is a factor based on the consideration of the illuminant variance. The idea of this factor can also be referred to [9].

3 VEHICLE TRACKING

3.1 Basic ideas of object tracking

Similar to motion segmentation, the tracking procedure is also oriented to the moving objects. However, *tracking* is more concerned about the position of the moving object while *motion segmentation* focuses on whether there is a moving object on the scene. Generally, tracking vehicles in video frames comprises two problems, target representation and target dynamics.

Many models have been proposed to represent a moving object. The 3-D model and the active-contour model can now be regarded as classical ones. In recent years, color based methods and feature based methods have been used. In this paper, we are inclined to utilize the region based method which has already been successfully used in many large sized rigid objects.

The function of target dynamics mostly lies in fusion of the temporal and spatial information. Because the accuracy of visual tracking is inevitably affected by different factors such noises and occlusions, we need to build the target dynamic model to combine the current observation and the history. Sometimes a target dynamic model is also called a 'Hypothesis & Test' model. In this paper, we use the particle filter to perform the target dynamics.

3.2 Target representation

We use a state vector to describe the rectangle region of a vehicle, such as

$$\mathbf{x}_t = [x, y, w, h, v_x, v_y]^T \tag{9}$$

where (x, y) are the center of the rectangle and (w, h) the length of the two sides. Finally, (v_x, v_y) represent the movement of (x, y).

Initialization of the state vector can be done with the result of motion segmentation. We can firstly define the initialization area in the scene. When the vehicle enters the area, it causes a cluster of motion pixels, which can be used to initialize the state vector (particle).

3.3 Target dynamics

The dynamic model of the particle filter can be presented as

$$\mathbf{x}_t = A\mathbf{x}_{t-1} + B\mathbf{w}_t \tag{10}$$

where A, B are the matrices representing the dynamics, $\mathbf{w}_t$ is a random vector representing the noise. In practice, A is defined by the average movements of the targets. The energy of the noise can be computed with BB^T.

4 EXPERIMENTS AND RESULTS

The traffic monitoring algorithm can be regarded as the integration of the two procedures of motion segmentation and vehicle tracking.

At first, we tested the effect of background learning and motion segmentation, the results are shown below in Figure 1.

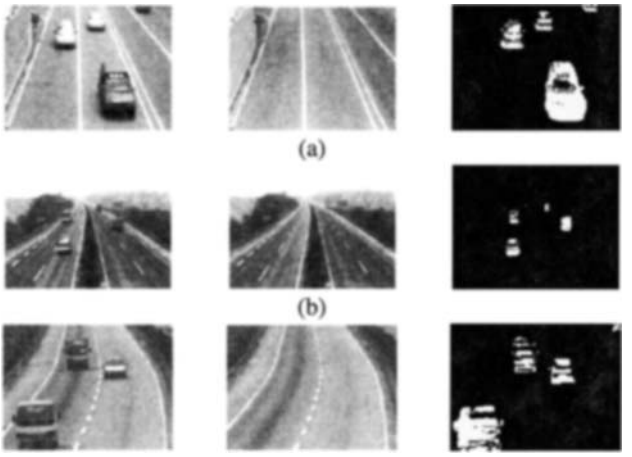


Figure 1 Results of motion segmentation

Secondly, we tested the process of vehicle tracking, the results are shown below in Figure 2.

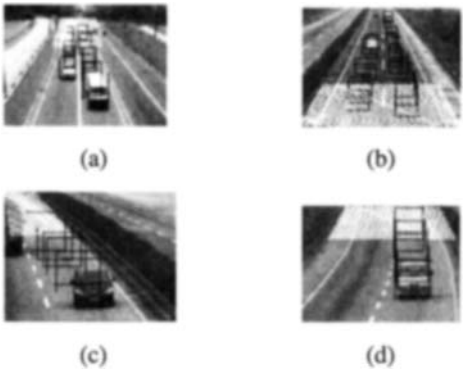


Figure 2 Tracking vehicles with particle filter

Finally, we combined the two procedures as an integrated method and used it to count vehicles with real-time video. The program was written in C++ language and executed on a PC with a CPU of P4-1.5GHz. The frame rate of each camera was 15 fps. The tested video samples were taken on different traffic load. The results are shown in Table 1, where the ground truth was manually defined for each sequence.

Table 2. Results of the experiments

Video Section	Description	Length	Ground truth	Output of the program	Errors
1	Free flow	10:00	56	60	4
2	Heavy congestion	10:00	121	132	11
3	Mix of free flow and congestion	10:00	86	90	4

REFERENCES

[1] J. Sol, B.T. Chun and M. Wang, "Analysis of road image sequences for vehicle counting", in *Proc. Systems, Man and Cybernetics. 'Intelligent Systems for the 21st Century', IEEE International Conference*, **Vol. 1,** (679 –683), Vancouver, Canada, 1995.

[2] Z. Yin, L. Cao "Traffic Parameters Detection Based on Video Virtual Loop Technologies", in *Proc. The Seventh International Conference on Applications of Advanced Technology in Transportation*, (885~893), Massachusetts, USA, 2002.

[3] J. van Leuven, M.B. van Leeuwen, and F.C.A. Groen, "Real-time vehicle tracking in image sequences", in *Proc. IEEE Instrumentation and Measurement Technology Conference*, Budapest, Hungary, 2001.

[4] J.B. Kim, C.W. Lee, K.M. Lee et al., "Wavelet-based vehicle tracking for automatic traffic surveillance", in *TENCON 2001 Proceedings of IEEE Region 10 International Conference on Electrical and Electronic Technology*, **vol.1**, (313~316), Singapore, 2001.

[5] N. Friedman and S. Russell, "Image Segmentation in Video Sequences: A Probabilistic Approach", in *Proc. the 13th Conference on Uncertainty in Artificial Intelligence*, Providence, USA, 1997.

[6] Y. Mea, Y. Shirai, J. Miura et al., "Object Tracking in Cluttered Background Based on Optical Flow and Edges", in *Proc. Fourteenth International Conference on Pattern Recognition,* **vol.2**, (1439~1441), Brisbane, Australia, 1998.

[7] D. Koller, J. Weber, T. Huang et al. "Towards Robust Automatic Traffic Scene Analysis in Real-time", *in Proc. the International Conference on Pattern Recognition*, Israel, 1994.

[8] C. Stauffer and W.E.L. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking", *in Proc. Computer Vision and Pattern Recognition*, (246~252), 1999.

[9] X. J. Tan, W. Shen and Z.H. Guo, "Vision-based method for traffic flow surveillance on highway", Journal of Computer Applications (in Chinese), 25(5), 1215-1218.

A Video-based Method for Evaluating Traffic Data from Detectors

Nan Zou[1] and Jianwei Wang[2]

[1,2]Department of Civil and Environmental Engineering, University of Maryland, College Park, 20742, USA. Phone: 301-405-2638. Email: nanzou@umd.edu

Abstract:

This paper presents a video-based method for evaluating and validating volume and speed data collected with traffic detectors. Assessing the detector data reliability is a critical task for all Intelligent Transportation Systems (ITS). Their performance will be significantly impacted by the data quality. Most existing studies mainly use volume as the only variable for evaluating the detector quality. A cost-efficient and rigorous method that can concurrently evaluate both volume and speed from the detectors is not available yet.

The video-based method presented in this study is both cost-efficient and sufficiently reliable for evaluating detector data for ITS system operations. For example, the performance of a travel time prediction system is very sensitive to the quality of the detector data. This paper will mainly detail the key features of the proposed system and its application in a case study of RTMS detector data on I-70 corridor in Maryland.

Background:

With the fast developments of Intelligent Transportation Systems (ITS), traffic detectors that provide traffic count, speed and occupancy have become increasingly popular. Regardless of the employed technologies, most detectors need to have their internal parameters calibrated property. Hence, how to efficiently and effectively evaluate the functioning of available traffic sensors has emerged as one of the essential tasks in the traffic control, especially in view of providing the real-time operational data. In review of related literature, it is noticeable variety of methods for detector data evaluation exist (1). Most of those focus mainly on volume count. A reliable and efficient method for concurrently evaluating the volume and speed data still remains to be developed.

Video-based methods have been widely used in the fields of the vehicle detection and the movement tracking (2, 3, 4). However, the speed measurement using video-based method has been found to be unreliable with traditional analog video capturing devices for short travel distances in the literature. Kuo and Machemehl (5) analyzed the probability of measurement errors and their relations with video frame rate and measurement distance, and found the speed trap distance needs to be at least 15 meters long for one of the scenarios. Bonneson (6) enlarged this distance to 20 meters after further analysis on the potential bias and errors. With such a long required distance, video-based method has been found to be not suitable for measuring speed in a segment where vehicles change speed frequently.

With the development of high quality digital video capture devices, the image quality has been significantly improved. At the same video recording frame rate of about 30 fps, the digital devices provide higher resolution (720 by 480) and sharper

edges of the objects in the video, which is one of the main concerns that contribute to measurement errors. Therefore, the analysis results in the previous literature that are based on traditional analog devices do not hold for the new devices. Furthermore, computer technologies have been playing more and more important roles in helping to improve the data measurement. This paper takes consideration of the abilities of new digital video capturing systems and evaluates their performance on the video-based speed measurement with the help of a developed video-processing program. The proposed video-based method is cost-efficient and reliable for validating the speed data from detectors for all ranges of speed. The potential errors have been carefully analyzed in this study, followed by a case study on I-70 corridor in Maryland.

System Features and Operational Procedures

System Components

The developed video-based evaluation system includes the following components:

- One MiniDV camcorder with the ability to record videos with the resolution of 720x480 pixels per frame and the recording rate of 29.97 frames per second with its 20x optical zoom lens;
- One heavy duty tripod, which is 72 inches high, to place the camcorder on the road side stably;
- Highlight spray to mark the parallel reference lines on the road side;
- A GPS receiver that is able to display time acquired from the satellites;
- A developed video-processing program with the ability to high light and extend reference lines and record timestamps when individual vehicle passes each reference line.

Field Survey Procedures

Step 1: Pre-survey preparation:

a) Select an optimal location of the camcorder considering the geometry constraints;
b) Mark the parallel reference lines on the roadside;
c) Measure the distance between parallel reference lines;
d) Check the time difference between the detector and the GPS satellites;

Step 2: Video recording

a) Adjust the zoom of the camcorder to make sure the survey segment occupies almost the entire screen
b) Start video recording and then place a GPS receiver with 3D fix status to show the time acquired from the satellite in front the camcorder for few seconds.
c) Try to avoid the vibration when recording the video
d) Record the GPS time every time a new recording starts after change of battery or tapes.

Step 3: Download the traffic data from the detector and then update the timestamps based on the time difference collected on step 1.d

Data Processing Procedures

Step 1: Capture the un-compressed video by using an IEEE1394 firewire card with video-editing software.
Step 2: Load the video in the developed software
Step 3: Sync the video time with the time on the GPS receiver shown at the beginning of the video
Step 4: Use the software function to highlight and extend the reference lines in the video to cross the road segment
Step 5: Use function keys to record the timestamps of individual vehicle passing each extended reference lines on the screen.

Analysis of System Reliability

Concept of the System

The concept of the system is to compute the average speed v at which one vehicle passes a segment by dividing the length of the segment L with the travel time t.

$$v = \frac{L}{t}$$

With the consideration of measurement errors, the computed speed v' can then be expressed as:

$$v' = \frac{L'}{t'} = \frac{L + \varepsilon_l}{t + \varepsilon_t}$$

Where L' is the measured distance and t' is the measured travel time. ε_l represents the measurement error of distance and ε_t represents the measurement error of travel time.

Measurement Error of Travel Time

With the developed video-processing software, the location of individual vehicle can be traced every 1/29.97 seconds at the frame rate of 29.97fps. When one vehicle is traveling at 65mph, the travel distance between two adjacent frames is 3.18 feet. Therefore, it is very frequently seen in the video that one vehicle's front wheel is before or after the reference line in two adjacent video frames. In order to estimate the time spot when vehicle actually steps on the line, the following rules were developed:

Distance from vehicle's front wheels to the reference line refers to the distance from the spot at which vehicle's visible front wheel touches the ground to the reference line along the direction of the road. If one vehicle's front wheels are before the reference line at frame n and after the reference line at frame n+1, then the timestamp t' is determined as follows:

$$t' = \begin{cases} t_n & , \text{when } | s_n | \leq | s_{n+1} | / 2 \\ t_{n+1} & , \text{when } | s_{n+1} | \leq s_n / 2 \\ (t_n + t_{n+1})/2 & , \text{otherwise} \end{cases}$$

Where t_n is the timestamp of frame n and s_n is the distance from vehicle's visible front wheel to the reference line at frame n.

$$t_{n+1} - t_n = \frac{1}{29.97} \text{ seconds}$$

$$s_n \cdot s_{n+1} \le 0$$

Based on the experience, mistakes may be made when

$$\frac{1}{5} \le \left|\frac{s_n}{s_{n+1}}\right| \le \frac{1}{2} \text{ or } \frac{1}{5} \le \left|\frac{s_{n+1}}{s_n}\right| \le \frac{1}{2}$$

With consideration of possible errors at both reference lines,

$$|\varepsilon_t| \le \frac{1}{29.97} \times \frac{2}{3} \approx 0.0222 \text{ seconds}$$

Measurement Error of Distance

As illustrated in Figure 1, two parallel 2-inch-wide reference lines with approximate lengths of l_R are marked on the road side. The distance between two marked reference lines is measured as L_M. The distance between two extended reference lines in lane i is L_{Ti}. The following issues may impact the computation results:

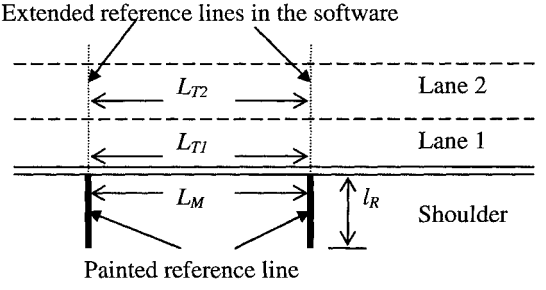


Figure 1. Illustration of the survey site

- ε_M, the measurement error of L_M
- ε_A, the difference between L_{Ti} and L_M

$$\varepsilon_L = \varepsilon_M + \varepsilon_A$$

ε_A is caused by the angel variation when extending the marked reference lines in the software. The possible variation of angle is determined by the width and the length of actual marked reference lines. When the reference lines are both 2-inch wide and 12 feet long, the maximum error can then be calculated as follows, assuming the width of the lane is 12 feet and vehicles' visible front wheels are within 4 feet from the lane marker at the same side:

$$|\varepsilon_A| \le \frac{2}{12} \times 75\% \times (12i + 4) - 1 = 1.5i - 0.5 \text{ inches}$$

Assuming ε_M is less than 0.5 inch, then $|\varepsilon_L| \le 1.5$ inches when the computation is for lane 1, the maximum relative errors of speed for different pairs of the actual speed and the segment length are shown in Table 1.

Max. Possible Relative Error		Actual Speed (mph)					
		10	20	30	40	50	60
Segment Length (feet)	20	3.58%	4.45%	6.24%	8.09%	10.01%	12.00%
	40	1.77%	2.19%	3.04%	3.91%	4.79%	5.69%
	60	1.17%	1.45%	2.01%	2.58%	3.15%	3.73%
	80	0.88%	1.08%	1.50%	1.92%	2.35%	2.78%
	100	0.70%	0.87%	1.20%	1.53%	1.87%	2.21%
	120	0.58%	0.72%	1.00%	1.27%	1.55%	1.83%

Table 1. Distribution of maximum possible relative errors

Case Study

A case study has been conducted for detector No.5 from ARAMPS (An **A**utomated **R**eal-Time **Tra**vel **Ti**me **P**rediction **S**ystem) project on I-70 eastbound in Maryland.

All 1,812 vehicles in lane 1 (the right lane) at detector 5, which covers a 2-lane freeway segment, were analyzed with the proposed video-based method with two reference lines 40.7 feet apart during the time period from 7:35AM to 8:45AM on 11/29/2005. A radar speed gun was used to measure some vehicles in the same time period.

Video-based Speed Measurement vs. Speed Gun: Comparison of Individual Vehicle

Speeds of 50 vehicles picked randomly were measured with a radar speed gun from 8:10AM to 8:23AM. Later, each of those 50 vehicles was identified in the video. A comparison of speeds from the video-based method and the speed gun was made for all 50 pairs. The speeds from the video-based measurement method range from 22.15 mph to 64.74 mph. The comparison between video-based method and speed gun measurement is shown in Table 2. The results indicate two sets of speed data match each other fairly good. Please note that the data from the speed gun is for reference only because the location of speed measurement may vary slightly between different observations.

Speed Category from Video	Average	Standard Deviation	Max. Positive Difference	Max. Negative Difference	Count
20-35	0.00	1.24	1.56	-1.88	9
40-55	-0.77	1.53	1.70	-3.31	13
>55	0.30	1.81	4.50	-2.13	28

Table 2. Speed Differences between Video-based Measurement and Speed Gun (mph)

Comparison between All Three Methods

Speeds have also been measured by a speed gun for most vehicles for about 14 minutes (8:01AM to 8:08AM and 8:25AM to 8:32AM). A total of 380 vehicles were detected by the detector of which 337 vehicles (88.68%) were measured by the speed gun. The video-based method detected 385 vehicles in the same time period. Each single measurement point from the speed gun was recorded with a timestamp from a handheld GPS receiver. The speeds from the video-based method and the speed gun have been averaged to match 27 30-second intervals in the 14-minute period from the detector. All three methods show consistent speed range when speeds are above 55 mph (Table 3).

Speed Category from the Video	Average Speed (mph)			# of Intervals
	Video-based	Speed gun	Detector	
50-55	53.91	53.09	58.33	3
55-60	57.51	56.94	59.44	16
>60	60.73	59.66	62.63	8

Table 3. Average Speeds from Three Measurement Methods (mph)

Video-based Measurement vs. Detector

Finally, 140 30-second intervals were determined in the video to match the intervals from the detector for the entire data collecting period from 7:35AM to 8:45AM. As shown in Table 4, detector data was close to those from the video-based method when speeds from the video are above 55mph, which is the speed range used to make the latest speed calibration on the detector. However, large variances have been found when speeds from the video are less than 55mph. The results indicate further calibration for lower speed conditions is needed for this detector to provide reliable speed data for the travel time prediction system.

Speed Category from Video	<25	25-35	35-45	45-55	>55
Average Difference	14.04	17.93	15.05	8.48	3.38
Standard Deviation	12.96	3.27	2.37	3.31	3.16
Max. Positive Difference	32.45	22.50	40.50	13.65	9.38
Max. Negative Difference	-2.41	-8.25	-35.03	-2.08	-4.59
Interval Count	5	34	23	27	47

Table 4. Speed Differences between Video-based Measurement and Detector (mph)

Conclusions

A video-based speed detection system, which is cost-efficient and reliable, has been developed in this study. Based on the detailed analysis on possible errors, the system is proved to provide reliable speed detection for all speed range under detection conditions mentioned in this paper. A case study shows that the system output is consistent with speed gun measurements and large variances have been identified from one detector when speed is less than 55 mph. The result indicates a further speed calibration is required for that detector to provide reliable speed data.

Reference:

1. Weber, N. A., "Verification of Radar Vehicle Detection Equipment", Technical Report SD98-15, South Dakota DOT, 1999
2. Bonneson, J., and Fitts, J., "Traffic Data Collection Using Video-based Systems," Transportation Research Record 1477, pp. 31-40, 1995
3. Setchell, C., "Applications of Computer Vision to Road-traffic Monitoring", Ph.D. Thesis, University of Bristol, 1997
4. Strong, C., Lowry, S., and McCarthy, P., "Collecting Vehicle-speed Data by Using Time-lapse Video Recording Equipment", Transportation Research Record 1855, pp 97-104, 2003
5. Kuo, C., and Machemehl, R., "Probabilistic Speed-Estimation Measurement Errors Through Video Images," J. Transp. Engrg., Vol. 123(2), pp. 136-141, 1997
6. Bonneson, J., Discussion: probabilistic Speed-Estimation Measurement Errors through Video Images, J. Transp. Engrg., Vol. 124(6), pp. 611-613, 1998

Uses of Airborne Imagery for Microscopic Traffic Analysis

Mark Hickman[1] and Pitu Mirchandani[2]

[1] Department of Civil Engineering and Engineering Mechanics, University of Arizona, P.O. Box 210072, Tucson, AZ 85721-0072, PH (520) 626-9420, email: mhickman@engr.arizona.edu.

[2] Department of Systems and Industrial Engineering, University of Arizona, P.O. Box 210020, Tucson, AZ 85721-0020, PH (520) 621-7284, email: pitu@sie.arizona.edu.

Abstract

Over the past several years, traffic analysts and researchers have been exploring the use of airborne imagery to understand traffic behavior and to analyze traffic speeds, volumes, densities, delays, queue lengths, and other traffic parameters. As part of the National Consortium on Remote Sensing in Transportation - Flows (NCRST-F), researchers at the University of Arizona have been developing software tools to assist in analyzing traffic behavior, at the level of the individual vehicle (or, "microscopic" traffic behavior). To do this, we have gathered digital video segments taken from a helicopter over traffic facilities, using a digital video camera. The video image sequence is then processed using software TRAVIS (Tracking and Registration of Airborne Video Image Sequence). The software registers the video sequence (eliminating the movement of the aircraft from the image), identifies vehicles in the images, and tracks the vehicles through the video image sequence. The vehicle coordinates in the image sequence can then be transformed into relevant traffic measures: location, speed, acceleration and deceleration, and position in terms of its lane on the roadway. With a large set of video segments, the software has been able to identify and track individual vehicles automatically, thereby significantly reducing the need for manual data reduction for some traffic applications. We also discuss the applications of the data from airborne imagery for understanding microscopic traffic flow and for the development and calibration of traffic simulation models.

Introduction

Over the past several years, traffic analysts and researchers have been exploring the use of airborne imagery to understand traffic behavior. The advantages of using an airborne platform, in contrast to a fixed platform mounted above or near the roadway, are numerous. First, the platform is mobile, allowing flexibility in scheduling data collection at a given site and allowing a single platform to collect data at a number of different sites, even within the same flight or on separate flights. Because the video is taken at approximately nadir (directly above the roadway), various image distortions and occlusions that are common in fixed platforms can also be avoided.

Analysts have used such imagery to analyze traffic speeds, volumes, densities, delays, queue lengths, and other traffic parameters. Angel et al. (2002, 2003) provide a preliminary summary of much of the early uses of airborne imagery (still imagery and video) for traffic surveillance. More recently, there has been considerable research in this area as part of the National Consortium on Remote Sensing

of Transportation – Flows (NCRST-F). As examples, Angel and Hickman (2002) used airborne video to analyze travel times over a 3-mile arterial highway, looking at both average travel times and the variance of travel times among vehicles in a platoon. In a subsequent paper, Angel and Hickman (2003) used airborne video captured while a helicopter hovered over a signalized intersection. This allowed direct computation of delays and level of service at these intersections. These applications required manual data reduction from the video. In contrast, Mirchandani et al. (2002) applied an automated image processing technique to derive vehicle speeds from a sequence of still images from a roadway. Their technique runs in real time, allowing the estimation of vehicle speeds directly. In addition, Agrawal and Hickman (2004) used still images taken from an airborne platform to measure vehicle queues at signalized intersections. They present a method to derive the vehicle queues by identifying the individual vehicles and the spatial extent of the queue, using semi-automated image processing techniques.

A parallel effort within the NCRST-F at the University of Arizona has examined automated methods of processing airborne imagery to derive individual vehicle trajectories. A method based on the Kanade-Lucas-Tomasi (KLT) feature tracking algorithm (Shi and Tomasi, 1994) was proposed by Shastry (2002) and Shastry and Schowengerdt (2002). This method registers the imagery to a single image, thereby removing the movement and jitter of the airborne platform. It then tracks specific features in the image that are moving, focusing on vehicle-sized objects, using the KLT feature tracker (Shi and Tomasi, 1994). Following on this research, Kadam (2005) developed a separate tracker, described more fully below. Once individual vehicles can be tracked in the imagery, subsequent processing is necessary to derive actual trajectories. The ability to derive vehicle trajectories in an automated fashion from airborne imagery is an important milestone in traffic data collection. As discussed below, such an achievement this allows researchers to then derive parameters for microscopic traffic behavior and to generate data for traffic simulation models.

Framework

The general process of data collection and processing is depicted in Figure 1. The process begins with data collection over a roadway or traffic facility of interest. A commercial digital video camera, mounted on the aircraft to collect imagery from near nadir, collects the imagery in the form of a digital video. Once collected, the raw digital video is then processed using software called TRAVIS (Tracking and Registration of Airborne Video Image Sequence) (Kadam, 2005). As described in more detail below, the software registers the video to a single image, identifies interesting features (vehicles) in the image sequence, and tracks these vehicles through the sequence. The results of this step are pixel coordinates of the vehicle in the image. Once these coordinates are known, these pixels can then be converted into parameters of individual vehicle trajectories. Specifically, the pixel coordinates can be converted into "ground" coordinates through the use of the image scale. In addition, the position of the vehicle relative to the roadway can be determined by using a "road mask", describing the geometry of the roadway within the image. From the roadway mask, the direction and the specific lane of a vehicle can be determined. Finally, the vehicle position and time (i.e., the trajectory) can then be used for specific traffic applications. Each of the steps is described in more detail below.

Data Collection

The video data collection conducted to date has used a commercial off-the-shelf digital camcorder with a pixel resolution of 720x480 pixels, collecting images at 30 frames (images) per

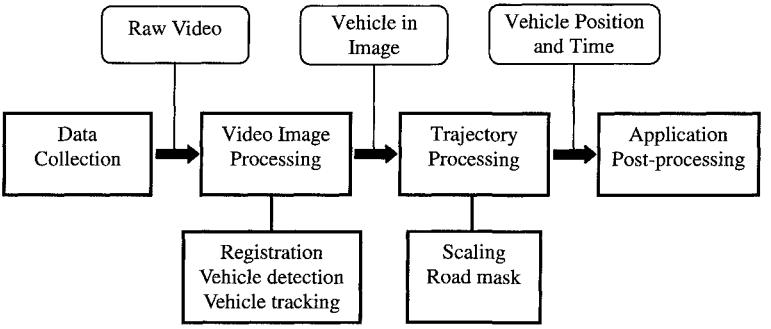


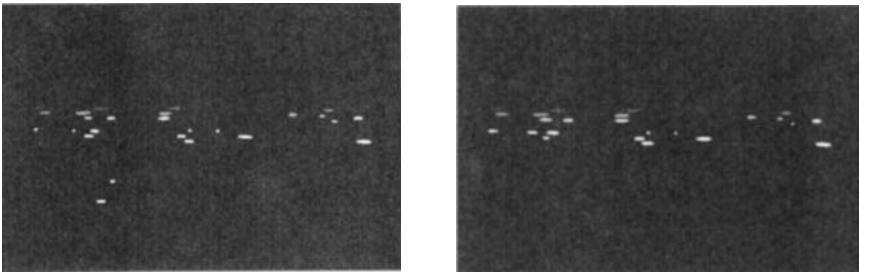
Figure 1: Process of Traffic Data Collection and Processing

second. In many of the experiments, we have flown at heights ranging from 500 ft (160 m) to 2000 ft (650 m) above the surface of the roadway. With a typical focal length on the camcorder, this allows us to capture distances on the ground approximately equal to the height; i.e., from 160 m to 650 m on the ground, in the longer direction of the camcorder's field of view. Generally, this provides a scale ranging from 1-4 pixels per meter (or 0.3-1.3 pixels per foot), or commonly around 2 pixels per meter with 360 m in the field of view. Such a scale allows sufficient pixels for each vehicle in the image so that the vehicle can be easily identified as a "good" feature in the image processing software. However, it also provides a large enough field of view to simultaneously capture numerous vehicles over a roadway segment of interest. More details on the data collection platform can be found in Angel et al. (2003).

Video Image Processing

The input to the video image processing is the sequence of images from the video, which can be large. In the video image processing in TRAVIS, the sequence of images in the video (up to 30 frames per second) are first registered. This step of registration identifies common features from one image that reappear in the second image. These are determined as "ground control points" in the imagery, using the registration technique in the KLT algorithm (Shi and Tomasi, 1994). Once the matching is performed, the algorithm performs an affine transformation of the coordinates of the second image to the first. This removes any movement of the airborne platform (both systematic movement and jitter) from the image sequence, allowing the images to appear as if they were taken from a fixed point of view.

In the second step, two adjacent images are then subtracted. This subtraction operation leaves only moving features in the image. Once these features are identified, the algorithm then identifies the "blobs" in each individual image that generate these moving features. Within each image, the highest difference in the red, green, and blue bands in the image is found, and used to threshold for "blobs" in the images. Simple dilation is also performed, and the objects are then screened using simple heuristics (i.e., using the approximate size and shape of a vehicle). The "blobs" are then labeled in each image, using connected component analysis. It is these features that are then tracked throughout the sequence. An example of adjacent images with connected components is shown in Figure 2.



(a) (b)

Figure 2: Connected Components in Adjacent Images (a) and (b)

The third step of vehicle tracking provides both short- and long-term filtering of these "blobs" to eliminate noise in the image, but to retain the "blobs" that represent moving vehicles. A short-term tracker is used to detect new moving vehicles in the sequence, and to screen out "blobs" that are not actually vehicles. Some small blobs can be seen in Figure 2 that are removed by the short-term tracker. At the same time, new vehicles entering the image can also be tracked: an example of a new vehicle is that on the lower left side Figure 2. The long-term tracker, on the other hand, maintains the coordinates of the vehicle over time. It also allows the tracking of vehicles even when they are not detected for short periods of time (e.g., the vehicles stops moving for a short time, or it is not detected in a short sequence of images).

The net result from the tracker is a sequence of pixel coordinates of vehicles as they are tracked through the image sequence. An example of the video output, in the form of vehicle coordinates, is shown in Figure 3. This illustrates the text output, in the form of pixel coordinates for each object (the rows) and for each frame in the video sequence (the columns).

Trajectory Processing

The subsequent processing of the vehicle coordinates, shown as "Trajectory Processing" in Figure 1, represents an area of ongoing research. The objective is to convert the vehicle "location" in pixels to ground coordinates. The most direct method is to scale the pixels to ground coordinates, using a simple transform (e.g., a linear transform of X meters per pixel). More sophisticated non-linear models can also be used, depending on the camera height, focal length and field of view. With our camera focal length and height, the resulting distortion at the edges of our imagery is small and of the same magnitude of other errors in the image processing. For this reason, a simple linear transformation has been used in the research to date. As yet another option, we have also experimented with a high-resolution camera platform with precision GPS, inertial measurements, and ground surface data; this has shown that vehicle locations and movements can be determined much more precisely. In either case, the result of this pixel-to-ground transformation is the vehicle coordinates in x and y, in meters relative to a given origin on the ground. From this, some basic vehicle parameters can be easily derived: location, speed, and acceleration and deceleration.

To derive more detailed measures of microscopic traffic behavior, the location of the vehicle on the roadway (i.e. its location relative to the traffic lanes) must also be determined. This allows for vehicle position determination not only relative to other vehicles and with respect to the roadway

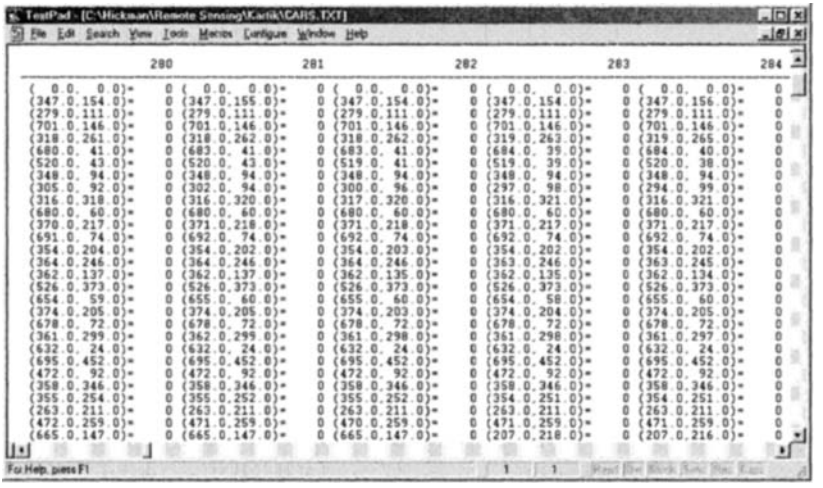


Figure 3: Sample Text Output of Vehicle Coordinates in Pixels

geometry: vehicle spacing in a single lane, lane changes, and other important parameters of the traffic behavior. However, because the airborne platform is moving, the determination of fixed references on the ground (i.e. lane locations) is not trivial. A "road mask" can be applied to the registered imagery, indicating roadway and lane boundaries for the trajectory processing. Such a road mask can be generated by a number of techniques, including: (1) image processing to detect the roadway as a feature in the imagery; (2) the use of existing geo-referencing information, such as roadway centerlines and/or road edges; or, (3) other image processing techniques to isolate the roadway within the imagery. We are also investigating other automated methods using partial vehicle trajectories to determine likely longitudinal motion (and hence lane-keeping behavior) in the imagery. The investigation of methods to generate this road mask is an ongoing area of research.

Applications

Once the vehicle trajectories are known, the data can then be used for a variety of applications. Our research was motivated by the need for microscopic traffic data. Specifically, the processing of vehicle trajectories gives direct information about vehicle acceleration and deceleration behavior and speeds. With more information on vehicle positions in the roadway, car following behavior, lane-changing behavior, merging and weaving behavior, and related traffic phenomena may also be studied in greater detail. The type of data developed in this process can be used to look at individual driver behavior, in a wide variety of situations and on a variety of traffic facilities.

The data provided in this way can lead to research in a variety of areas. Most notably, in can lead to the development of new models of microscopic traffic behavior (e.g., lane-changing behavior described by Toledo, 2003). It may also provide data sets for the calibration and validation of existing and emerging traffic simulation models. As part of the USDOT's Next Generation SIMulation

(NGSIM) program, comparable data sets are already being developed (Cambridge Systematics, 2005) to provide microscopic data for traffic modeling and simulation purposes, but using fixed camera locations. The use of airborne platforms, together with software tools to derive vehicle trajectories as outlined here, provide yet another opportunity to collect and distribute such microscopic traffic data. As we note, the airborne platform provides greater flexibility in the location and frequency of data collection, further diversifying the types of data sets available for traffic modeling.

References

A. Agrawal and M. Hickman (2004). "Automated Extraction of Queue Lengths from Airborne Imagery." Paper presented at the IEEE Intelligent Transportation Systems Conference, October 4-7, 2004.

A. Angel, M. Hickman, D. Chandnani, and P. Mirchandani (2002). "Application of Aerial Video for Traffic Flow Monitoring and Management." In *Applications of Advanced Technology in Transportation, Proceedings of the Seventh International Conference*, C.P. Wang, S. Madanat, S. Nambisan, and G. Spring (eds.). ASCE, Reston, VA, pp. 346-353.

A. Angel, M. Hickman, P. Mirchandani, and D. Chandnani (2003). "Methods of Analyzing Traffic Imagery Collected from Aerial Platforms." *IEEE Transactions on Intelligent Transportation Systems*, Vol. 4, No. 2, pp. 99-107, June.

A. Angel and M. Hickman (2002). "Experimental Investigation of Travel Time Estimation Using Geo-Referenced Aerial Video." National Research Council (U.S.), Transportation Research Board Meeting (81st : 2002 : Washington, D.C.), Preprint CD-ROM.

A. Angel and M. Hickman (2003). "A Method for Analyzing the Performance of Signalized Intersections from Airborne Imagery." National Research Council (U.S.), Transportation Research Board Meeting (82nd : 2003 : Washington, D.C.), Preprint CD-ROM.

Cambridge Systematics (2005). *NGSIM* web site, http://ngsim.camsys.com/, Accessed Dec. 27, 2005.

K. Kadam (2005). Detection and Tracking of Vehicles in an Airborne Video Sequence. MS Thesis, Department of Electrical and Computer Engineering, University of Arizona, May.

P. Mirchandani, M. Hickman, A. Angel and D. Chandnani (2002). "Application of Aerial Video for Traffic Flow Monitoring and Management." *Proceedings of the Pecora 15 Conference* (CD-ROM), November 2002.

A. Shastry and R. Schowengerdt (2002). "Airborne Video Registration for Visualization and Parameter Estimation of Traffic Flows." *Proceedings of the Pecora 15 Conference* (CD-ROM), November.

A. Shastry (2002). Airborne Video Registration and Traffic Flow Parameter Estimation in a Multimedia GIS. MS Thesis, Department of Electrical and Computer Engineering, University of Arizona, December.

J. Shi and C. Tomasi (1994). "Good Features to Track." *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593-600.

T. Toledo (2003). Integrated Driving Behavior Modeling, PhD Dissertation, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, February.

Temporal Variations in Traffic Flow and Ramp-Related Crash Risk

Chris Lee[1] and Mohamed Abdel-Aty[2]

[1]Department of Civil and Environmental Engineering, University of Central Florida, Orlando, FL 32816-2450; PH (407) 823-4902; FAX (407) 823-4676; email: cclee@mail.ucf.edu
[2]Department of Civil and Environmental Engineering, University of Central Florida, Orlando, FL 32816-2450; PH (407) 823-5657; FAX (407) 823-3315; email: mabdel@mail.ucf.edu

Abstract

This study proposes the method of predicting the temporal variation in crash risk on freeway ramps and at the intersections of ramps (the junction of ramps with cross-roads). Using a 5-year ramp-related crash data on a freeway in Orlando, Florida, the study develops probabilistic models that relate the frequency of crashes with the factors such as ramp type (off-ramps and on-ramps), ramp configuration, daily and hourly ramp traffic volume, and 5-minute average speed and volume in the mainline close to ramps 5~10 minutes prior to the time of crashes. The study estimates crash rates (the expected number of crashes divided by ramp traffic volume) as surrogate measures of ramp-related crash risk using log-linear models.

The parameters of the models showed that the frequency of ramp-related crashes generally increased with ramp traffic volume and crash rates were higher on off-ramps than on-ramps. It was also found that higher volume and lower speed at the locations near ramps significantly contributed to higher crash risk. Considering the temporal variation in traffic condition, it was found that crash rates were higher during dawn (12~6 am) compared to the other time periods of day and they were higher on loop and outer connection ramps than diamond ramps.

The study demonstrates how the models can be applied to freeway traffic management in predicting temporal variation in ramp-related crash risk based on ramp geometric and traffic flow characteristics. Findings of the study suggest some traffic control strategies that may mitigate high ramp-related crash risk.

Introduction

A substantial proportion of freeway crashes occur on ramps and in the vicinity of ramps. For instance, approximately 15% of total freeway crashes occurred on ramps in California during 1992-1994 (Khorashadi, 1998). In this regard, many researchers have examined crashes on ramps and interchanges. Lundy (1967) found that crash rates were higher on exit and curved ramps than entrance and straight ramps. Twomey et al. (1993) found that a decrease in spacing between interchanges increased crash rates in urban areas. McCartt et al. (2004) observed that the rates of rollover crashes were higher on off-ramps and the rates of rear-end crashes were higher on on-ramps. They also found

that the common types of crashes differ by the location of crashes and by ramp configuration. Bauer and Harwood (1997) developed negative binomial regression models and showed that ramp length, mainline traffic volume and area type (rural/urban) are also significant factors affecting crashes on ramps. All these studies commonly found that ramp type and configuration are closely associated with crashes on ramps.

However, there exist some limitations in these studies. First, the crashes at the intersections of ramps were excluded from the analysis. However, the vehicles entering on-ramps from cross-roads or exiting to cross-roads via off-ramps can collide with the vehicles traveling in cross-roads at the intersections. Although these crashes do not occur on ramps, they are partially contributed by ramp geometric and traffic factors. Second, in spite of significant variation in traffic flow over time, the studies mostly did not account for actual traffic condition at the time of crashes due to lack of data. However, this actual traffic condition leading to crashes on ramps can be estimated using the traffic flow data obtained from loop detectors in the lanes close to ramps.

The objective of this study is to identify the relationship between crashes on ramps and at the intersections of ramps (defined as "ramp crashes"), and a variety of ramp geometric and traffic characteristics including mainline traffic characteristics in the lane close to ramps. In the crashes at the intersections of ramps, this study only includes those crashes where at least one vehicle entering or exiting ramps is involved.

Description of data

This study used the crash data for the ramps located at a 25-mile stretch of Interstate-4 freeway (I-4). A total of 522 ramp crashes occurred on 98 ramps on this section of I-4 over 5 years, 1999-2003. To identify the factors associated with ramp crashes, detailed information on ramp geometric and traffic characteristics was collected.

Ramp geometric characteristics include the type of ramps and the configuration of ramps that were identified from topographical maps and aerial photographs. Based on the classification in Bauer and Harwood (1997), ramp configuration was classified into 1) diamond, 2) free-flow loop, 3) outer connection, and 4) direct/semi-direct connection as shown in Figure 1. For brevity, free-flow loop, outer connection and direct/semi-direct connection ramps are termed "loop", "outer" and "direct" ramps, respectively.

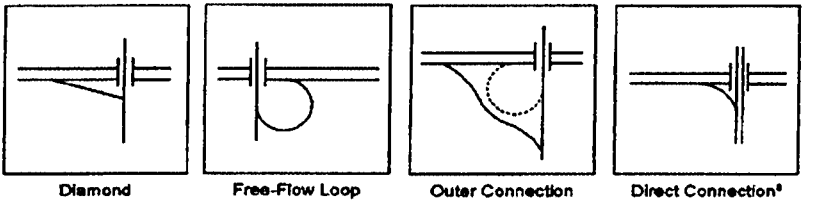


Figure 1. Configuration of ramps. [Source: Bauer and Harwood (1997)]

Traffic characteristics include average daily traffic volume on each ramp and the short-term traffic flow data in the mainline near each ramp. Since ramp traffic is mainly affected by traffic condition in the lane close to ramps, traffic flow data in the right lane (or left lane if left-side ramps) were extracted from the detector stations immediately upstream of off-ramps or downstream of on-ramps. This is because merging vehicles from on-ramps are more likely to be affected by traffic congestion downstream of the merging point whereas diverging vehicles to off-ramps are more likely to be affected by traffic conditions upstream of the diverging point. Using 30-second average speed and volume in the lanes close to ramps, 5-minute average speed and volume for 5~10 minutes prior to the time of crashes were calculated. Due to a malfunction of loop detectors at some stations during some time periods, loop detector data were available for only 260 ramp crashes among 522 crashes.

Development of models

This study develops log-linear models to identify the relationship between ramp crash frequency and the associated ramp geometric and traffic characteristics. The log-linear models describe log-transformed crash frequency in the function of a linear combination of these characteristics. In the models, crash frequency was adjusted by a average daily ramp traffic volume used as a covariate (continuous variable) to represent crash rate. All other explanatory factors were used as categorical variables. In case of 5-minute average speed and volume in the lanes close to ramps, they were categorized into low and high regimes. To objectively classify these two regimes, the distributions of average speed and volume during normal traffic condition should be observed. For this purpose, 1,888 five-minute average speed and volume in the right and left lanes were extracted from all loop detectors in randomly selected time periods when crashes did not occur. As a result of *t*-tests, the average values of these random non-crash samples were statistically different from the average values of crash cases at a 95% confidence level. Given that ramp traffic volumes were collected in the form of daily and hourly volume, the two models with daily and hourly ramp traffic volume were developed as follows:

Model with daily ramp traffic volume

The log-linear model with daily ramp traffic volume is defined as follows:

$$\ln(F) = \theta + \lambda_{Type} + \lambda_{Config} + \lambda_{AV} + \lambda_{AS} + \beta \cdot \ln(RTV) \qquad (1)$$

with F = the expected number of ramp crashes over 5 years; θ = constant; λ_{Type}= effect of ramp type (1 = off-ramp, 2 = on-ramp); λ_{Config} = effect of ramp configuration (1 = diamond, 2 = loop, 3 = outer, 4 = direct); λ_{AV}, λ_{AS} = effect of 5-minute average volume and speed in the lane close to ramps (1 = low, 2 = high), respectively; RTV = ramp traffic volume in million vehicles over 5 years; and β = parameter for ramp traffic volume.

Table 1. Estimated parameters of log-linear model (Daily ramp traffic volume)

Parameters	Estimate	Standard Error	Z-value
Constant (θ)	-4.28	2.17	-1.97
Off-ramp (λ_{Type}=1)	0.69	0.15	4.72
On-ramp (λ_{Type}=2)*	0	.	.
Diamond (λ_{Config}=1)	0.43	0.18	2.39
Loop (λ_{Config}=2)	0.22	0.64	0.35
Outer connection (λ_{Config}=3)	-0.27	0.66	-0.42
Direct/Semi-direct connection (λ_{Config}=4)*	0	.	.
Low volume in mainline (λ_{AV}=1)	-0.50	0.14	-3.55
High volume in mainline (λ_{AV}=2)*	0	.	.
Low speed in mainline (λ_{AS}=1)	0.54	0.14	3.84
High speed in mainline (λ_{AS}=2)*	0	.	.
Ramp traffic volume (β)	1.05	0.38	2.76

*The parameters of aliased cells are set to zero.

The parameters of the log-linear model (Equation 1) were estimated using iterative proportional fitting as shown in Table 1. As a result, the model yielded a good fit to the observed data as indicated by low log-likelihood ratio χ^2 (= 21.35) and high p-value (= 0.62). The parameters in the table represent the relative impact of each level of categorical variables on crash rates compared to the aliased cell (base case). Thus, positive parameters indicate an increase in ramp crash rates relative to the base case and negative parameters indicate a decrease in ramp crash rates relative to the base case.

It was found that off-ramps and diamond ramps contribute to higher ramp-related crash rate. Higher crash rates on off-ramps than on-ramps are mainly due to vehicle speed on ramps – diverging vehicles travel faster than merging vehicles at the beginning of ramps. Crash rates are higher on diamond than loop ramps because a substantial number of crashes occurred at the intersections of diamond ramps.

It was also found that higher volume and lower speed led to more ramp crashes. Higher volume represents more chances of vehicles exiting via off-ramps and more vehicles downstream of on-ramps. This implies high exposure to crashes on off-ramps and higher mainline volume downstream of the merge point leading to more chance of collision between the mainline and ramp traffic. Lower speed represents more chance of congested traffic condition upstream of off-ramps and downstream of on-ramps. Due to this traffic congestion, vehicles are required to suddenly reduce speed or stop at the end of ramps before merging (on-ramps), and vehicles are hindered from diverging at the beginning of ramps and may increase the speed on ramps (off-ramps).

However, it should be noted that some ramp configurations such as loop and outer ramps were not statistically different in terms of their effects on ramp crashes. Thus, it is unclear whether the model truly represents the effects of different ramp configurations and ramp traffic volume on ramp crashes. For this reason, the log-linear models were developed using hourly ramp volume rather than total daily volume.

Model with hourly ramp traffic volume

The log-linear model with hourly ramp traffic volume defined is as follows:

$$\ln(F) = \theta + \lambda_{Type} + \lambda_{Config} + \lambda_T + \lambda_{AV} + \beta \cdot \ln(p_T \times RTV) \quad (2)$$

with λ_T = effect of time of day, and p_T = proportion of ramp traffic volume during corresponding time period T of day. Considering the typical pattern of hourly variation in traffic volume, the time of day was divided into the following 5 periods: dawn (12 am – 6 am), morning peak (6 am – 10 am), mid-day (10 am – 3 pm), afternoon peak (3 pm – 7 pm) and night (7 pm – 12 am). Ramp traffic volume during each time period was estimated by applying the hourly proportion to the total daily volume.

Table 2. Estimated parameters of log-linear model (Hourly ramp traffic volume)

	Estimate	Standard Error	Z-value
Constant (θ)	-4.57	1.09	-4.20
Off-ramp (λ_{Type}=1)	0.44	0.15	2.90
On-ramp (λ_{Type}=2)*	0	.	.
Diamond (λ_{Config}=1)	0.41	0.17	2.50
Loop (λ_{Config}=2)	0.95	0.48	1.96
Outer connection (λ_{Config}=3)	1.17	0.59	1.97
Direct/Semi-direct connection (λ_{Config}=4)*	0	.	.
Dawn (λ_T=1)	2.10	0.41	5.06
Morning peak (λ_T=2)	-0.85	0.26	-3.25
Mid-day (λ_T=3)	-0.23	0.25	-0.90
Afternoon peak (λ_T=4)	-0.40	0.25	-1.58
Night (λ_T=5)*	0	.	.
Low volume in mainline (λ_{AV}=1)	-0.44	0.15	-2.88
High volume in mainline (λ_{AV}=2)*	0	.	.
Ramp traffic volume (β)	1.47	0.28	5.33

*The parameters of aliased cells are set to zero.

The parameters of the modified log-linear model (Equation 2) were estimated as shown in Table 2. The model fit was good (log-likelihood ratio χ^2 = 67.97, p-value = 0.51) and most variables were statistically significant. Average speed was excluded to avoid excessively large number of categories in a contingency table. It should be noted that the effects of ramp configurations changed in the modified model. More specifically, loop and outer ramps are more associated with higher ramp crash rates than diamond ramps and their effects relative to the base case (direct ramp) became marginally significant at a 95% confidence level unlike the previous model with daily ramp traffic volume. This indicates that the model with hourly ramp traffic volume better represents the effect of ramp configuration on ramp crash rates. Thus, in spite of high

number of crashes at the intersection of diamond ramps, curved ramps (loop and outer) contributes to higher crash rates than straight ramps (diamond).

Also, it can be seen that the time of day has significant effect on ramp crashes – crash rates are highest during dawn and relatively lower during peak periods. The reason for high crash rate during dawn can be partially explained by very high proportion of single-vehicle crashes. This indicates that drivers' careless driving and loss of control more frequently occur in very low traffic volume during dawn and this contributes to high crash risk on ramps. It was also confirmed that the proportions of single-vehicle crashes were higher on curved ramps than diamond ramps during dawn.

Conclusions and recommendations

This study develops log-linear models that identify the relationship between ramp crashes and ramp geometric/traffic flow characteristics. The findings of the analysis are summarized as follows: 1) Temporal variation in ramp-related crash risk can be predicted using 5-minute traffic data in the mainline close to ramps, typically obtained from loop detectors; and 2) Hourly ramp traffic volume is a better exposure that can reflect differential effects of ramp configurations on crash risk than total daily ramp traffic volume by accounting for high number of single-vehicle crashes during dawn.

Based on the findings, we recommend the following countermeasures. First, more strict regulation should be imposed on drivers' obedience to traffic signals particularly at the intersections of diamond and outer ramps. In case of off-ramps, warning signs that advise drivers of possible stops at the end of ramps will help drivers avoid collisions with the stopped vehicles. Second, traffic conditions in the lanes close to ramps need to be monitored to detect high risk of ramp crashes. Third, drivers' careless driving during dawn should be more regulated on loop and outer ramps.

References

Bauer, K. M. and Harwood, D. W. (1997). "Statistical models of accidents on interchange ramps and speed-change lanes." Report No. FHWA-RD-97-106, Federal Highway Administration.

Khorashadi, A. (1998). "Effects of ramp type and geometry on accidents." Report No. FHWA/CA/TE-98/13. California Department of Transportation.

Lundy, R. (1967). "The effects of ramp type and geometry on accidents." *Highway Res. Rec. 163*, 80-117.

McCartt, A. T., Northrup, V. S., and Retting, R. A. (2004). "Types and characteristics of ramp-related motor vehicle crashes on urban interstate roadways in Northern Virginia." *J. Saf. Res. 35*, 107-114.

Twomey, J. M., Heckman, M. L., Hayward, J. C., and Zuk, R. J. (1993). "Accidents and safety associated with interchanges." *Transport. Res. Rec. 1385*, TRB, National Research Council, Washington, D.C., 100-105.

Application of data mining techniques for real-time crash risk assessment on freeways

A. Pande[1] and M. A. Abdel-Aty[2]

[1]Department of Civil and Environmental Engineering, University of Central Florida, 4000, Central Florida Blvd., Orlando, Fl 32816-2450; PH (407) 823-4902; FAX (407) 823-3315; email: anurag@mail.ucf.edu.

[2]Department of Civil and Environmental Engineering, University of Central Florida, 4000, Central Florida Blvd., Orlando, Fl 32816-2450; PH (407) 823-5657; FAX (407) 823-3315; email: mabdel@mail.ucf.edu.

Abstract

Data mining is the analysis of large "observational" datasets to find unsuspected relationships that might be useful to the data owner. It typically involves analysis where objectives of the mining exercise have no bearing on the data collection strategy. Freeway traffic surveillance data collected through underground loop detectors is one such "observational" database maintained for various ITS (Intelligent Transportation Systems) applications such as travel time prediction etc. In this research data mining process is used to relate this surrogate measure of traffic conditions with rear-end crash occurrence on freeways. Crash and dual loop detector data from 36.25-mile instrumented Interstate-4 corridor in Orlando (FL) are used in this study. The research problem is set up as a classification problem and separate data mining based classifiers are developed to discriminate crashes belonging to different categories from normal conditions on the freeway. Based on the models developed in this study one can identify the traffic conditions prone to rear-end crashes 5-10 minutes prior to the crash. The findings of this research are proposed to be used as a proactive traffic management system which could warn the drivers about potential rear-end crashes.

Introduction

The objective of this research is development of a framework to detect crash prone conditions in real-time. To achieve these objectives loop data collected from randomly selected non-crash locations have been used in this study along with the crash data. These data most commonly include speed, vehicle counts, and lane occupancy provided every 30 seconds by loop detectors installed beneath the freeway pavement. To establish relationships between real-time traffic data, geometric parameters, and rear-end crashes a data mining approach is adopted. It essentially means that tools from a range of fields such as machine learning (e.g., clustering algorithms), statistics (e.g., classification tree), and/or artificial intelligence are used in a step by step manner to analyze the data.

This research is part of a new trend in freeway traffic management which until recently was focused on timely detection of incidents. With the enormous increase in mobile phone usage in the recent past relevance of incident detection is diminishing and traffic management authorities are looking for proactive strategies. The basic element of a proactive traffic management system would be reliable models separating crash prone conditions from 'normal' traffic conditions in real-time. Most of the existing real-time crash 'prediction' models available in the literature are generic in nature, i.e., single generic model has been used to identify all crashes (such as rear-end, sideswipe, or angle). Conditions preceding crashes are likely to differ by type of crash and therefore the approach towards proactive traffic management should be type (of crash) specific in nature. The disaggregate models would also be useful in devising specific countermeasures for crashes. In this research the focus is on the most frequent group of crashes on the freeways, i.e., the rear-end crashes. The rear-end crash data for this study are collected over a five year period (1999 through 2003) from 36.25-mile corridor of Interstate-4 in Orlando metropolitan area along with information about geometric design features, such as ramp locations, curvature, etc. The corridor has a total of 69 loop detector stations in each direction,

spaced out at nearly half a mile. Each of these stations consists of dual loops and measures average speed, occupancy, and volume over 30-second period on each of the three through travel lanes in both directions.

Background

Lee *et al.* (2002, 2003) developed and refined log-linear models to predict crashes using crash precursors estimated from loop detector data. It was found that the coefficient of temporal variation in speed has a relatively longer-term effect on crash potential than density while the effect of average variation of speed across adjacent lanes was found to be insignificant. A study by Oh *et al.* (2001) also showed the 5-minute standard deviation of speed value to be the best indicator of 'disruptive' traffic flow leading to a crash as opposed to 'normal' traffic flow. Garber and Subramanyan (2002) demonstrated the feasibility of developing a methodology in which real-time data can be used to formulate traffic management strategies also incorporating crash risk. In our previous work (Abdel-Aty et al., 2004, 2005) case-control logistic regression models were developed with matched sampling of traffic flow characteristics for crash and non-crash cases while controlling for other external factors such as the roadway geometry, time of the day, etc. These generic logistic regression models achieved satisfactory classification accuracy.

The major shortcoming of these studies was that the inferences were made based on a 'one-size-fits-all' approach. It means all types of crashes were sought to be identified using a single generic model. In this study we try to overcome these deficiencies by examining traffic data from a series of loop detectors in order to explore their relationship with rear-end crashes. The choice of rear-end crashes was obvious due to their high frequency and significant impact on freeway operation. Loop data belonging to rear-end crashes have been used with non-crash data that are collected from randomly chosen corridor locations over the 5-year period (1999 through 2003). The random sampling of non-crash locations enables us to explore the impact of 'off-line' factors (e. g, presence of ramps, time of day, horizontal curvature), along with real-time traffic parameters, on occurrence of a rear-end crash.

Data Collection

There were *2179* rear-end crashes reported in the study area during the five year period (from 1999 through 2003). From the FDOT (Florida Department of Transportation) crash database we extracted information such as the date and mile-post location for each crash. Scanned copies of individual crash reports were then used to extract the reported time of the crashes. Based on a shock-wave progression based methodology and the precise location of the crash (known from the FDOT crash database) we ascertained that the reported time was in fact very close to the time of occurrence (Pande, 2005). Loop data corresponding to crashes would be used to train the models about the crash prone conditions while a sample of randomly selected non-crash cases would be used to 'teach' the model about what constitutes 'normal' freeway traffic.

Loop data collection for crash and non-crash cases

Loop data were extracted for every crash in a specific format, for example, if a crash occurred on April 12, 1999 6:00 PM, I-4 Eastbound and the nearest loop detector was at station 30, data were extracted from station 30, two loops upstream and two loops downstream of Station 30 for 20-minute period prior to the reported time of the crash. Hence, this crash case will have loop data table consisting of the 30-seconds averages of speed, volume, and occupancy for all three lanes at stations 28 through 32 (on eastbound direction) from 5:40 PM to 6:00 PM on April 12, 1999. The choice of four stations and 20 minutes was based on results from our previous studies (Abdel-Aty et al., 2004, 2005).

The raw 30-second data have random noise and are difficult to work with in a modeling framework. Therefore, the 30-second raw data were combined into 5-minute level in order to obtain averages and standard deviations. For 5-minute aggregation 20-minute period was divided into four time slices. The stations were named as "*D*" to "*H*", with "*D*" being farthest station upstream and so on. It should be noted that "*F*" is the station closest to the location of the crash (Station of the crash) with "*G*" and "*H*" being the stations downstream of the crash location. Similarly the 5-minute intervals were also given "IDs" from 1 to 4. The interval between time of the crash and 5 minutes prior to the crash was named as time-slice 1, interval between 5 to 10 minutes prior to the crash as time-slice 2, and so on. The parameters were also aggregated across the three lanes and the averages (and standard deviations) for speed, volume, and lane-occupancy at 5-minute level were calculated based on 30 (10*3 lanes) observations. It is worth mentioning that if at a freeway location the detector from a certain lane reported missing/invalid data, the observations from that lane were not used for calculating averages and standard deviations. In such scenario (with missing/invalid data from one or two out of the three lanes) there would be less observations (either 10 from one lane or 20 from two lanes) available to get a measure of traffic flow at that location.

The format of the traffic data collected with respect to time and location of crashes and the nomenclature for independent variables is shown in Figure 1. The variable *"SSD2"*, for example, represents the standard deviation of 30 speed observations during the 5-minute period of 5-10 minutes prior to a crash at station "D", which is the farthest upstream station. Note that, due to random intermittent failure of certain detectors, traffic data were not available for all *2179* rear-end crashes. Hence, the analysis presented in this paper is based on 1620 crashes which had the corresponding loop data available after disregarding the cases with some missing loop data.

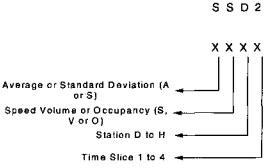


Figure 1. Nomenclature of independent variables with respect to time and location of the crash

As mentioned earlier, the sample of non-crash cases was generated randomly from the loop database consisting of traffic data from the year 1999 through 2003. The sample was expectedly almost uniform over the times of day, day of week, and freeway location. The details of the process of generating the sample may be found in Pande (2005). These random combinations were extracted in the same format as the crash data, i.e., sets of 20-minute loop data prior to the assigned time of the non-crash from 2 stations upstream and 2 stations downstream of a station assigned as station of non-crash. The random non-crash data were aggregated to 5-minute level and traffic parameters similar to crash cases (refer to Figure 1) were generated for this sample as well. The variable "y" was given the value 0 for these cases. After the assembly of traffic parameters, geometric features of the freeway at the locations of aforementioned crash and non-crash cases were collected. This information includes the distances of nearest on and off-ramp from crash (and non-crash) locations, in both upstream and downstream direction and horizontal curvature of the freeway. The database assembled in this study is by far the most comprehensive database for development of a proactive traffic management strategy.

Data Analysis

As part of preliminary analysis, distributions of average speeds just before the crash were examined at loop detector locations surrounding the crash location. The histogram distributions for variables *ASD1, ASF1*, and *ASH1* (5-minute average speeds at Station D, F, and H, respectively) over all rear-end crashes appear to have the shape of two adjacent approximately mound-shaped distributions. These distributions suggest that the crashes belonging to each peak need to be analyzed separately. In one of our previous studies (Abdel-Aty et al., 2005), crashes were separated by simply splitting the crash data based on the average speeds at station F just before the crash (time slice 1, 0-5 minutes before the crash). In this analysis, the idea of separating crashes by prevailing conditions only at station of the crash (station F) is refined. It is imperative because rear-end crashes at freeway locations are expected to be affected not only by the prevailing speeds at that location but also by the interplay between traffic speeds at the locations upstream and/or downstream of it. To reflect this fact, it was decided to cluster the rear-end crashes into two segments/clusters/groups, based not only on *ASF1* but also on traffic speeds measured at the extremities of the 2-mile stretch around crash location (i.e., *ASD1, ASF1* and *ASH1*).

Kohonen vector quantization (*KVQ*) technique (Kohonen, 1988) was used to cluster the crash data into two groups with three average speed parameters *ASD1, ASF1,* and *ASH1* as inputs. It is intended that separate models will be applied to predict the two groups (segments/clusters) of rear-end crashes. From an application perspective, one must be able to identify the cluster to which the real-time data under consideration belong, so that appropriate model(s) may be applied to assess whether or not it is a crash prone pattern. It can not be achieved through an unsupervised learning algorithm such as the *KVQ* method. Therefore, a set of classification rules were needed that may be used to assign real-time traffic speed patterns into one of the two clusters. Classification tree was selected as the tool to formulate these rules. The rules formulated by the classification tree model to separate rear-end crashes belonging to one cluster from the other are summarized in Table 1. Note that although the clusters in rear-end crashes were obtained based on traffic speeds prevailing right before the crash (0-5 minutes; time-slice 1) the rules in Table 1 use average traffic speeds from time-slice 2 (5-10 minutes before the crash). In a real-time application, it would allow more leverage in terms of time available to analyze the data before the crash actually occurs.

TABLE 1: The series of rules to identify clusters in rear-end crash data

Leaf	Conditions (Series of Rules)	Cluster Assigned
1	ASF2<44.146 and ASD2<51.26	cluster 1
2	ASF2<44.146 and ASD2>=51.26 and ASH2<46.8	cluster 1
4	ASF2>=44.146 and ASH2<32.941 and ASD2<53.165	cluster 1
6	ASF2>44.146 and ASH2>=32.941 and ASD2<27.30	cluster 1
3	ASF2<44.146 and ASD2>=51.26 and ASH2>=46.8	cluster 2
5	ASF2>44.146 and ASH2<32.941 and ASD2>=53.165	cluster 2
7	ASF2>44.146 and ASH2>=32.941 and ASD2>=27.30	cluster 2

From these rules it may be inferred that the cluster 1 rear-end crashes generally belong to low speed traffic regime, while those in cluster 2 belong to medium to high speed traffic regime. Hence, most of cluster 2 crashes occur under relatively free flow conditions that commonly prevail on freeways. Based on these observations one could infer that cluster 1 rear-end crashes occur during congested conditions that prevail on the freeway for small part of the day and have very low exposure. Based on these classification rules about 45.8% crashes were identified as cluster 1 while the remaining 53.8% were identified as cluster 2. If we apply classification rules from Table 1 to a dataset with random non-crash cases then only 6.27% were classified as cluster 1. It means that although cluster 1 makeup 45.8% of the crash dataset, it only makes 6.27% of the random non-crash sample. It indicates that the crashes belonging to cluster 1 may be 'predicted' (or anticipated) using the classification tree rules shown in Table 1. If we assign all traffic patterns belonging to cluster 1 as rear-end crashes, we would be able to identify about 46% of rear-end crashes by issuing warnings just over 6% of the times. Same procedure, however, would not work for cluster 2 rear-end crashes since cluster 2 traffic conditions are way more frequent (94% in the randomly selected loop data patterns) on the freeway. Hence, further classification models are needed to separate crashes from the non-crash cases within the traffic data belonging to cluster 2.

Classification models for cluster 2 rear-end crashes

In this section the data mining process is extended to develop classification model(s) for cluster 2 rear-end crashes. Classification models developed for cluster 2 rear-end crashes belong to multi-layer perceptron (MLP) and normalized radial basis function (NRBF) neural network architecture. Theoretical details of the two architectures and training procedures may be found in any standard neural network text, e.g., Christodoulou and Georgiopoulos (2001). Inputs to the classification models are decided based on a classification tree based variable selection procedure developed by Brieman et al. (1984). Details of the procedure and the variables included may be found in Pande (2005).

The neural network modeling was repeated in three steps. In first step, the independent variables included were the off-line factors and the traffic parameters measured only at station nearest to the crash location (i. e., Station F). In the next step, traffic parameters were included from three stations, station of crash and one station each in the upstream and downstream direction. In the third step traffic parameters were included from five stations, i.e., Station D through H. The input dataset has 878 cluster 2 rear-end crashes along with the non-crash cases which made 85% of the sample. The most critical parameter affecting the performance of neural networks is the number of nodes in the hidden layer (Cybenko, 1986). To select appropriate number of nodes in the hidden layer, the performance of ten different networks with hidden nodes varying from 1 through 10 were examined for *MLP* as well as *NRBF* architecture using Enterprise Miner from SAS Institute (SAS Institute, 2001).

To evaluate the performance of neural networks these models were applied to the validation dataset. The output of these models (for any observation) is the posterior probability (0<posterior probability<1) of the event of interest (i.e., a rear-end crash). The closer it is to unity the more likely, according to the model, it is for that observation to be a rear-end crash. 30% observations with maximum posterior probability are classified as crash and various models were examined based on the proportion of the validation dataset crashes captured within those observations.

NRBF with four hidden neurons were found to be best models when traffic parameters from one and three stations were included inputs. MLP with 8 hidden neurons were the best model when traffic parameters from 5 stations were used as inputs. The performance of the models was improved by combining the best models in each category. To combine two or all three of the best models in each category the output posterior probability was averaged for the individual models. It was found that

hybrid model created by combination of the three models performed better than the individual model and identified 55.4% of crashes in the validation dataset. The combination of best 1-station and 3-station model identified 53.05% crashes and the best 1-station models was the NRBF model with 4 hidden neurons, which identified 50.06% crashes from the validation dataset within the 30% observations with maximum posterior probability.

It is worth mentioning at this point that 55.4% identification of cluster 2 crashes was achieved through the model that uses traffic data from five stations (combination of best 1-station, 3-station, and 5-station models). Since data from five stations may not be simultaneously available due to intermittent failure of loops; performance of the models must be seen in terms of their data requirements as well. Sometimes it would be more practical just to use data from one station to identify these crashes. Therefore, even though the model provides better identification of cluster 2 crashes it would not make it an automatic choice for field implementation.

Real-time identification of rear-end crashes

Based on the data analysis presented here a real-time application strategy to identify conditions prone to rear-end crashes may be formulated. The strategy may be used to flag locations which are experiencing high risk of rear-end crashes. The application first starts by applying classification tree model based rules shown in Table 1. Those rules may be used to identify whether traffic data belong to cluster 1 or cluster 2. If the patterns belong to cluster 1 a rear-end crash warning is issued for the location without any further application. If the patterns are identified to be cluster 2 then we need to apply the neural network based hybrid models. As mentioned earlier, the hybrid models that combines best 1-station, 3-staton, and 5-station MLP/NRBF models provided optimal crash identification over the validation dataset and hence is preferred over other models. This model, of course, would need data from five stations around the section where we are trying to assess the crash risk. Therefore, in the next step check for data availability over five stations is applied. If data from five stations are available then the data are subjected to the hybrid model. If the requisite data are not available then the patterns may be subjected to the models with less accuracy but more tolerant data requirements (requiring inputs from 1 or 3 loop detector stations).

With this strategy one can identify *46%* of rear-end crashes (percentage of cluster 1 crashes among all rear-end crashes) by issuing warnings for about 7% cases. 55.4% of cluster 2 crashes, which make 54% of the rear-end crash data, may be identified by issuing warnings 30% of the times among the remaining 93% cases. It essentially means that about ¾ (46 + (54*55)/100 ≈ 75%) of the crashes could be identified by issuing warnings for about one third of cases (7 + (93*30)/100 ≈ 34%). It roughly translates into 66% accuracy on non-crash data for identification of 75% crashes. Further research is recommended to estimate the frequency and the nature of these warnings.

Conclusions

The paper presents a step by step approach of data analysis to develop a strategy to identify real-time traffic conditions prone to rear-end crashes using freeway loop detector data. It was concluded that the rear-end crashes on the freeway may be grouped into two distinct clusters based on the average speeds prevailing in approximately 2-mile section around the crash location 5-10 minutes before a crash. One cluster (group) of crashes occurs under extended congestion on the freeway while the average speeds are relatively higher during the 5-10 minute period before a cluster 2 crash (refer Table 1 for specific traffic speed conditions for each group of rear-end crashes). It was noticed that conditions belonging to cluster 1 occur very rarely and hence whenever such conditions are encountered in real-time then a crash warning may be issued. For cluster 2 rear-end crashes further neural network based classification models were developed. Based on the performance of the

classification models and the proposed real-time application strategy, 75% of the rear-end crashes may be identified 5-10 minutes before their occurrence with just 34% positive decisions (i.e., crash warnings). Since crashes (however frequent on the I-4 corridor under consideration) are rare events; these positive decisions would result in a significant number of 'false alarms'. However, it should be noted that 'false alarms' are not as detrimental in the present application as they are for incident detection algorithms. Crash prone traffic conditions, which have been identified in this paper, would not always result in a rear-end crash occurrence. The conditions, however, are worth warning the drivers and drivers need to be more attentive under such traffic conditions even if they may not always culminate in a rear-end crash. A reasonable number of warnings, which the drivers do not consider excessive, based on the models developed can potentially play a critical role in proactive traffic management. These warnings may be issued to the motorists driving on the freeway locations through VMS (variable message signs). However, the frequency and impacts of such warnings on driver behavior would need to be carefully estimated before implementing such measure. Another application for the findings of this research could be the formulations of VSL (variable speed limit) implementation strategies that can reduce the probability of rear-end crashes.

References

Abdel-Aty, M., Uddin, N., Abdalla, F., Pande, A., and Hsia, L. (2004). Predicting freeway crashes based on loop detector data using matched case-control logistic regression. *Transportation Research Record* 1897, pp. 88-95.

Abdel-Aty, M., Uddin, N., and Pande, A. (2005). Split models for predicting multi-vehicle crashes under high speed and low speed operation conditions on freeways. *Transportation Research Record* 1908, pp. 51-58.

Breiman, L., Friedman, J., H., Olshen, R., A., and Stone, C., J. (1984). Classification and regression trees. *Chapman and Hall,* New York.

Christodoulou, C., and Georgiopoulos, M. (2001). Applications of Neural Networks in Electromagnetics, *Artech House*, Boston.

Cybenko, C. (1989). Approximations by superposition of sigmoid functions. *Mathematics of Control Signals and Systems, Vol. 2*, pp. 303-314.

Garber, N.J., Subramanyan, S. (2002). Feasibility of incorporating crash risk in developing congestion mitigation measures for interstate highways: a case study of the Hampton roads area. Report No. FHWA/VTRC 02-R17. Virginia Transportation Research Council, Charlottesville, VA.

Hand, D., Mannila, H., and Smyth, P. (2001). Principles of data mining. *The MIT Press,* Cambridge, Massachusetts.

SAS Institute, (2001). Getting Started with Enterprise Miner Software, *Release 4.1, SAS Institute*, Cary, NC.

Kohonen, T. (1988). Learning vector quantization, *Neural Networks, 1 (suppl 1),* 303.

Lee, C., Saccomanno, F., and Hellinga, B. (2002). Analysis of crash precursors on instrumented freeways. *Transportation Research Record 1784*, pp. 1-8.

Lee, C., Hellinga, B., and Saccomanno, F. (2003). Real-time crash prediction model for the application to crash prevention in freeway traffic. *Transportation Research Record 1840*, pp. 67-78.

Oh, C., Oh, J., Ritchie, S., and Chang, M. (2001). Real-time estimation of freeway accident likelihood. Presented at the *80th annual meeting of Transportation Research Board,* Washington, D.C.

Pande, A., Applying hybrid models for real-time crash risk assessment on freeways. Ph. D. Dissertation, *University of Central Florida*, 2005.

Estimating Pedestrian Counts in Urban Areas for Transportation Planning and Safety Analyses

Srinivas S. Pulugurtha[1], Shashi S. Nambisan[2], and Pankaj Maheshwari[3]

[1]Department of Civil Engineering, University of North Carolina, Charlotte, 9201 University City Boulevard, Charlotte, NC 28223-0001; Tel. (704) 687-6660; Fax (704) 687-6953; Email: sspulugu@uncc.edu

[2] Transportation Research Center, University of Nevada, Las Vegas, 4505 Maryland Parkway, Las Vegas, NV 89154-4007; Tel. (702) 895-1325; Fax (702) 895-4401; Email: shashi@ce.unlv.edu

[3] Parsons Transportation Group, 840 Grier Drive, Suite 140, Las Vegas, NV 89119; Tel. (702) 435-2116; Email: pankaj47@gmail.com

Abstract

This paper investigates factors that can be used to quantify pedestrian counts in urban areas. The best subset regression is used to develop models to estimate pedestrian counts considering variables identified using general linear regression. The F-test is used to support the analysis. The models are developed using data collected at 15 selected locations with high pedestrian activity in the Las Vegas metropolitan area. The findings show that the pedestrian counts are a function of number of lanes, average annual household income and residential area proximate to the study location. Results show that the pedestrian counts are independent of the commercial area and the number of bus stops in the vicinity of the location. The developed models can be used to estimate pedestrian counts at any high pedestrian activity location provided the socioeconomic and demographic characteristics are known. The methodology is also applicable to other urban settings.

Introduction

The focus of this paper is to develop statistical models to estimate pedestrian counts during peak periods. These counts are generated due to pedestrian related activities in urban areas. The number of pedestrian counts provides valuable information for the planning, design, operations and management of transportation networks for pedestrians. The number of pedestrian counts generated depends on key factors such as low income group population, number of housing units, square foot of development, number of bus stops and bus routes within the walking distance, and school zones near the vicinity of the proposed study area. Roadway characteristics such as signalization, presence of side walks, and numbers of lanes are also determinants in affecting pedestrian counts.

Various candidate models for morning and evening peak hours are evaluated depending on the type of the facilities near the locations. General linear regression modeling is used to identify the significant factors. Pedestrian volume data are collected at 15 selected locations with high pedestrian activities and then depending on the surrounding socio-economic, demographic, land use and traffic characteristics, regression models are developed to estimate pedestrian counts. The models are developed using data from the Las Vegas metropolitan area. The outcomes of the study are expected to help better estimate pedestrian counts. In turn, this provides valuable information for the planning, design, operations and management of transportation networks for pedestrians.

Literature Review

Pedestrians walk more in urban areas than in suburban areas because of the route directness and convenience of pedestrian facilities (Mouden et al., 1997). On the other hand, urban land use patterns play an important role in influencing walking trips (Handy, 1996; Loutzenheiser, 1997). People who live in spread out development patterns (sprawl) areas spend less time walking than people who live in mixed-land use, neo traditional neighborhood, or well established neighborhoods, because sprawl requires more frequent and longer trips (Berkovitz, 2001). It should be noted that walking is particularly risky in metropolitan areas with a large degree of sprawl. In such areas, the development tends to increase the number of automobile owners, hence high speed and high volume roadways, which create unfriendly pedestrian environments. Several authors have worked on topics related to modeling pedestrian trips in the past. Examples of such studies on modeling pedestrian trips include examining pedestrian crash data using "time spent walking" and "number of roads crossed" as exposure measures (Keall, 1995), studying the relationship between site design and pedestrian travel in a mixed-use, medium density environment (Hess et al., 1999), and, forecasting pedestrian volumes in high-density urban areas based on existing land use characteristics and pedestrian volumes for specific locations (Pushkarev et al., 1971).

Data

The objective of this research is to develop models to help estimate pedestrian counts in urban areas during key periods of the day such as the morning peak and evening peak hours. Data required for modeling include pedestrian volume data, on-network characteristics (number of lanes and number of bus stops), and off-network characteristics (population, average household income, and square foot of commercial and residential development).

Pedestrian Volume Data

Pedestrian volumes were collected at high pedestrian activity locations during the morning peak, and evening peak hours. Fifteen high crash locations were considered for the analysis out of which 10 are intersections and 5 are point locations along corridors. Data collected along one leg of an intersection were considered in this study. This was done to keep it consistent with data collected at point locations along corridors. Counts include pedestrians crossing in both the directions at each location. The morning peak hour is considered to be 7:00 AM to 9:00 AM while the evening peak is defined as 4:00 PM to 6:00 PM. Pedestrian volume data were collected in an urban setting of the Las Vegas metropolitan area located in Clark County, Nevada. Since Las Vegas is one of the top tourist destinations in the United States, sites are chosen so as to minimize the possibility of erroneous data collection. Typically, the visitor population peaks over the weekends. Thus, data were collected on a typical weekday, generally on Tuesdays and Thursdays to make sure that they do not account for the weekly peaking of tourist data.

On-network and Off-network Characteristics

On-network characteristics considered for modeling pedestrian counts include number of lanes and number of bus stops. The number of lanes was obtained from the field. The bus stops data were obtained in a GIS format from the Regional Transportation Commission of Southern Nevada. The bus stops within a quarter mile of the specified location were identified by creating a buffer. The number of bus stops is an important indicator in estimating the pedestrian counts. To consider a more general aspect of this variable, the number of bus stops within a quarter mile buffer zone is considered for the analysis.

Off-network characteristics considered for modeling pedestrian counts include population, average household income, and square foot of commercial and residential development. Data was extracted using ArcMap, a Geographic Information Systems (GIS)

software. The street centerline network was combined with the census block group polygon layer of the Las Vegas metropolitan area. Then point shape files were created for the selected sites and a buffer of a quarter mile was created around the sites. The quarter mile was selected based on the anecdotal evidence that people would at most walk for 5 minutes in the study area. The next step was to compute the population density, average household income, square foot of commercial and residential development within the buffer zones. Within the buffer zones, there are several parcel layers with certain characteristics such as parcel area, total number of households, income groups ranging from less than $10,000 to greater than $150,000, and population. The block group polygon layer is clipped with the buffer layer to compute the shape area of the parcel within the buffer zone.

Estimation of Population within the buffer layer

The population is assumed to be distributed uniformly throughout the census block. As a result, the population is assumed to be distributed in proportion of the census block and is calculated using by the following equation.

$$P_i = \frac{A_i}{A_j} \times P_j$$

where, P_i = Population of census block inside buffer zone, P_j = Population of the census block 'j', A_i = Area of the census block inside buffer zone, and A_j = Area of the census block 'j'.

If the buffer comprises multiple census blocks, the population of each census block inside buffer zone is summed to estimate the population within the buffer layer.

Estimation of Average Household Income

The average income for each parcel layer inside the buffer zone is computed using the following relationship.

$$I = \frac{\sum (N_j \times I_j)}{\sum N_j}$$

where, I_j = Average income of block group 'j' $= \frac{\sum MeanIncomeGroups \times N_i}{\sum N_i}$ and

N_i = number of households in each income category.

N_j is the number of households within a parcel layer inside the buffer zone. It is assumed to be uniformly distributed in proportion to the parcel area and is represented as $N_j = N \times \frac{A_i}{A_j}$, where, N is total number of households within a parcel layer.

Estimation of Square Foot of Development

Land Use data were obtained from the Clark County GIS Manager's Office (GISMO) website where it is classified into various categories including residential, commercial and industrial. Separate shape files were made for all the land use categories and then each layer was clipped with the buffer layer to compute the development area within the buffer zone. The development area is the total square footage, and it represents the area which is developed and not the area allotted for development.

Models to Estimate Pedestrian Counts

Data for each selected location are shown in Table 1. Linear regression was used to build a model considering pedestrian volume as a dependent variable and variables such as average number of lanes (NLANES), population density in a buffer zone (POPDEN), average annual income (AVEINC), commercial development (ln(COMMAREA)), residential development (ln(RESAREA)), and number of bus stops (NSTOPS) as independent variables. Linear regression was done on the variables using commercial software (Minitab). An initial examination of the correlation matrix showed a strong correlation between variables POPDEN and ln(RESAREA). Thus, variable POPDEN was eliminated from further consideration as it was felt that residential development is better explanatory variable when compared to population density in a buffer zone (which is estimated based on the assumption that population is distributed uniformly throughout the census block). The Best Subset Regression was used to model different peak hour pedestrian trips and an attempt was made to select the best model using F-Test.

Table 1. Pedestrian Volume and Site Characteristics.

Site #	Location (Street 1 / Street 2)	Pop in Buffer Zone	Pop Density in Buffer Zone (1,000)	Avg Annual Income (1,000's of $)	Morning Peak Hour Ped Vol	Evening Peak Hour Ped Vol	Avg # Lanes per Leg	Commercial Sq. ft. (10,000)	Residential Sq. ft. (10,000)	# Stops
1	Maryland Pkwy / Sierra Vista Dr	2,469	43.991	40.449	90	65	6	78.589	56.125	10
2	Maryland Pkwy / Dumont St	2,571	49.551	40.787	55	59	5	119.326	51.886	8
3	Maryland Pkwy / Twain Ave	2,033	90.806	41.256	95	97	7	106.223	22.389	13
4	Harmon Ave / ParadiseRd	1,577	18.354	38.546	88	157	6	170.345	85.922	6
5	Flamingo Rd / Koval Ln	1,091	11.803	45.085	132	184	8	410.933	92.432	8
6	Flamingo Rd / Paradise Rd	2,318	25.241	41.247	125	172	8	183.548	91.834	6
7	Bonanza Rd / F Street	291	221.478	30.587	38	37	4	10.079	1.314	5
8	Bonanza Rd / D Street	317	338.097	27.619	39	78	5	33.688	0.938	8
9	Lake Mead Blvd / Las Vegas Blvd	1,653	447.410	41.654	74	93	8	22.694	3.695	7
10	Lake Mead Blvd / Pecos Rd	2,163	99.064	42.752	68	140	7	5.310	21.834	8
11	Lake Mead Blvd: Belmont St to McCarran St	2,653	1323.588	47.658	13	13	7	11.359	2.004	5
12	Fremont St: 11th St to 8th St	1,338	23.476	27.841	180	118	4	111.561	56.994	13
13	Fremont St: 8th St to 6th St	1,670	42.887	27.448	129	103	4	251.016	38.939	9
14	Twain Ave: Cambridge St to Swenson St	3,637	17.751	38.222	70	109	5	18.730	204.886	7
15	Twain Ave: Swenson St to Palos Verde St	4,469	23.260	38.213	92	76	5	13.005	192.133	8

Models for Morning Peak Hour Pedestrian Counts (MPHPC)

The regression analysis of variable MPHPC using different combinations of independent variables was done and the summary of those models is described in Table 2. Only those models which can clearly describe the effects of adding or deleting any other variable in the model are selected. The F-Test uses the concepts of reduced and full model to select the best regression model. Variance Inflation Factor (VIF) was used to estimate the amount of correlation among the covariates. Mean square error of prediction (R^2_{Pre}) and predicted sum of squares ($PRESS$) statistics are used as measures to indicate the predictive power of a model. Generally, a model with a small value of $PRESS$ is preferable to one with a large value. For more details on tests and statistics used in this paper, interested readers can refer to "*Applied Linear Regression* Models (Kutner et al., 2004)".

The main objective was to select a model that has best explanatory and predictive power. As seen from Table 2, the t-statistics are not significant at the 90 percent confidence level for model 1. Hence, this model was rejected without any further analysis. Model 2 has variable NSTOPS significant at 90 percent confidence level, and it has good R^2_{Adj} and R^2_{Pre} values as compared to model 3. Even the F-test shows that model 2 is better than model 3. In addition model 3 has a large intercept which might create some doubt about the validity of the model. Hence, model 2 is considered as good for explanatory and predictive purposes.

The final regression equation from model 2 is
MPHPC = 87.6+(21.4 NLANES)–(5.92 AVEINC)+(38.4 ln(RESAREA))+(5.32 NSTOPS)

Table 2. Models for Morning Peak Hour Pedestrian Counts.

Variable	Model 1			Model 2			Model 3		
	Estimate	t-stat	VIF	Estimate	t-stat	VIF	Estimate	t-stat	VIF
Intercept	68.20	1.31		87.6	1.86		149.12	3.74	
NLANES	18.32	2.37	3.8	21.35	3.09	3.1	23.47	3.06	3.0
AVEINC	-5.206	-2.79	4.2	-5.923	-3.55	3.5	-6.969	-3.91	3.1
ln(COMMAREA)	11.66	0.90	1.7	--	--	--	--	--	--
ln(RESAREA)	33.81	3.42	1.7	38.36	4.56	1.3	19.31	5.04	1.1
NSTOPS	4.90	1.79	1.3	5.31	1.99	1.2	--	--	--
R^2	0.831			0.816			0.744		
$R^2_{(Adj)}$	0.738			0.743			0.674		
F Stat	8.87			11.10			10.64		
S	22.065			21.856			24.606		
C_p	6.0			4.8			6.7		
PRESS	14705.8			14642.5			14881.7		
$R^2_{(Pre)}$	0.434			0.437			0.427		

Note: --indicates the variables not included in the model
$t_{critical}$ 90/95 = 1.77/2.16
The numbers in shaded boxes are not significant at 95% confidence level

Models for Evening Peak Hour Pedestrian Counts (EPHPC)
Regression analyses of the variable EPHPC using other independent variables was done and a summary of the resulting models is presented in Table 3. As shown in Table 3, the t-statistics are not significant at 95 percent confidence level for variables ln (COMMAREA) and NSTOPS in model 1. Hence, model 1 was rejected without any further analysis. For model 2, the reduction in RMSE (root mean square error, S) from model 1 is not sufficient enough to make the variable NSTOPS significant at 95 percent confidence level. So, model 2 was also rejected without giving any further consideration. Finally, model 3 seems to be the best as all the variables are significant at 95 percent confidence interval. This is also supported by the F-test, which shows that model 3 is better than model 2.

The final regression equation from model 3 is
EPHPC = 86.6+(39.3 NLANES)–(7.47 AVEINC)+(19.8 ln (RESAREA))

Summary and Conclusions
In this paper, the use of off-network and on-network characteristics to estimate pedestrian counts is evaluated. Multiple Linear Regression Techniques were used for this purpose. Models were developed for estimating morning and evening peak hour pedestrian counts. The results indicate that commercial area is insignificant and is not important in estimating pedestrian counts. It was seen that the number of bus stops was a significant factor in only the morning peak hour model. This could be due to the high percent of captive riders during the morning peak hours when compared to the evening peak hour. The average annual household income was significant in both the models, which generally was not seen in earlier research. As expected, the peak hour pedestrian counts were a function of number of lanes and residential area. The validity of the model was checked by estimating counts and comparing them with pedestrian trips at test locations other than those considered for modeling. The

mean squared prediction error is fairly close (2 percent error) to the error mean square based on the regression fit to the model-building data set supporting the validity and predictive ability of the model The methodology could be adopted to estimate daily or other off-peak hour pedestrian counts.

Table 3. Models for Evening Peak Hour Pedestrian Counts.

Variable	Model 1			Model 2			Model 3		
	Estimate	t-stat	VIF	Estimate	t-stat	VIF	Estimate	t-stat	VIF
Intercept	120.10	1.98		124.56	2.37		86.56	2.17	
NLANES	39.94	4.43	3.8	40.63	5.26	3.1	39.33	5.11	3.0
AVEINC	-7.949	-3.66	4.2	-8.113	-4.34	3.5	-7.467	-4.17	3.1
ln(COMMAREA)	1.164	0.18	1.7	--	--	--	--	--	--
ln(RESAREA)	20.99	4.19	1.7	21.45	5.25	1.3	19.82	5.16	1.1
NSTOPS	-3.38	-1.06	1.3	-3.28	-1.10	1.2	--	--	--
R^2	0.820			0.819			0.797		
$R^2_{(Adj)}$	0.720			0.747			0.742		
F Stat	8.19			11.33			14.44		
S	25.726			24.448			24.672		
C_p	6.0			4.0			3.1		
PRESS	15272.8			10970.1			10268.4		
$R^2_{(Pre)}$	0.538			0.662			0.689		

Note: --indicates the variables not included in the model

$t_{critical}$ 90/95 = 1.77/2.16

The numbers in shaded boxes are not significant at 95% confidence level

References

Berkovitz, A. (2001). The Marriage of Safety and Land Use Planning: A Fresh Look at Local Roadways. *Public Roads*, Vol. 65(2).

Handy, S.L. (1996). Urban Form and Pedestrian Choices: Study of Austin Neighborhoods. In *Transportation Research Record: Journal of the Transportation Research Board, No.1552*, TRB, National Research Council, Washington D.C., pp. 135-144.

Hess, P.M., A.V. Mouden, M.C. Snyder, and K. Stanilov (1999).. Site Design and Pedestrian Travel. In *Transportation Research Record: Journal of the Transportation Research Board, No.1674*, TRB, National Research Council, Washington D.C., pp. 9-19.

Keall, M.D. (1995). Pedestrian Exposure to Risk of Road Accident in New Zealand. *Accident Analysis and Prevention*, Vol.27, 1995, pp. 729-740.

Kutner, M.H., Nachtsheim, C.J., Neter, J. (2004). Applied Linear Regression Models. 4th Edition, McGraw-Hill/Irwin Publishers, Chicago, Il.

Loutzenheiser, D.R. (1997). Pedestrian Access to Transit: Model of Walk Trips and their Design and Urban Form Determinants around Bay Area Rapid Transit Stations. In *Transportation Research Record: Journal of the Transportation Research Board, No.1604*, TRB, National Research Council, Washington D.C., 1997, pp. 40-49.

Mouden, A.V., P.M. Hess, M.C. Snyder, and K. Stanilov (1997). Effects of Site Design on Pedestrian Travel in Mixed Land-Use, Medium-Density Environments. In *Transportation Research Record: Journal of the Transportation Research Board, No.1578*, TRB, National Research Council, Washington D.C., pp. 48-55.

Pushkarev, B., and J.M. Zupan (1971). Pedestrian Travel Demand. *Highway Research Record* Vol.355, Highway Research Board, Washington, D.C.

A Robust Dynamic Segmentation Tool for Highway Safety Analysis

Kaiyu Liu[1] and Albert Gan[2]

[1] Research Associate, Lehman Center for Transportation Research, Florida International Univ.; PH (305) 348-4103; FAX (305) 348-2802; email: liuk@fiu.edu
[2] Associate Professor, Lehman Center for Transportation Research, Florida International Univ.; PH (305) 348-3116; FAX (305) 348-2802; email: gana@fiu.edu

Abstract

This paper describes a dynamic segmentation application that allows roadways to be quickly segmented for highway safety analysis. After roadways are segmented (either variable-length or fixed-length), the system can automatically calculate the crash statistics associated with each segment, which can then be used to identify high-crash locations, to model the relationships between crash experience and geometric features, etc. The system can output segmented records in different formats, including in tables, 2-D and 3-D charts, and GIS shape files.

Introduction

A major duty of traffic safety engineers is to identify high crash locations and to perform crash analysis on these locations to identify potential countermeasures for implementation. The process requires the use of both roadway and crash information, which is currently stored in separate databases. In Florida, traffic and geometric information for its state roadway system is stored in the Roadway Characteristics Inventory (RCI) system, which includes more than 200 geometric features (e.g., functional class, AADT, number of lanes, median width, etc.) with over one million records [1]. Relating crash data to these segments makes it possible for various safety analyses to be performed, including identifying high crash locations, modeling the relationships between crash experience and various geometric and traffic design features, etc. While many software applications have been developed for planning and operational analysis, software systems to aid in such safety analyses have been noticeably missing. This paper describes an innovative desktop system that can dynamically generate roadway segments from the RCI roadway data based on selected roadway features and to automatically aggregate crash information to compute crash statistics for these segments.

System Components

Figure 1 shows the architecture of the system, which consists of four major functional components: Preprocessor, Crash Selector, Segment Generator, and Result Viewer. Each of these components is detailed below.

The Preprocessor component perform three major functions: (1) to reorganize the original RCI data to create an efficient database suitable for the query operations of

this application; (2) to merge separate traffic or geometric feature data files into a common RCI database; and (3) to create coordinates for crash records.

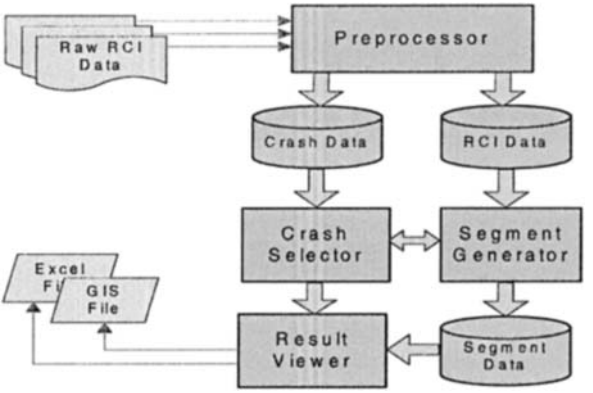


Figure 1. System Architecture.

The Crash Selector component allows specific crash records to be used for analysis. The current crash database includes the 1990-2002 crash data for the Florida state roadway system. It includes more than 30 variables that can be used as filters to include only certain types of crashes in the analysis. Examples of these variables include district, county, time, vehicle type, roadway type, crash severity, harmful event, driver's age, driver's gender, alcohol/drug involvement, etc.

The Segment Generator allows the users to select from more than 100 features such as number of lanes, median width, functional class, etc. The users can also choose to regroup existing RCI feature options such as grouping trucks and buses as one vehicle category. Crash statistics can be dynamically summarized based on the new category defined. Based on the selected RCI features or categories, segment type, and crash types, the Segment Generator can automatically generate uniform roadway segments (i.e., none of the selected roadway features change within 1-mile segment) and compute the crash statistics associated with each segment. In this case, the uniform segments are of variable lengths. The users can also choose to use fixed-length segments, which further divide each uniform segment into smaller sub-segments of a selected length. These sub-segments are usually used in modeling of relationships between crash experience and traffic and geometric features.

The Result Viewer component provides various tools to view the segmentation result and to output segmented records or crash records in different formats, including tables, charts, and GIS shape files, for various analyses.

The system is a typical Windows desktop program designed to run on Microsoft Windows operating systems. It is fully stand-alone and was developed using Visual Basic 6.0, Microsoft Access 2000, and ESRI MapObjects 2.1 Developer Library.

Spatial Data Model and Structure

The system spatial data model is built on the state road lines that are in ESRI's shape file format. Major data sets include RCI data and crash data, with segment data being derived from both data sets by the system.

Roadway Data. About 100 RCI features are available in the current roadway database of the system. The database was first restructured to reduce data redundancy and increase query efficiency. Tables 1 and 2 show the original and redefined RCI records, respectively. In the original RCI data table, the record for a roadway segment is assigned "L" or "R" if the roadway segment is divided as the left or the right side of the road, or "C" if a roadway undivided (i.e., composite). The restructured table merges both sides of a roadway segment into one record, assigned "D", which means that the segment is divided. The records assigned a "C" originally are kept the same, except that the roadway side is re-assigned to "U" for "Undivided". The restructured records offer obvious advantages. Firstly, it reduces the database records significantly, which improves the query and search speed. Secondly, it avoids overlapped segment records so that the three basic fields for roadway definition, Roadway, BeginMilepost and EndMilepost, can be used as the primary key to uniquely identify records and be indexed for more efficient database operations.

Table 1. Example of Original Roadway Records (with partial fields shown).

Roadway	Begin Milepost	End Milepost	Rdwy Side	Functio n Class	No. of Lanes	Shld TypeL	Shld TypeR
87000247	0	0.938	C	17	2	3	3
87000247	0.938	0.949	C		0		
87000248	0	0.6	R	17	2		6
87000248	0	0.6	L	17	2	3	

Table 2. Example of Restructured Roadway Records (with partial fields shown).

Roadway	Begin Milepost	End Milepost	Rdwy Side	Function Class	No. of Lanes	Shld TypeL	Shld TypeR
87000247	0	0.938	U	17	2	3	3
87000247	0.938	0.949	U		0		
87000248	0	0.6	D	17	2	3	6

Crash Data. The key fields in the crash database are roadway and milepost, both fields are referred to as crash location spatially. This database also includes other crash-related information, such as crash time, area, crash type, roadway conditions, first harmful event, etc.

Segment Data. The Segment Generator generates this table dynamically based on the selected RCI features and segmentation method. Three types of segment records can be dynamically generated. Table 3 shows an example of the roadway database in the system. The first type of segment records considers only the segment length

requirements. Table 4 shows an example based on a one-mile maximum length and 0.1-mile minimum length. The first smallest segment in Table 3 goes from milepost 0.000 to 0.029. Because this segment is shorter than one mile, the next segment, from 0.029 through 1.127, is concatenated. When the concatenated segment reaches one mile, it stops and splits this segment at milepost 1.000. The next new segment starts at 1.000. This segmentation process continues until the end of the roadway. If the last segment of this roadway is shorter than 0.1 mile, it would be dropped because it does not meet the 0.1-mile minimum segment length.

Table 3. Example of Roadway Records (with partial fields shown).

Roadway	BeginMilepost	EndMilepost	RdwySide	Function Class	No. of Lanes
87001000	0	0.029	D	6	4
87001000	0.029	1.127	D	6	4
87001000	1.127	1.625	D	16	4
87001000	1.625	2.007	D	16	4
87001000	2.007	2.129	D	16	4
87001000	2.129	2.372	D	16	4
...	...	...	...	...	...

Table 4. Output Segment Data Based Only on Segment Length Requirements.

Roadway	BeginMilepost	EndMilepost	...
87001000	0.000	1.000	...
87001000	1.000	2.000	...
87001000	2.000	3.000	...
...	...	...	...

In the second type, only the RCI features are considered in segmentation. As shown in Table 5, the first record stops at milepost 1.127 because the next segment has a different functional class. As more features are selected, more output records will be generated as the probability of roadway changes in one of the selected features increases.

Table 5. Output Segment Data Based Only on Roadway Features.

Roadway	BeginMilepost	EndMilepost	Function Class	No. of Lanes	...
87001000	0.000	1.127	6	4	...
87001000	1.127	2.543	16	4	...
87001000	2.543	3.842	16	6	...
...	...	...	...	...	...

The third type of segment output considers both the RCI features and the segment length requirements. Using the same minimum length requirements as for the first type, the resulting table is shown in Table 6. The first record shows that, instead of going to 1.127, the record stops at milepost 1.000 to meet one-mile maximum length. This table also shows some calculated crash statistics for each segment, including the overall crash rate and crash numbers stratified by crash severity.

Table 6. Output Segment Data Based Only on Both Roadway Features and Minimum Segment Length Requirements.

Roadway	Begin Mile-post	End Mile-post	Function Class	No. of Lane	ADT	Crash Number	Crash Rate	Fatal	Injury	PDO
8700100	0.000	1.000	6	4	1141	8	64.02	0	3	5
8700100	1.000	1.127	6	4	1210	2	118.8	0	2	0
8700100	1.127	2.127	16	4	1693	8	43.13	0	5	3
8700100	2.127	2.543	16	4	4360	5	25.18	0	3	2
8700100	2.543	3.543	16	6	4579	27	53.84	2	15	10
8700100	3.543	3.842	16	6	4333	3	21.15	0	3	0
...	...	...	...	...	...	...	...	...	...	...

Input User Interfaces

Figures 2 shows the main user interface for specifying methods for roadway segmentation and crash types to be included. The segmentation screen allows the user to select the needed RCI features and to define the segmentation rules. This screen also allows the users to select crash features for calculations, so besides the total crash numbers and rates for each segment, the system would tally crash numbers for each type of selected crash feature value. The crash selection screen (not shown) provides various options that can be used as filters to specify the needed crash data for analysis. The category selection screen (not shown) allows the users to define new categories for output segments such as combining fatal and injury crashes if it is desirable to only differentiate between crashes involving casualties and crashes involving only property damages.

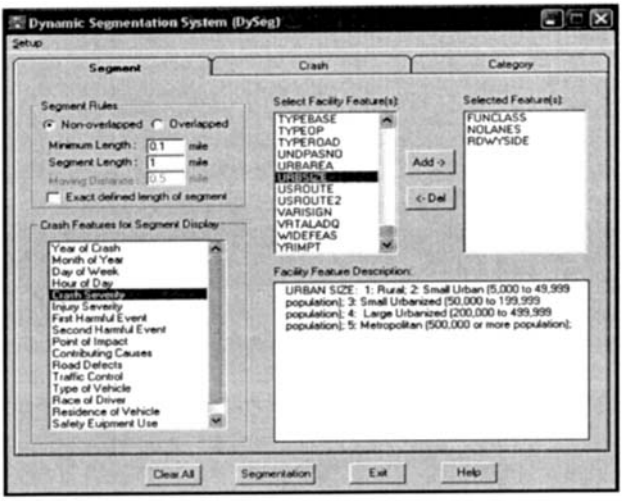


Figure 2. User Interface for Segmentation Specifications.

Sample Output

Figure 3 shows a sample output of the system, which includes a map display showing the segmented result in a GIS thematic style. The Variable dropdown box shown on the screen allows the user to select to display certain types of crashes dynamically. Optionally, the users can select to display the same output by either table or charts. In addition, the users can display 2- and 3-dimensional maps showing locations of individual crashes. All outputs can be saved to Excel and/or ESRI GIS shape file formats for further analyses, including statistical modeling relationships between crash experience and geometric features.

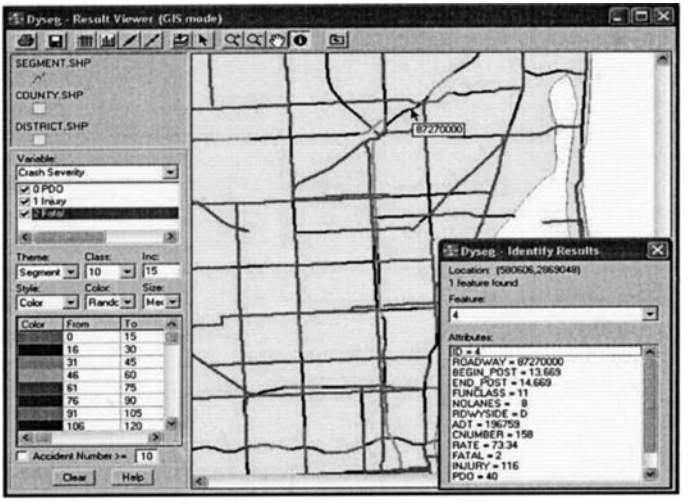


Figure 3. GIS Display of Segmented Results.

Summary

This paper has described a user-friendly software system for highway safety analysis. The system is able to combine roadway and crash information to dynamically generate segment records that contain calculated crash statistics, which can be used for analyses such as identifying high crash locations, serving as input to crash modeling, etc. Although the system is currently applicable to only Florida RCI and crash data, the system architecture, data model and structure can be helpful for the development of similar applications in other states.

Reference

[1] *RCI Features and Characteristics Handbook*, Office Handbook, Transportation Statistics Office, Florida Department of Transportation, June 2003.

3D Traffic Simulation for Intermodal Safety and Security

Lee-Fang Chow [1], Fang Zhao[2], L. David Shen[3], Hongbo Chi[4], Xuemei Liu[5], Jingtao Shan[6]

[1]Lehman Center for Transportation Research, Dept. of Civil & Env. Eng., Florida International University, Miami, FL 33199; PH 305-348-6577; FAX 305-348-2802; email: chowl@fiu.edu
[2]Lehman Center for Transportation Research, Dept. of Civil & Env. Eng., Florida International University, Miami, FL 33199; PH 305-348-3821; FAX 305-348-2802; email: zhaof@fiu.edu
[3]Lehman Center for Transportation Research, Dept. of Civil & Env. Eng., Florida International University, Miami, FL 33199; PH 305-348-1869; FAX 305-348-2802; email: shenll@fiu.edu
[4]Lehman Center for Transportation Research, Dept. of Civil & Env. Eng., Florida International University, Miami, FL 33199; PH 305-348-3822; FAX 305-348-2802; email: hongbo.chi@fiu.edu
[5]Lehman Center for Transportation Research, Dept. of Civil & Env. Eng., Florida International University, Miami, FL 33199; PH 305-348-3822; FAX 305-348-2802; email: xuemei.liu@fiu.edu
[6]Lehman Center for Transportation Research, Dept. of Civil & Env. Eng., Florida International University, Miami, FL 33199; PH 305-348-3822; FAX 305-348-2802; email: jshan002@fiu.edu

Abstract

This project involved the construction of a 3-D Miami downtown transportation simulation model to study and visualize the traffic operations of different transportation modes and investigate the impacts of potential security and safety events in the Miami downtown area. The 3-D model (including buildings, highways, and elevated guideways) was constructed using AutoCAD and a Geographic Information system (GIS). A traffic simulation model was built based on Cube Base/Voyager/Dynasim. In this study, the 3-D simulation model was used to assess the elapsed time for a specific evacuation scenario. The results show that it would take at least six hours to evacuate the traffic in the downtown area to the nearest external stations for a given emergency event.

Introduction

There has been an increasing concern over the safety and security of the transportation infrastructure in the U.S. Much attention has been directed to airports, seaports, public transit, etc. However, in an urban area where multi-modes in a transportation system are tightly integrated, disruptions caused by either an incident or terrorist attack will affect the regional network including all modes. This is the

case of downtown Miami, where interstates, guideway transit, seaport, railroad, and local roadway system come together. As a critical link in the southeast Florida transportation network, one major disaster to this transportation hub will mean the paralysis of the entire downtown or even the regional transportation system and subjecting workers and visitors in the downtown area to life-threatening dangers.

Technological solutions to address security threats should identify critical transportation assets and infrastructure, proactively mitigate risks, and develop real-time response to reduce impacts after a security incident. An emergency management information system is a decision support system that integrates these phases of emergency management (Tzemos and Burnett 1995) and should be able to display and analyze the spatial relationships among emergency management facilities and resources, transportation routes, and population at risk. Emergency planners should also be able to visualize the animated temporal progression of both the hazard situation and the evacuation of the affected population from a disaster site (Kwan and Lee 2005).

This project involves the construction of a 3-D Miami downtown transportation simulation model to study and visualize the operations of different transportation modes and investigate the impacts of potential security and safety events in Miami downtown as well as on the regional transportation network. Traffic simulation is performed in Dynasim, designed for visualizing traffic conditions based on simulated driver behaviors (Citilabs 2005). The program performs operational analysis of complex traffic on roadways and animates the flow of automobiles, trucks, buses, rail, and pedestrians in a 3-D transportation system. It is also possible to render high quality 3-D animations.

Study Area

Downtown Miami is a critical link in the southeast Florida transportation network due to its strategic location and concentration of activities. In the downtown area, there are federal, state, county, and city government offices. In addition to diplomat missions, it has a financial center with a concentration of national and international banks and financial institutions. It is also a popular tourist destination and an art and cultural center, with hotels, marinas, stores, shopping malls, parks, arenas, museums, schools, universities, and a new performing art center in the area. Approximately 66,000 people work in Miami downtown. The area is also close to the Port of Miami, which is recognized as the Cruise Capital of the World and the Cargo Gateway to the Central and South Americas, with nearly four million cruise passengers passed and over nine million tons of cargo and over one million TEUs (twenty-foot equivalent unit containers) in fiscal year 2003.

3-D Modeling and Animation Process

To create a 3-D simulation involves first creating a traffic layer in Cube Base/Voyager and then adding extra 3-D rendering layers in a graphical format

compatible with Dynasim for buildings and street landscaping, resulting in a multi-layer file describing a 3-D real-world transportation system. The process of creating a 3-D transportation animation system has the follow steps:

- *Traffic network data collection.* The traffic data collected include draft 2-D network alignment developed from the highway network created in the 2000 transportation model for Miami-Dade County, geometric information from field survey, signal timing plans from Dade County Public Works (DCPW) Traffic Control Center (TCC), and vehicle O-D table estimation from the 2000 transportation model for Miami-Dade County.
- *Construction of transportation model.* A total of 946 highway links, including 234 centroid connectors, are created in Cube.
- *Traffic assignment.* Cube is used to perform user equilibrium traffic assignment to obtain the traffic flow on each link and the turning volumes at each intersection in the network. The turning volumes are the crucial information in the subsequent processes in Dynasim.
- *Exporting 3-D traffic layer.* Traffic flows generated from the traffic assignment process in Cube are exported to Dynasim.
- *Creating 3-D rendering building/object layers.* Three-dimensional images of buildings and landscaping are created and overlaid with the traffic layer. Roads are created based on edge-of-road GIS data and the building footprints are extruded into 3-D according to the building heights from Light Detection and Ranging (LIDAR) data. Elevated guideways for Metrorail, Metromover, highway ramps and interchanges are also created.
- *Perform simulation and animation.* The simulated traffic flows as well as the rendered building and landscaping represent a realistic transportation network as shown in Figure 1.

Figure1 Morning Traffic Condition on Biscayne Blvd in Front of Bay Front Park

Simulation of Emergence Evacuation

With the necessary traffic data, the 3-D transportation network developed in this study allows the users to better orient themselves to the geo-spatial content in the

network. The decision makers are thus able to speed up the process in the evaluation of different alternatives that are developed for responding to a given scenario for transportation management, emergency evacuation, etc. In this study, a scenario is developed to illustrate the implementation of the 3-D transportation system for evaluating alternatives for emergency evacuation. The scenario is described as an explosion occuring at 9:00 AM on a typical weekday inside a building located at the corner of 1st Street and 1st Avenue. All non-emergency vehicles in the downtown area must be evacuated and the time needed to evacuate the downtown traffic needs to be estimated.

The first step in assessing the elapsed time needed to evacuate the downtown is to determine the travel demand in the study area at the time of the explosion. The procedures to estimate the number of vehicles that may need to be evacuated is described below.

1. *Calculation of K-factor for each time period.* In a typical travel demand model, daily volumes in term of origin-destination (O-D) vehicle trips at the zonal level are assigned to the network links. For this study, daily O-D volumes are first stratified into hourly O-D volumes. Hourly volumes are then aggregated to estimate the number of vehicles in the study area prior to the emergency event. The traffic count data available from the 2004 Florida Traffic Information (FTI) CD-ROM, published by the Transportation Statistics Office of the Florida Department of Transportation (FDOT), are used to estimate the K-factors, namely the proportion of Annual Average Daily Traffic (AADT) occurring in a given hour. A total of 30 portable traffic monitoring sites (PTMSs) surrounding the study area are selected to calculate the area-wide K-factors by dividing the total traffic volumes collected from these PTMSs at a given hour with the total 24-hour counts. The aggregated K-factor from midnight to 9 AM is 0.266. Using this factor, the O-D matrix in terms of vehicle trips during the specific time period is calculated. The resulting OD table represents the traffic demand for the downtown area before 9 AM.

2. *Calculation of D-factor.* The O-D vehicle trip matrix for traffic demand before 9 AM estimated in the previous step include external-external (E-E) trips, internal-internal (I-I), external-internal (E-I), and internal-external (I-E) trips. Both the E-E and I-E vehicle trips are the traffic exiting from the downtown area. In reality, some of the trips would be traveling on the network links to their destinations via external stations when the incident occurs. For simplification, these trips are assumed to have already exited the downtown area and are thus excluded from the trip matrix. The evaluation traffic is hence composed of the E-I trips. Since the K-factors are calculated in terms of two-way traffic, the E-I trips are adjusted to reflect the directional distribution in the traffic flow. The PTMS traffic counts are again used to calculate the D-factors, namely the percentage of total, two-way peak hour traffic occurring in the peak direction. The traffic count data show that the D-factors are approximately equal to 0.58 during the morning peak for vehicles entering the downtown area. These factors are then applied to estimate the number of vehicles that went into the study area. For a given E-I zonal pair, for example, the number of

vehicle trips is equal to 0.58 of the sum of the original E-I trips and the corresponding I-E trips.

3. *Compilation of vehicle trip table for evacuation traffic.* Each column sum in the vehicle trip matrix from Step 2 for a given internal zone is the total traffic demand prior to the explosion. No demands are derived for the external zones since all E-E and I-E trips are set to zero. The integrated vehicle trips then become those that need to be evacuated.

4. *Distribution of vehicles from the zones they locate.* The vehicle trips from Step 3 for a given internal zone are re-allocated to the nearest external station that is identified using the shortest path method. Long travel time is assigned to those highway links near the zone of the explosion site to prohibit vehicles traversing through this particular area.

5. *Determination of the elapsed time for vehicle evacuation via iterative trials.* Simulation does not provide exact time length required for evacuation. Therefore, after the destinations for each internal zone are defined, the simulation time specified as the elapsed duration for the evacuation traffic is gradually increased with a one-hour increment. For example, the entire network is congested when the demand is specified in terms of hourly volumes. Less congested traffic condition is observed when the same demand is specified in the simulation for a period of 6 hours instead of one hour.

Statistics such as traffic flows, travel times, and speeds may be calculated by placing data collection points in the strategic areas of the simulation. Cube Dynasim constantly tracks every vehicle at the data collection points. The statistics are collected and aggregated. In reality, traffic flows vary from one day to the next but they fluctuate around an average value. The simulation produces vehicle movements in the network according to realistic driver behavior based on statistical observations. The traffic fluctuations experienced day to day are reflected through multiple simulation runs. Ten simulation runs were performed in Cube Dynasim using different series of random numbers and the statistical results are averaged. Table 1 shows the simulation results at several on-ramps of the freeways, Brickell Avenue, and the west bound of Port Boulevard.

Table 1. Travel Speeds (mph) for Different Evacuation Times

Evac. Duration	Ramp to I-395 from US1	Ramp to I-395 from SE 1st Ave	Ramp to I-95 from SW 3rd Ave	Ramp to I-395 from Miami Ave	Brickell Ave	WB Port Blvd
1 hour	0.89	1.83	4.14	3.91	1.83	1.83
2 hours	3.13	11.33	3.70	4.68	2.17	2.17
3 hours	6.23	13.11	4.43	3.65	3.35	3.35
4 hours	7.71	13.31	3.94	7.43	7.00	7.00
5 hours	10.09	12.67	8.28	8.92	10.53	10.53
6 hours	11.06	13.59	9.19	9.82	12.27	12.27

Conclusions

A network for the Miami downtown area has been created in Cube Base and converted to Cube Dynasim using the process described in this paper. The animated 3-D transportation system being developed may be integrated with other components to help decision-makers visualize the effects of newly proposed transportation management strategies, traffic accidents, terrorist attacks, etc., on the transportation infrastructure.

It is demonstrated in this study that the 3-D simulation model may be used to assess the elapsed time for a specific evacuation scenario. The results show that it would take at least six hours to evacuate the traffic in the downtown area to the nearest external stations under a given emergency event. One limitation of this study is that the vehicle trips in the study area is derived under simplified condition as the E-E and I-E trips are excluded from the vehicle trip matrix. Additionally, only the motorized evacuation trips are considered and the time for evacuees to access their vehicles are ignored. The allocation of evacuation traffic to the nearest external stations may also be deviate from the reality. The variations in the time that road users react to an evacuation order are also not considered in the study. Finally, different responsive strategies such as allowing evacuees to use opposite traffic lanes, adjusting signal timing plans, etc., that would reduce evacuation time are not evaluated in the analysis. These strategies may be analyzed to identify the optimal responsive plan to the given evacuation scenario.

References

Citilabs, Inc. (2005b), Cube Dynasim: Dynamic Multimodal Microsimulation, http://www.citilabs.com/dynasim/index.html.

Federal Emergency Management. Information System (FEMIS) (2002). *System Adminstration Guide for FEMIS* Version 1.5.3. Prepared for the CSEPP Office United States Army Soldier and Biological Chemical Command under a Related Services Agreement with the U.S. Department of Energy.

Kwan, M. and J. Lee (2005). "Emergency Response after 9/11: the Potential of Real-Time 3D GIS for Quick Emergency Response in Micro-Spatial Environments," *Computers, Environment and Urban Systems*, Volume 29, pp. 93-113.

Tzemos, S. and R. Burnett (1995). "Use of GIS in the Federal Emergency Management Information System (FEMIS)." *In Proceeding of 1995 ESRI User Conference*, Palm Springs, California.

Effect of Vehicle to Vehicle and Vehicle to Infrastructure communication systems on Transportation Safety

S. Diwan[a], B. Dalla Chiara[b,*], F. Deflorio[b]

[a] *Transportation Systems Engineering Group, IIT Bombay, Powai, Mumbai 400076, India.*
[b] *Dept. DITIC-Transport, I Faculty of Engineering Politecnico di Torino, 10129, Torino, Italy.*

Abstract

The world of telecommunications is providing us with a lot of communication paradigms that can be utilized as a means of wireless communication between mobile and stationary agents or between mobile agents themselves. As transportation engineers, we are interested in knowing the schemes that can be applied to the field of transportation engineering to alleviate traffic problems like safety, congestion, environmental degradation and off late, energy consumption by the vehicles, to name a few.

Intelligent Transportation Systems are undergoing a transition from demonstration projects to becoming part of the mainstream set of options available to transportation planners. Hence, evaluation of ITS is one of the most critical and important steps to be taken before any ITS technique can be deployed.

Safety has been recently emerging as an area of increased concerns, attention and awareness within transportation engineering. It has been extremely difficult to evaluate safety for new and innovative traffic treatments.

In this paper, we first review all the available techniques for communication between various vehicles among themselves and with the infrastructure. The aim of this is to have a complete overview of all the possible communication techniques that the world of electronics and telecommunications have proposed in the last years that can be applied for accomplishing the task of enabling vehicles communicate and interact with other vehicles as well as with the infrastructure. Then we propose a method to evaluate safety based on safety parameter called Risk Index (R.I.) to evaluate the use of vehicle to vehicle and vehicle to infrastructure communication technologies for safety.

Keywords: ITS, Vehicle to vehicle communications, Safety, Risk Index Car-following models

---

* Corresponding Author. Tel.: +39-011-5645621, Fax: +39-011-5645699.

E-Mail address: sarvee@iitb.ac.in (S.Diwan), bruno.dallachiara@polito.it (B. Dalla Chiara), francesco.deflorio@polito.it (F. Deflorio).

1. Introduction

Intelligent Transportation Systems continue to blossom in many dimensions. Amongst the most interesting of these is the collection of applications that involve interaction and co-operation between the infrastructure and in-vehicle systems and the interaction between the in-vehicle systems of different vehicles. These kinds of applications fall under vehicle to vehicle and vehicle to infrastructure communication systems.

The research in the field of ITS has been going on since the 80's but the time frame over which the research is going on is changing. We may define the time frame as the reaction time that the driver gets to respond to any emergency on the road. Till date, various technologies have been proposed to increase the time available for the driver to react to any emergency situation or an incident on the road (e.g. VMS, On-board message broadcasting, etc.). Still many incidents take place which give the driver much less reaction time than that made available by the modern technologies, which results in an accident. An accident occurs when the reaction time of the driver is more than the time available for the driver to react to an emergency situation on the road. The vehicle to vehicle and vehicle to infrastructure communication technologies aim to deal with such a small time interval that even if the incident has occurred at such a distance that it is difficult for the driver to react quickly, then the information is quickly transmitted to the follower and an automatic system does the needful, then possibly without the intervention of the driver, to avoid any accident.

For the expansion of ITS, 'ITS communication systems' are expected to be important subsystems to support various ITS. For example, comfortable driving with requested information such as traffic conditions, a weather forecast, and multimedia information are expected to be universally available when driving. Vehicle to Vehicle (V2V) and Vehicle to Infrastructure (V2I or IVC) communication are the primary means for supplying information to the drivers.

The structure of the paper is as follows: in section 2, basic definitions are taken into account in order to clearly understand the concept of vehicle to vehicle and vehicle to infrastructure communication systems; section 3 contains previous works; section 4 gives the reason for choosing the safety application of V2V and V2I communication technology for our analysis; section 5 deals with our proposed approach and demonstration of the use of V2V and V2I communication for safety; section 6 gives the results of the analysis; section 7 and 8 give the discussions and conclusions drawn from the analysis respectively.

2. Vehicle to Vehicle and Vehicle to Infrastructure communication systems

2.1 Definitions:

a) Inter-Vehicle Communication: Inter-vehicle communication is here defined as communication and cooperative exchange of data between vehicles via wireless technologies, within a range varying from a few meters to a few hundreds of meters, in

order to improve road safety, mobility, efficiency and augmenting road capacity. Hereafter, the related technologies will be identified as V2V or IVC.

In V2V communication, data is transmitted and consumed where it is generated and needed. IVC has two key advantages:

- **Direct communication:** Since the vehicles communicate directly without any intermediate base stations, the communication delay is low. IVC is also suitable for the distribution of time-critical data such as emergency notifications in the area of an accident. Furthermore, vehicles can communicate even in rural areas not covered by cellular systems—the communication network is established by the cars themselves and available everywhere.
- **Low service fees:** Normally, IVC does not require communication infrastructure or service provider. Service charges are, generally, completely avoided.

A schematic diagram of the IVC is shown in Figure 1: A typical V2V communication model below.

Figure 1: A typical V2V communication model

(Source: Omae *et al*, 2000)

b) Vehicle to Infrastructure Communication: Vehicle to Infrastructure communication is, here, defined as the co-operative interaction, typically via wireless communications, based on infrastructure and vehicle based systems to improve the safety and performance of the transportation infrastructure and of vehicles along with their operators and users. Hereafter in this document, the related technologies will be identified as V2I.

V2I communication needs infrastructure to be built and then all the participating mobile units (vehicles) can transmit their data to the control center where it can be processed and again sent to the various vehicles in the vicinity as useful information. V2I is not strictly a communication only from an infrastructure to a vehicle but it is a communication of any kind between the vehicle and the surrounding that is related to any kind of useful information for the driver e.g. little lighted panels, blinking panels etc. A schematic diagram is shown in Figure 2: A typical RVC Communication model.

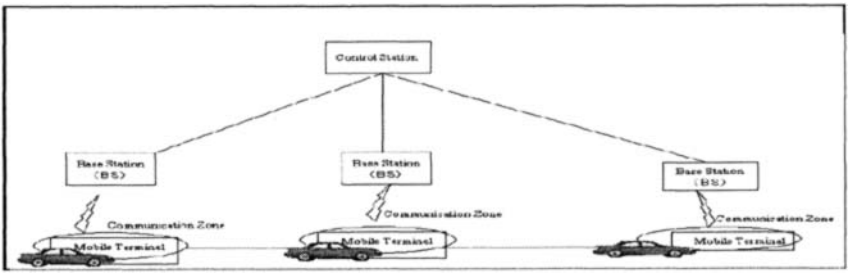


Figure 2: A typical RVC Communication model

(Source: Kubota et al, 2000)

2.2 Application of IVC and RVC technologies

The development and proposal of a lot of new communication techniques among mobile agents as well as between a fixed and a mobile agent enables us to use those techniques for improving the current standard of transportation facilities. The three most talked about applications of IVC and RVC as seen from a transportation point of view are as follows:

1) *Comfort Applications*: This type of application improves passenger comfort and traffic efficiency and/or optimizes the route to a destination. Examples for this category are: Traffic-information system, weather information, gas station or restaurant location and price information, and interactive communication such as Internet access or music download.

2) *Safety Applications*: Applications of this category increase the safety of passengers by exchanging safety relevant information via IVC. The information is either presented to the driver or used to activate an actuator of an active safety system. Example applications of this class are: emergency warning system, lane-changing assistant, intersection coordination, traffic sign/signal violation warning, and road-condition warning. Applications of this class usually demand direct vehicle-to-vehicle communication due to the stringent delay requirements.

3) Capacity Enhancement: Applications of this type aim to augment the capacity of the highways by transmitting information using the IVC technology and hence maintaining the minimum safe distance between the vehicles, thereby increasing the capacity of the highways manifolds. Yet, these types of applications have just been proposed and not tested in a real-time situation. It still is debatable if such kind of applications is possible in future or not but at present this is definitely one possibility. In these kinds of applications, delay requirements are very stringent and a highly reliable communication system needs to be used.

3. Literature Overview

V2V and V2I communication has been a subject of intense study from a feasibility point of view. There have been many studies done in this field as regards the telecommunication point of view. For instance, in [12], the authors have successfully demonstrated that use of Bluetooth protocol and wireless sensor networks for intervehicle communication can increase the safety on the road through data exchange among the vehicles.

There have been a few safety related studies like [3] which uses V2V and V2I communication protocol for cooperative collision warning. The main problem addressed was that of delay in warning messages and removal of redundant warning messages.

There have been many studies relating to the dissemination of data in mobile networks like [4],[1],[6]. In the first one the authors propose a method for scalable information dissemination in highly mobile ad hoc network: segment oriented data abstraction and dissemination (SODAD). The authors have also presented an example application (SOTIS) of SODAD and evaluated it using the network simulation software. It worked well even in cases of low penetration like 2%-3%.
In [1], the authors propose a DOLPHIN protocol for Intervehicle communication. In [6], SPOT communication technique is used.

There have been studies like [9] which propose an IVC protocol in which the information is exchanged in a peer-to-peer manner without involving any public infrastructure utilizing IEEE 802.11 wireless standards.

In [7], the authors propose a device that can be assembled in a vehicle to help the river see beyond what he can physically see and to communicate with other vehicles using short range wireless communication.

There are numerous other studies like [2], [5] etc. to establish that V2V and V2I communication is possible using different protocols. The former uses MAC protocol while the latter uses optical two way LASER system for communication.

4. Reasons for the safety Analysis

There are many benefits that might be expected from V2V and V2I communication. In this section, we try to give reasons for choosing safety benefits to analyze the effectiveness of V2V and V2I communication systems.
Basically there are three elements of any transportation system: the user or driver, the vehicle and the environment or surroundings including the infrastructure. The ways in which these key elements interact with each other is very essential in understanding the effect of communication in ITS. Figure 3: Schematic Diagram of Interaction of key elements of a transportation system shows the schematic diagram of this interaction.

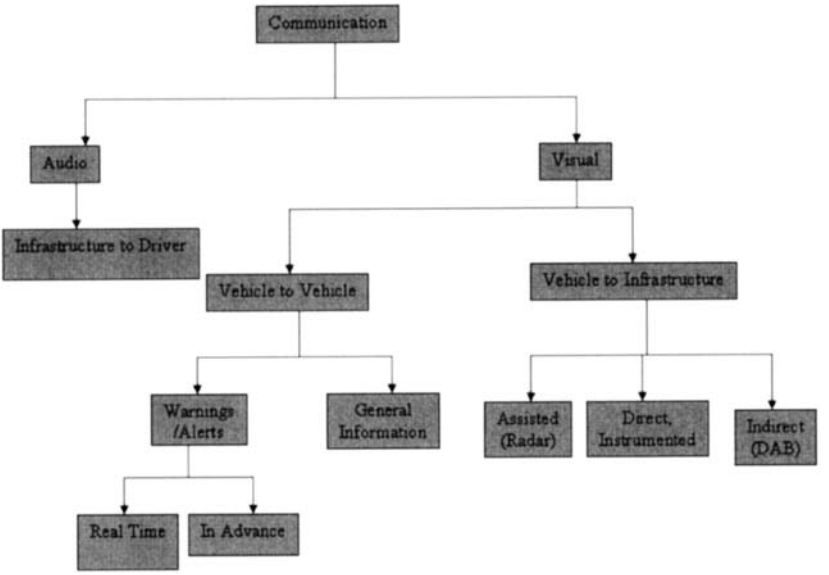


Figure 3: Schematic Diagram of Interaction of key elements of a transportation system

We also took an in-depth look at the interests of various actors of the transportation system. We define there are four major participating groups whose interests are being taken into account. They are: the driver, the authorities who represent the rights of the citizens like the judiciary, the companies managing the infrastructure, most frequently the highways as regards ITS, and the car manufacturers. The most frequent or typical aims of the various participants can be summarized as follows:

1. Driver:
    - Safety
    - Avoid Congestion and Queuing
    - Travel Fast
2. The judiciary/Magistrate
    - Ensuring safety of the citizens
    - Protecting the environment
    - Enforcement of laws
3. The highway authorities or the companies managing highways:
    - To obtain a good income for managing the infrastructure by tolls or other ways
    - Maintaining the infrastructure
    - Providing a good Level of Service
    - To guarantee, directly or indirectly, SOS services
4. Car Manufacturers
    - Obtain good incomes by selling cars and providing maintenance services

- Providing a direct or indirect assistance in case of damage and, possibly, a remote diagnosis of the vehicle
- Respond quickly to SOS, as far as they are concerned.

Table 1: Interests of Various Participants in Transportation system

	Safety	Avoid Congestion	Speed	Law Enforcement	Respect Environment	Earn Money/ Profits	Maint. Of Motor Ways	Good LOS	SOS	Assistance
Driver										
Judiciary										
Companies Managing Infrastructure										
Car Constructors										

On analyzing carefully the interests as summarized in Table 1: Interests of Various Participants in Transportation system, we can see that there is one common interest to all of the above-said players. The driver's main aim is to be safe. The main aim of the judicial authorities is to ensure safety of the citizens. The aim of the highway authorities is to react to the SOS messages, maintain highways properly for safety of the drivers. For the car manufacturers, safety is an important aspect indirectly because nobody would buy their car if it is not safe; drivers ask for safety, hence the car manufacturer's primary aim of making profits can't be fulfilled. They also have to provide a good diagnosis and assistance in case of an emergency for the public to lure into buying their cars.

So we see that, either directly or indirectly, a common aim of everyone involved in the transportation system can be outlined: it is the *safety* of the passengers and of the vehicles.

5. The approach of the research

Traffic accidents have been taking thousands lives every year, outnumbering any deadly disease or natural disasters. As previously discussed and shown, safety is the major concern and interest of the major players of the Transportation system. In the previous section it was shown that one of the major benefits of Vehicle to Vehicle and Vehicle to Infrastructure Communication systems is that it improves safety: in this study, we analyze the effect of V2V and V2I Communication systems on it.

Drivers suffer from natural perception limitations on roadway emergency events, resulting in a delay in propagating and reacting to the emergency warnings. Besides, a person may devote his entire attention to some activity; still, if this person carries on more than one activity at the same time, his attention is fragmented. A loss in concentration of a driver may, according to the probability of finding any obstacle along the path of his car, lead to many unfavorable and dangerous situations on the road, which can be avoided if the driver's attention is to its full capacity throughout the journey. The

fragmentation of driver's attention may be attributed to various factors as roadside distractions like an advertisement board, talking on a cellular phone either with or without a hands-free, talking with co-passenger, listening to the radio or music, thinking of something while driving or even due to tiredness. This is where V2V and V2I Communication systems come handy and useful. They keep the driver alert and attentive throughout his journey or at-least near the incidents on the road by propagating various emergency warnings and signals at appropriate times so as to bring up the attention level of the driver.

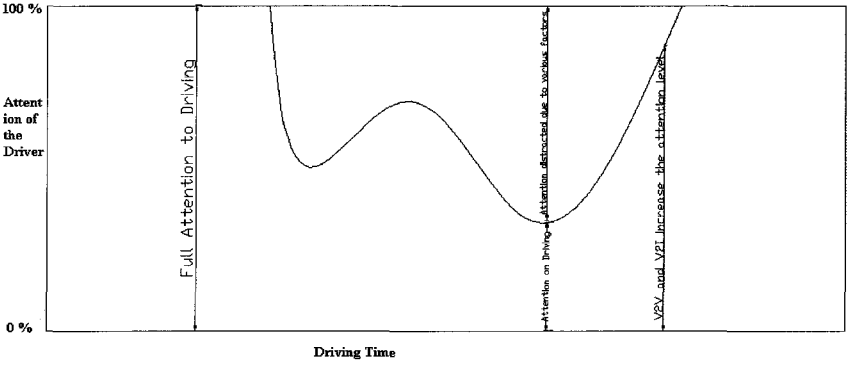


Figure 4: Variation of Driver's attention with time

Attention level of a typical driver can be drawn against time as shown in Figure 4: Variation of Driver's attention with time. Initially, the driver is supposed to drive with full attention for a certain part of the journey and then he may begin to get distracted due to one or more of the various factors mentioned above. V2V and V2I Communication systems bring this graph (that is, attentiveness of the driver) back to 100 % and keep it there till any emergency is there by sending continuous emergency warnings/signals at appropriate times.

As already discussed previously that V2V communication systems are preferred because of low installation and operational costs, the main propagation of warnings may take place through vehicles only. V2I communication systems may act as backup only in case V2V communication system does not work. There may be various reasons for the V2V communication systems to be unable to send information:

- the communication system may get disrupted in the vehicle due to its being involved in an accident;

- Some vehicle may be outside the communication/transmission range of the vehicle, than the message may not get transferred to the follower which is at a great distance.

In short, V2I communication may be useful during a low traffic density and also as a back up of the V2V communication system.

After showing how V2V and V2I communication systems affect safety, influencing driver's attention, we now move on to the main aspects of accidents, as far as this paper is concerned.

Accidents may be classified into two categories: primary and secondary. Primary accidents are the accidents that take place at the first place without the influence of any other incident on the road. We define secondary accidents as the accidents that take place due to the presence of a primary accident on the road. Examples of such accidents are plenty and there is a need to study these classes of accidents. Secondary accidents often take place in big numbers due to poor weather conditions like dense fog, heavy rainfall/snowfall or due to bad light conditions when the visibility is very poor.

At present, the authors are not very sure of use of V2V and V2I communication technology to reduce primary accidents because it is a very complex case. Consider the following example.

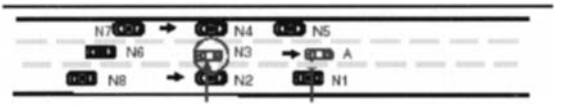


Figure 5: A typical highway scenario

Taking the case as shown in Figure 5: A typical highway scenario, suppose that vehicle N3 breaks suddenly and sends a warning message like STOP to all its neighbors. This message is perfectly fine for vehicle N6 and for vehicle A to some extent. But just think what will happen if vehicle N4 and N2 stop suddenly and N7 and N8 are outside the range of N3. It is very likely that N7 and N3 will crash into their leaders causing an incident on the road, which was not desired. So, a lot of study still needs to be done for selective message broadcasting based on the actual sharp position of the vehicle as it moves on the road. As of now, we think that V2V and V2I communication may be extremely useful for preventing secondary accidents.

So, from now on the attention is focused to the effects of V2V and V2I communication systems on safety and more specifically on the prevention of secondary accidents.

5.1 Safety Indicator

The implementation of an ITS application cannot be justified only by the increase in performance of a network if it implies a decrease in user safety. Therefore, the knowledge of the safety level is crucial for making good decisions. However, safety analysis also has an indirect impact on the evaluation process. Besides compromising health, the presence

of new accidents will create congestions, thus a decrease in network performance. The safety and performance evaluations are directly linked, the first having an important influence on the second.

Safety needs to be evaluated before ITS application can be deployed, therefore, we cannot depend on accident data sheets to evaluate safety. Safety Evaluation of ITS applications needs to be done beforehand.
An analytical Gipp's car-following model has been utilized to demonstrate the safety effects of V2V and V2I communication systems. A simulation environment has not been used because the vehicles are simulated and analyzed using various models, which makes it virtually impossible to simulate an accident with a general simulator. All the vehicles follow Gipp's car-following model, which makes the simulation world an ideal one where no accident can take place. So we resorted to the analytical technique.

We use simple Kinematics to define a Risk Index (R.I.). It is obvious that *a collision between two cars will happen if the time required to stop the leader is greater than the time required for the follower to reach the leader.*

This means that the limiting condition for an accident to take place is:

Time required to stop the follower = Time taken by the follower to reach the leader
i.e.;
$T_{stop} = T_{follower}$

$T_{stop} = D_{rel} / V_{rel}$

We have to define an indicator to indicate the safety index or on the contrary, the Risk Index.

We define the Risk Index (R.I.) as follows:

$$R.I. = (T_{stop} * V_{rel})/D_{rel}$$

Where,
R.I. = Risk Index
T_{stop} = Time required by the follower to stop completely
V_{rel} = Relative Speed of the follower with respect to the leader
D_{rel} = Distance between the follower and the leader

T_{stop} consists of three parts. It is composed of the driver's reaction time, the vehicle characteristics, like its braking capacity and condition of tires and the pavement conditions, like the coefficient of friction, presence of oil or ice, etc. V2V and V2I communication can somehow interfere in all the three components to reduce their value. The influence of V2V and V2I communication on driver's reaction time is the most important one; it is discussed and shown in the first part of this section. V2V and V2I communication also influence the second component if they act directly on the vehicle

without the driver's interference. The third component might be influenced in the future by ITS if we have smart highways and advanced pavements such that they can improve their condition by themselves without any human intervention.

It is important to note at this point that R.I gives a value that has got a physical meaning. If the R.I. is greater than 1, it means that there is a danger of an accident. If it is less than 1, there is no danger of an accident. An R.I. value less than 1 indicates a safe condition.

5.2 Demonstrative Problem

To demonstrate the use of V2V and V2I communication, let us consider a simple scenario with five cars traveling in a single lane so that we can also demonstrate the propagation of risk with time along a lane.

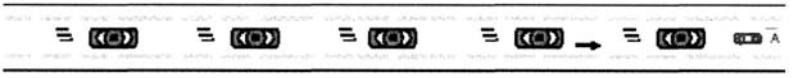


Figure 1: Figure representing a typical highway scenario with five cars traveling in a single lane

Firstly, a typical highway scenario is described hereafter.
Vehicle A is the endangered vehicle that is either broken down on the road or decelerating very sharply so that it becomes a danger for the following vehicles. Vehicle 1 reacts sharply to the incident on the road caused by vehicle A. Vehicle 1 decelerates and is followed by other vehicles in the same lane.
The properties of all the vehicles like maximum acceleration, maximum deceleration, etc. are assumed to be the same. The speeds and distances of all the five vehicles are as follows:

Table 2: Vehicle Characteristics

Vehicle Number	Velocity (m/s)	Position from a reference point (m)
1	40	80
2	40	65
3	50	40
4	50	25
5	50	10

The length of each vehicle is assumed to be 5m and the deceleration capacities of vehicles are $4m/s^2$ and $3m/s^2$.
The Gipp's car following model was implemented in MS Excel using a simple Visual Basic code.

Now, we define what happens when V2V and V2I Communication system is there and when it is not present. When vehicle 1 perceives (without the intervention of the driver) an accident in front of it, it breaks and decelerates with its full capacity, which in turn gives signal to vehicle 2 to decelerate. Vehicle 3 can decelerate only when vehicle 2 decelerates because it cannot physically see vehicle 1 decelerating and so vehicle 3 gets

less time to react and if emergency situation is grave it may even collide to vehicle 2 due to less time to react. But if V2V and V2I communication is there, as soon as vehicle 1 decelerates, it sends emergency warnings to all the vehicles as well as to the infrastructure, which in turn sends it to all the concerned vehicles. Because of this message, vehicle 3 need not wait for vehicle 2 to react. Instead, it can react directly to vehicle 1's action as soon as vehicle 1 decelerates. So the basic difference in this case is that, instead of vehicle 2 being the leader of vehicle 3 and influencing its dynamics, it is the vehicle 1, which influences the vehicle dynamics of 3.
To put it simple terms, in presence of V2V and V2I communication systems, vehicle 1 becomes the leading vehicle for vehicle 3 instead of vehicle 2 and hence vehicle 3 has ample time to react and avoid any dangerous situation. Similar is the case for vehicles 4 and 5. In presence of V2V and V2I communication technologies, these vehicles can directly react to the actions of vehicle 1 and treat vehicle 1 as their leader vehicle.

6. Results:

The results from the analysis can be summarized in the following tables. Risk Index (R.I.) was calculated in absence and presence of V2V and V2I communication.

Table 3: Without V2V and V2I communication

TIME	R.I. for Vehicle 3	R.I. for Vehicle 4	R.I. for Vehicle 5
0	0	0	0
1	5.698	14.151	0
2	0.519	17.36	14.236
3	0.277	9.711	22.675
4	0.151	3.513	18.088
5	0.042	1.832	3.593
6	0.051	1.057	1.6325
7	0.133	0.617	0.8695
8	0.204	0.33988	0.4743
9	0.265	0.1556	0.2416
10	0.316	0.031	0.0964

Table 4: With V2V and V2I communication

TIME	R.I. for Vehicle 3	R.I. for Vehicle 4	R.I. for Vehicle 5
0	0	0	0
1	0.370	0.581	0
2	0.103	0.217	0.340
3	0.140	0.172	0.303
4	0.288	0.150	0.306
5	0.344	0.150	0.303
6	0.323	0.160	0.292
7	0.246	0.169	0.273
8	0.134	0.165	0.242
9	0.0001	0.142	0.197
10	0.152	0.105	0.142

From the above tables, graphs were plotted for vehicles 3, 4 and 5. The obtained curves were as follows:

Vehicle 3:

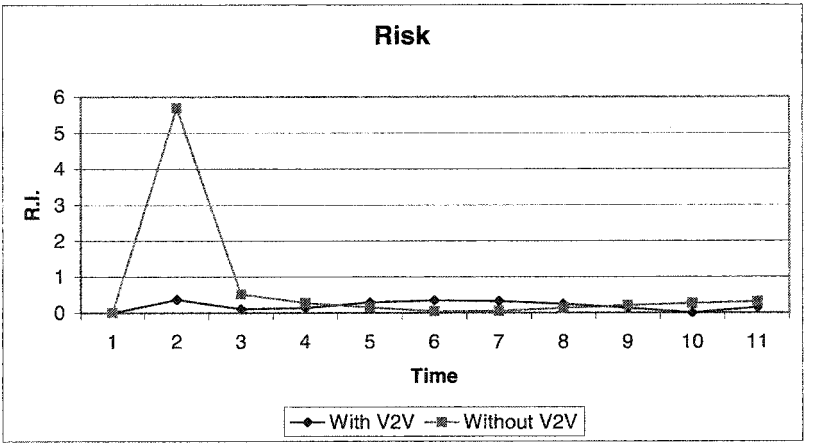


Figure 2: Variation of Risk Index with time for vehicle 3

Vehicle4:

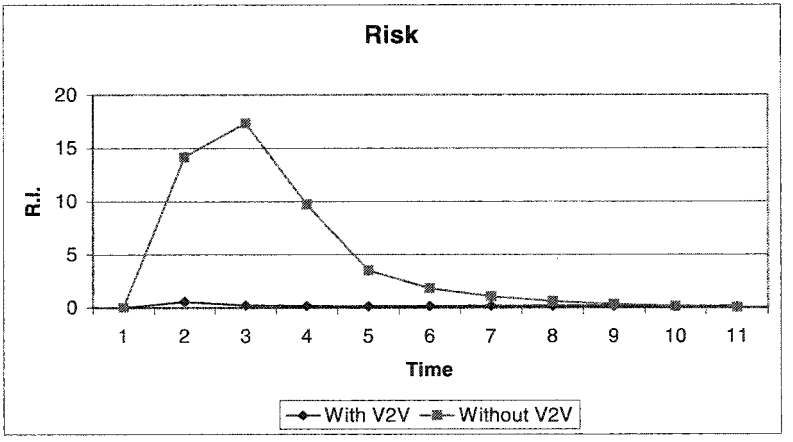


Figure 3: Variation of Risk Index with time for vehicle 4

Vehicle 5:

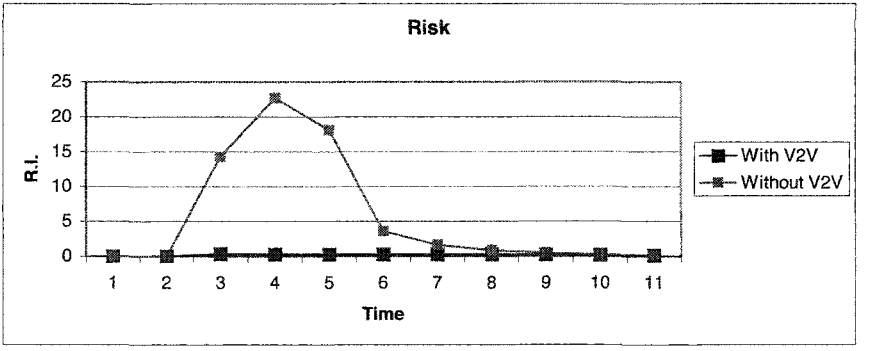


Figure 4: Variation of Risk Index with time for vehicle 3

To study the propagation of risk, the graphs of all the three vehicles when no V2V and V2I communication is used was superimposed.

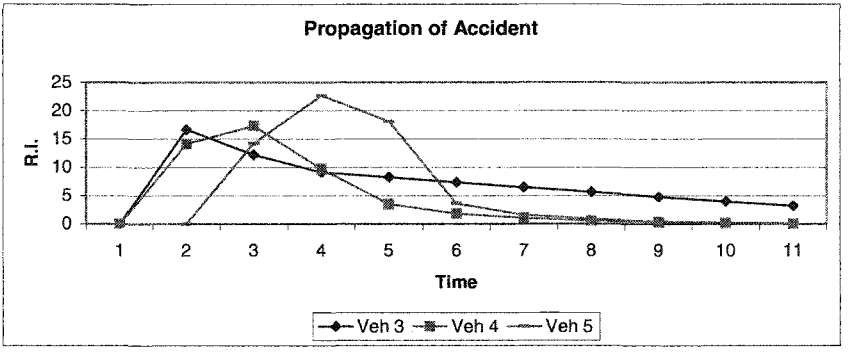


Figure 5: Propagation of peak Risk along the following vehicles

7. Discussion:

It is clear from the results that the Risk Index is less than 1 every time we use V2V and V2I communication which indicates that there is no danger of colliding in the longitudinal direction. It can be seen that percentage reduction in R.I. for the peak times for various vehicles are as shown in the table.

Vehicle	% Reduction
3	93.5
4	98.75
5	98.66

More important than these values is the value of R.I. for different vehicles at different times. It shows that R.I. is always less than 1, which in turn means that there is no danger of longitudinal collision if we use V2V and V2I communication.

It can be seen from the above graphs that the peak danger condition for the vehicles comes according to the position of the vehicle. The nearest vehicle gets its peak danger earliest and vehicle 5 gets to its peak danger last which is quite obvious as the nearest vehicle to the endangered vehicle gets in the danger zone earlier than the vehicles which are at a distance from the endangered vehicle. It can also be seen from the graph that in the initial time after vehicle 1 breaks abruptly, there is a high danger for the following vehicles (3, 4 and 5). As the time passes by, the danger is reduced and hence the two curves meet at higher times.

It can also be seen that if no V2V and V2I communication is used then there is a propagation of the risk/danger. The peak value of R.I. and its time is as shown in the table.

Vehicle	Peak R.I. Value	Time for Peak R.I. (s)
3	5.69	1
4	17.36	2
5	22.68	3

It is highest for vehicle 5 and least for vehicle 3 which is expected because vehicle 5 is the farthest from the abnormal vehicle and hence is caught unaware at the time of sudden braking of vehicle 1 while vehicle 3 which is closer has a lower peak risk index. The occurrence of Peak R.I. indicates the propagation of an accident or Risk in the same lane.

The huge gap between the two curves in the same graph shows how effective V2V and V2I communication could be in reducing the risk of secondary accidents.

8. Conclusion:

This paper proposes to use V2V and V2I communication for reducing the number of secondary accidents which occur in large numbers due to poor weather conditions in which traditional systems like brake lights etc. do not work. A case study of five vehicles traveling in a single lane has been carried out to demonstrate the effect of V2V and V2I communication in reducing secondary accidents by quantifying the Risk using a term known as Risk Index. From the above results and discussions, it can be concluded that V2V and V2I communication systems may be an efficient way to reduce the number of

secondary accidents. The analysis for a case of five vehicles has been done just to demonstrate the use of V2V and V2I communication systems in reducing the number of secondary accidents

However, some limitations should be kept in mind while applying V2V and V2I communication technology for safety purposes. For it to work perfectly to prevent primary accidents, it needs to be updated with the vehicle's sharp dynamic position, which we think is not possible nowadays, keeping in mind the available technologies at present. Also, the R.I. defined above is just a parameter. It indicates a safe condition if its value is less than 1, irrespective of its value. We have assumed that due to the repeated warnings that the V2V and V2I communication systems give, the driver's attention reaches its peak. In this analysis, we have not incorporated any deviation in the driver behavior.

Keeping in mind the above limitations, further studies should be done in order to investigate the use of V2V and V2I communication for reducing the primary accidents as well as for their use in multi-lane roads.

9. References:

[1]. Tokuda, K.; Akiyama, M.; and Fuji, H.; "DOLPHIN for Inter-Vehicle Communications System," Proceedings of the IEEE Intelligent Vehicles Symposium 2000, Dearborn (MI), USA, October 3-5, pp. 504-509, 2000.

[2]. Fujimura, K.; Hasegawa, T.; "A Collaborative MAC Protocol for Inter-Vehicle and Road to Vehicle Communications," Proceedings of IEEE Intelligent Transportation Systems Conference, Washington, D.C., USA, October 3-6, 2004.

[3]. Yang, X.; Liu, J.; Zhao, F.; and Vaidya, H.N.; "A Vehicle-to-Vehicle Communication Protocol for Cooperative Collision Warning," Proceedings of the First Annual International Conference on Mobile and Ubiquitous Systems: Networking and Services (MobiQuitous'04), 2004.

[4]. Wischhof, L.; Ebner, A. and Rohling, H.; " IEEE Transactions On Intelligent Transportation Systems," Vol. 6, No. 1, pp. 90-101, March 2005.

[5]. Tsumura, T.; "Optical Two-way Vehicle-to-Vehicle Communication by the use of Laser and Corner Cube," Journal of Robotics and Autonomous Systems, Vol. 17, pp. 81-86, 1996.

[6]. Kubota, S.; Mizuno, M.; Kaneko, T.; and Nagafusa, M.; "An Effective Data Communication
Method For Mobile Terminal In Spot Communication Systems," IEICE Transactions on Fundamentals of Electronics," Vol. E83-A, No. 7, July 2000.

[7]. Dashtinezhad, S.; Nadeem, T.; Dorohonceanu, B.; Borcea, C.; Kang, P.; and Iftode, L.; "TrafficView: A Driver Assistant Device for Traffic Monitoring based on Car-to-Car

communication," Proceedings of the 59th Semiannual IEEE Vehicular Technology Conference (VTC'S04), 2004.

[8]. Korkmaz, G.; Ekici, E.; Ozguner, F.; Ozguner, U.; "International Conference on Mobile Computing and Networking, Proceedings of the first ACM workshop on Vehicular ad hoc networks," Philadelphia, PA, USA, pp. 76-85, 2004.

[9]. Zhang, J.; Ziliaskopoulos, K.A.; Wen, N.; Berry, A.R.; "Design and Implementation of a Vehicle-to-Vehicle Based Traffic Information System," In Press.

[10]. Fax, J.A.; Murray, M.R.; "Information Flow and Cooperative Control of Vehicle Formations," Proceedings of IEEE Transactions On Automatic Control, Vol. 49, No. 9, pp. 1465-1476, Sep. 2004.

[11]. Blum, J.J.; Eskandarian, A.; and Hoffman, J.L.; "Challenges of Intervehicle *Ad Hoc* Networks," Proceedings of IEEE Transactions On Intelligent Transportation Systems, Vol. 5, No. 4, December 2004.

[12]. Sawant, H.; Tan, J.; Yang, Q.; Wang, Q.; "Using Bluetooth and Sensor Networks for Intelligent Transportation Systems," Proceedings of the 2004 IEEE Intelligent Transportation Systems Conference, Washington, D.C., USA, October 3-6, 2004.

[13]. Veschambre, Y.; Darriet, H.; and Fernier, M.; "DAB Project on Tours-Poitiers Highway, " Proceedings of 7th World Congress on ITS, Turin, 6-9 November, 2000.

[14]. Kroh, R.; Held, A.; Aldinger, M.; Keller, R.; Lohmar, T.; Kovacs, E.; " High Quality Interactive and Broadband IP Services for the Vehicle on the Net –The COMCAR Approach," Proceedings of 7th World Congress on ITS, Turin, 6-9 November, 2000.

[15]. Omae, M.; Fujioka, T.; Miyake, T; "Relative position measurement of neighboring vehicles using DGPS and inter-vehicle communication," Society of Automotive Engineers of Japan, 2000

[16]. CarTALK Website, 25 May 2005, available online at http://www.cartalk2000.net

[17]. Dalla Chiara B., "A study on motorway safety: exogenous variables and perception of on-board telematics relevance", Proceedings of 10th World Congress on Intelligent Transport Systems and services, 16-20 November 2003, n. 4143 - pagg.1-20, Madrid (Spain), Brussels

Intelligent Decision Support System for Traffic Incident Management

Hongqiang Li [1], Huapu Lu[2] and Intikhab Ahmed Qureshi[3]

[1]Postdoctor, Institute of Transportation Engineering, Department of Civil Engineering, Tsinghua University, Beijing 100084, P. R. China; PH (86-10) 62772615; FAX (86-10) 1062795339; email: lhqlch@yahoo.com.cn
[2]Professor, Institute of Transportation Engineering, Department of Civil Engineering, Tsinghua University, Beijing 100084, P. R. China; PH (86-10) 62795339; FAX (86-10) 62795339; email: luhp@mail.tsinghua.edu.cn
[3]Ph.D student, Institute of Transportation Engineering, Department of Civil Engineering, Tsinghua University, Beijing 100084, P. R. China; PH (86-10) 62772615; FAX (86-10) 62795339; email: intikhab_mce@yahoo.com

Abstract

In China, people have been paying more and more attention to incident response since the outbreak of SARS in the spring of 2003. Traffic incident impacts the normal operation of traffic system, even results in paralysis of partial functions of one city. Researches have shown that automated real-time intelligent decision support system can reduce the work intensity of operators in Traffic Operations Center (TOC). This paper introduces the Intelligent Decision Support System (IDSS) for traffic incident operators in Beijing. The prototype system and its main modules are presented. It identifies and verifies incidents that occur on the road network, then regenerate incident response strategies and render advice on what action should be taken. Preliminary applications in fire incident show it is effective to save time and avoid confusion.

Introduction

Traffic incidents are unplanned events that occur randomly in time and space. These random and unpredictable lane-blocking incidents significantly reduce the capacity of the roadway section in which they occur, and resultantly cause traffic congestion. Incident-related delay accounts for 50 to 60 percent of the delay on the roads in most metropolitan areas (FHWA, 2000).

Traffic incident management is defined as the coordinated, pre-planned and/or real-time use of human, institutional, mechanical, and technical resources for reducing the duration and consequently incident-related impacts, and improving the safety of motorists, crash victims, and responders.

Traffic incident management is a complex task for operators in TOC, especially to inexperienced incident operators. The lack of traffic management experience and unfamiliarity with traffic incident response process make it difficult for operators to

use and handle the valuable information effectively, such as traffic data, knowledge, models, and methods to obtain correct solutions when a traffic incident occur. This aggravates traffic operation condition and cause more problems.

Expert System (ES) has been proposed by a number of researchers as a tool to provide vital aid to decision makers. As one of the most major problems in advanced transportation management systems, traffic incident management using real-time knowledge-based expert systems (KBES) was first proposed by Ritchie (1990). To date, several prototypes and tools, FRED, ARTIST CARTESIUS, etc., have been developed to support real-time decision making in traffic incident management (Chang et al., 1993; Logi et al., 2002). The feasibility of assisting the operators in TOC responding to traffic incidents has been demonstrated by field-tested.

As the host city of the 29th Olympics in 2008, Beijing, the political and cultural capital of China as well as a center of international activities, is facing heavy traffic pressure. Until the end of 2004, the vehicle possessing quantity is over 2.3 million, including 1.11 million private cars. Moreover, it is increasing at the speed of 800~1000 vehicles per day. The number of drivers is 3.63 million approximately. Forecast shows the number of vehicles in Beijing will add up to 3.5 million by 2008. Therefore, it is the need of the time to improve traffic condition by using advanced technologies of Intelligent Transportation System (ITS). One is to develop an intelligent decision support system.

Characteristics of Urban Traffic-related Incidents

The emergence of urban traffic-related incidents has the following traits: (1) Suddenness and randomness. The time (when) and place (where) traffic-related incidents will take place is not certained. It is difficult to prevent the occurrence of them. (2) Diversity. There are various types of urban incidents, including traffic accidents, spilled loads, vehicle breakdowns, truck overturns, disabled vehicles, temporary maintenance and construction activities, signals and detectors breakdown, and other unusual events. (3) High frequency. For instance, there are about 3000 cars damaged because of self-ignition in Beijing annually. (4) Connectivity. One kind of incidents generally can bring other kinds of incidents. Generally, the severity of secondary incident is greater than that of the original incidents.

The disposal of urban traffic-related incidents has two features: (1) Limited time. Road traffic is a regular facility. Therefore, urban traffic-incidents must be dealt immediately to mitigate the impacts of traffic flow. (2) Heavy task. In short time, operators have to collect relative information, assemble personnel, resources and execute a proper response plan. The workload of them is very heavy.

Cities play important roles in politics, economy, cultures and society. As an important system in cities, urban traffic system is complicated and huge. It has many associated subsystems, such as road network, transit, bicycles, and logistics. But, it also is a very delicate system (see Figure 1). Any awry of it will impact other parts of

the whole traffic system, even result in the paralysis of some functions of the city. Moreover, the concentration of urban population and association among urban functions deteriorates the impacts.

Figure 1. Delicate urban transportation system.

IDSS for Traffic Incident Management

Intelligent decision support system

Intelligent decision support system is an integrated system of conventional decision support system with Artificial Intelligence (AI) especially with Expert Systems (ES) (Huang, 2001). As its name implies, IDSS is used to provide assistant function to improve system performance since the computer-based IDSS can speed up decision-related analysis and calculations.

Functional modules of IDSS for traffic incident management

IDSS for traffic incident management consists of the three interconnected components which can be mended by users through their maintenance and management systems (see Figure 2).

(1) The database includes all historical and real-time data needed for the operation of the various system functions. The validation of data input by users is examined automatically. The knowledge base (KB) contains encoded knowledge related to traffic incident and traffic management, which is built after knowledge acquisition by knowledge engineers. The main sources used in this study were the experience of domain experts by interviews and literature (articles, papers, books, rule of law, etc.). Knowledge is represented as an appropriate symbolic structure. Rule base contains more than two thousand rules in the form of if/then collected from specialist. Case base includes the past typical solved traffic incident cases for reusing. Rule base and case base also serve as routine analysis and training of new operators. Model base contains models used for traffic incident-related calculations, for instance, regression forecast models of traffic incident duration. While method base contains various

methods for traffic incident analysis, such as probability method, Bayes method, etc.

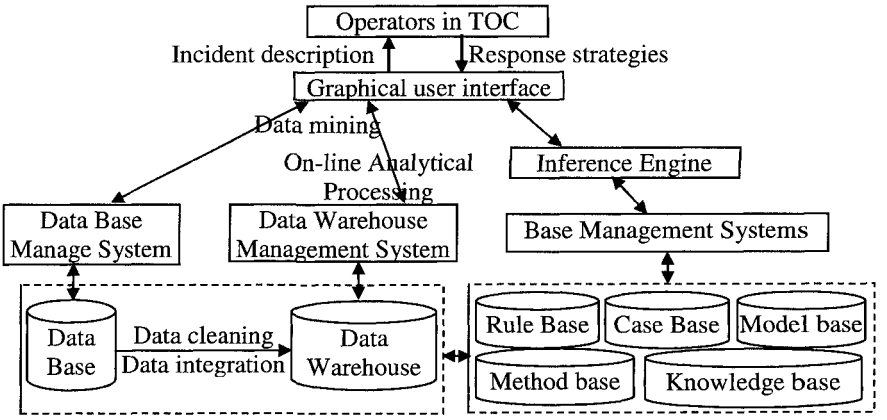


Figure 2. Framework of the IDSS for urban traffic incident.

(2) The inference engine is responsible for the traffic incident decision reasoning. Rule-based reasoning (RBR) matches the rules against the target problems. Rules represent general knowledge of the domain. The advantages of RBR are its naturalness, modularity and ease of explanation. But, the rules are difficult to be acquired. Case-based reasoning (CBR) is the process of solving new problems based on the solutions of past similar problems. Cases are natural and usually easy to obtain. New cases can be inserted into a case base without making changes to the pre-existing cases. The major drawback of CBR is that it does not provide concise representations of the incorporated knowledge. In this system, they are simply combined to improve accuracy of inference based on the collected information. CBR is used to solve regular incidents, whereas RBR is applied for occasional incidents.

(3) The graphical user interface (GUI) supports the functions of data input (knowledge base maintenance and management, modification of model parameters, incident-related information, etc.) and result outputs (inference process, response strategies for traffic incidents, dispatch of resources, etc.). It is user-friendly by using menu, push-button, and drop-list to facilitate personnel in TOC to interactive with the system. And the most known software packages, such as Microsoft Word and Excel, are used to present the decision results. It also can provide system help for personnel in TOC when they come into contact with operation difficultness.

Hardware structure of IDSS for traffic incident management

The general traffic incident decision process includes the following steps: (1) incident detection; (2) incident verification; (3) motorist information; (4) response; (5) site management; (6) traffic management; (7) clearance. Figure 3 shows the hardware configuration, including Closed Circuit TV (CCTV), communication

system, incident command center, signal control system, variable message sign, and rescue vehicles, etc. Figure 4 is a detection and verification interface. The picture on-scene is given by CCTV and displayed on the top-left. The place where the incident occurs is displayed on the Geographical Information System (GIS) map.

Figure 3. Hardware configuration.

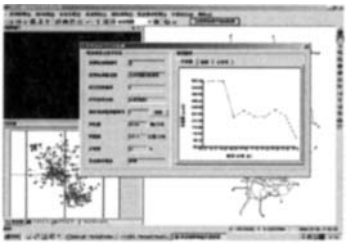


Figure 4. Incident detection/verification.

Validation of IDSS for Traffic incident Management

Incident types and classification

Urban traffic incidents are divided into four types. (1) Public incident, such as leak of poisonous gas, electricity cutoff at a big area, fire, medical treatment and epidemic prevention, parade, and detonation of danger. (2) Natural calamity, such as earthquake, floodwater, mashed vegetable, coast of mountain. (3) Bad weather, such as wind, rain, snow, fog, hailstone, and ice. (4) Incidents in traffic system, such as traffic accidents, traffic congestion, disabled vehicles, spilled cargo, peccancy of vehicles, destruction of road facilities.

Urban traffic incident results from drivers, vehicles, roads, and environment. In Beijing city, traffic incidents are classified into four levels, namely fatal, momentous, biggish, and normal. Four colors (red, orange, yellow, blue) are used to represent the severity of an incident. The severity is divided according to incident type, occurrence time period (rush hour or not), and occurrence place (Central business district or not). A light accident in rush hour or central business district (CBD) also can obtain the highest level alarm.

Testing, verification and validation of the system

This part introduces the application of IDSS for traffic incident management in Beijing. Traffic incident are mainly reported by telephones calls from motorists mainly because of the rapid development of wireless communication in China. Loop-detectors with algorithms detecting traffic abnormalities, learning vector quantization (LVQ), are also used in a few occasions for researches.

The input for traffic incident includes the basic information and expert information

for traffic incident decision. Operators can accomplish the task by rapid choice. Some options directly given by system for user's convenience can be modified, for example incident start time. The main outputs of the system include five kinds of strategies: (1) Motorist information (information content, information media, etc.); (2) Traffic management (traffic flow guidance route, lane closure, etc.); (3) Signal control (timing plan, ramp control, etc.); (4) Resource dispatch (Response agencies and service providers, the appropriate number of personnel and equipment, etc.); (5) Traffic analysis (Incident duration time, Incident impact analysis, etc.). Since the complexity of traffic incident, fire incident (self-ignition of vehicles) decision support was mainly developed. Preliminary applications show the countermeasures from IDSS can decrease about one-third of response time (average 2 minutes). Its benefit may be insignificant for one incident, but considerable for the whole urban traffic system.

Conclusions and Future Researches

In this paper, the features of urban traffic incidents are analyzed. The functional modules and hardware configuration of the system are introduced. Preliminary applications show it is effective for traffic incident management. Future researches will be aimed at more rigorous testing of the system about its reliability and comprehensive evaluation of economic benefits. When it has been validated by traffic incident domain experts from various response agencies and service providers, it may help inexperienced operators to operate the processes more effectively within due time.

Acknowledgments

This research was sponsored in part by Beijing Municipal Science & Technology Commission (H020620530120) and China Postdoctoral Science Foundation (200537367).

References

Chang, E. C. - P., and Huarng, K. (1993). "Freeway incident management expert system design." Transportation Research Record, no. 1399, 36-44.

Federal Highway Administration Office of Travel Management. (2000). "Section 1: Introduction to incident management." *Traffic Incident Management Handbook*. P.B. Farradyne, 1.7-1.10.

Huang, T.Y. (2001). *Intelligent Decision Support System*. Beijing: Publishing House of Electronics Industry. (in Chinese)

Logi, F., and Ritchie, S. G. (2002). "A multi-agent architecture for cooperative inter-jurisdictional traffic congestion management." Transportation Research Part C 10(5-6): 507-527.

Ritchie, S. G. (1990). "A knowledge-based decision support architecture for advanced traffic management." Transportation Research Part A, 24(1): 27-37.

Traffic Flow Data Extracted from Imagery Collected using a Micro Unmanned Aerial Vehicle

B. Coifman, [a,b], M. McCord, [a], R. G. Mishalani, [a], M. Iswalt, [a], and Y. Ji, [a]

[a] Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, Columbus, OH, 43210. USA, http://transportation.eng.ohio-state.edu/

[b] Department of Electrical and Computer Engineering, The Ohio State University, Columbus, OH, 43210, USA

Abstract

Roadway networks span large distances and can be difficult to monitor. Most efforts to collect roadway usage data either require a large fixed infrastructure or are labor intensive. Advances in electronics and communication have recently enabled an alternative, Unmanned Aerial Vehicles (UAVs). UAVs capable of carrying sensors and communications hardware to relay data to the ground are now commercially available. In this paper we investigate the feasibility of using a UAV to monitor roadway traffic and develop several applications. UAVs can cover large areas, travel at high speeds and are not restricted to traveling on the road network. This paper demonstrates several applications using data from a UAV flying in an urban environment: determining level of service (LOS), estimating average annual daily travel (AADT), documenting intersection operations, and measuring origin destination flows. The descriptions of these empirical tasks are intended to provide the main concepts and methods applied to derive useful information for both off-line planning and real-time management, and to illustrate the challenges and opportunities such applications pose.

Introduction

Roadway networks span large distances and can be difficult to monitor. Most efforts to collect roadway usage data either require a large fixed infrastructure or are labor intensive. Conventional traffic surveillance uses a set of detectors (including cameras) deployed at fixed locations, requiring a high density of detectors to monitor changing conditions throughout the network. When information is needed from beyond the range of these fixed detectors, personnel are usually deployed to assess conditions. Unmanned Aerial Vehicles (UAVs) capable of sophisticated autonomous flight, carrying video cameras, geo-positioning sensors and communications hardware to relay data to the ground are now commercially available (e.g., MLB, 2005; GDS, 2005). These UAVs can supplement inflexible, fixed networks of sensors, and labor-intensive, potentially slow deployment of personnel.

In this paper we investigate the feasibility of using a UAV to monitor traffic and develop several applications. As noted in Coifman et al (2004), UAVs can travel at high speeds, covering a large area, unrestricted by the road network to focus resources, reducing the need for fixed sensors and cameras. With the autonomous flight capabilities they can potentially free up personnel from time that would otherwise be spent in transit to remote field locations.

To explore the benefits of UAVs in this context, a set of experiments was conducted on July 22, 2003 using the BAT III technology (MLB, 2005) carrying a payload of two video cameras. The UAV flew over the campus of The Ohio State University in Columbus and the surrounding area at an altitude of 150 m and an air speed of 50 km/h while transmitting video images to a ground station in real-time. The flight lasted almost two hours and data were collected from several facilities. Fig. 1A shows a map of the study area, roughly 2 km east-west and 2.2 km north-south (adapted from OSU, 2003). This urban setting included many intersections and a freeway, SR 315, running north/south through the middle of the map. Many applications were examined, including: level of service (LOS), average annual daily travel (AADT), intersection operations, and origin destination (OD) flows on a small network. Fig. 1B shows a schematic of the primary roadways used in this study.

The subsequent descriptions are intended to provide the main concepts and methods applied to derive useful information for both off-line planning and real-time management applications. The extracted information is presented in some detail to demonstrate the feasibility of extracting useful information from images sampled from an UAV, and to illustrate the challenges and opportunities such applications pose.

LOS Measurement and AADT Estimation

The UAV made one round trip along the southern 1.5 km of SR 315 shown in Fig. 1. The video images from this trip were used to investigate LOS and AADT, respectively reflecting instantaneous and long-term traffic conditions. Density (veh/km) was used to measure LOS, while flow (veh/h) was used to estimate AADT.

At any given instant the UAV only views a small portion of the roadway. As the UAV flies over the traffic it observes conditions over a large time and distance, but the field of view (FOV) continually changes, making difficult the task of measuring conventional density or flow. The generalized definitions over time and space (Edie, 1963) can accommodate this changing FOV. For an arbitrary region, A, of the time-space plane the flow, q, density, k, and space mean speed, $\bar{v}$, are defined as,

$$q(A) = \frac{d(A)}{|A|},\ k(A) = \frac{t(A)}{|A|},\ \bar{v}(A) = \frac{d(A)}{t(A)}, \tag{1}$$

where,

$d(A)$ = the total distance traveled by all vehicles in region A,

$t(A)$ = the total time spent by all vehicles in region A,

$|A|$ = the "area" covered by region A in the time-space plane.

Sampling video from the northbound flight over SR 315 at two second intervals, e.g., Fig. 2A, vehicle trajectories were constructed for southbound traffic. Positions were calculated using a geo-referenced aerial photo and a geographic information system (GIS) to measure distance between observed landmarks. This procedure was repeated in each lane. Similarly the FOV of the southbound lanes was extracted from the video. Fig. 2B shows the resulting trajectories from the inside lane in the time-space plane with solid lines. The FOV for 20 seconds is shown with a shaded region. When a vehicle leaves the FOV before the next observation the trajectory is extended linearly to the FOV boundary and depicted with dashed lines, evident at the ends of most trajectories. The flow and density over the region spanned by the FOV were obtained from the vehicle trajectories by applying Eq. 1 (separately, the research

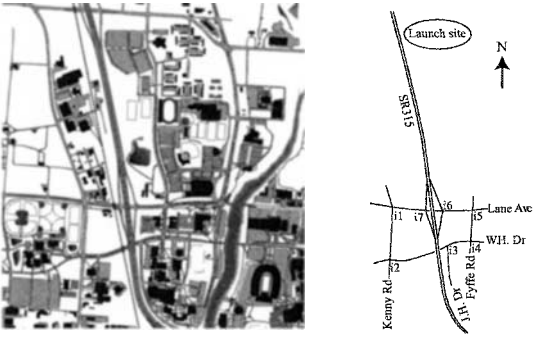


Figure 1, (A) Map of the study area, roughly 2 km across, (B) schematic of the primary network from the map (at the same scale) used in this study.

considered measures of flow and density from still frames, finding similar results).

In fact one does not need the vehicle trajectories, only the times and locations where vehicles enter and leave the FOV since time and distance are summed across all vehicles. It is only important to differentiate between entering and leaving the region, without tracking the vehicles through the FOV, i.e.,

$$d(FOV) = \sum_i \left(x_{exit}(i) - x_{enter}(i)\right) = \sum_i x_{exit}(i) - \sum_i x_{enter}(i), \tag{2A}$$

$$t(FOV) = \sum_i \left(t_{exit}(i) - t_{enter}(i)\right) = \sum_i t_{exit}(i) - \sum_i t_{enter}(i). \tag{2B}$$

The FOV boundaries must be specified accurately to measure $|FOV|$ precisely. Eqs. 2A and 2B can be summed across individual lanes or the entire roadway can be processed without regard to which lane a vehicle is traveling in.

Using the density across all lanes, the LOS can be determined directly from the *Highway Capacity Manual* tables, TRB (2000). In this case the density was 29.8 veh/km/lane, resulting in LOS F. AADT is estimated following the methodology of McCord et al (2003), using $q(A)$ as an estimate of hourly volume:

$$AADT = q(A) \cdot F^{8-9} \cdot 24 \cdot F^{Tu,July}, \tag{3}$$

where,

F^{8-9} = time-of-day factor published by the Ohio Department of Transportation (ODOT) for 8-9am, (the time of day when the video was taken),

$F^{Tu,July}$ = seasonal factor published by ODOT for a Tuesday in July, (the day and month when the video was taken).

Naturally, the estimates from Eqs. 1 and 3 are expected to be somewhat noisy. In this case the flow was measured at 8,399 veh/h, with an estimated AADT of 124,581 veh/day. This estimate is very close to the value of 117,180 veh/day reported by ODOT for the same segment using conventional, terrestrial techniques. As shown in Jiang et al (in press), even such noisy estimates can be valuable when they are combined appropriately with estimates based on older, commonly used terrestrial data. In addition, the impact of this noise can be reduced by integrating information across several independent observations.

Intersection Operating Conditions

The majority of the flight time was devoted to monitoring intersections. First the UAV circled the small network shown in Fig. 1B, counter clockwise past i1-i2-i4-i5, then reversed direction, flying clockwise past i1-i5-i4-i2. Next the UAV circled individual intersections for several minutes, first i1, then i4, i5, and the diamond interchange of i6 and i7. These video

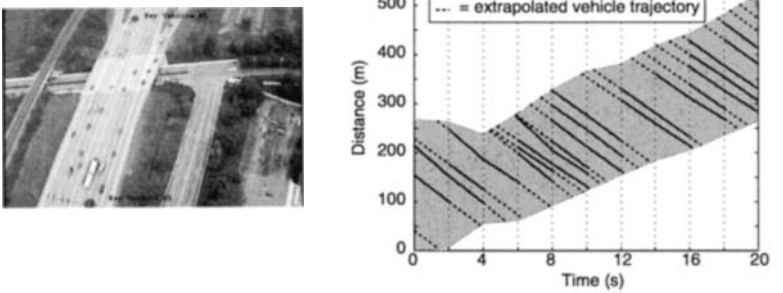


Figure 2, (A) Image from video while flying north along SR 315 (i3 is visible on right), (B) extracted vehicle trajectories in lane 1 and UAV FOV on SR 315-South.

segments were used to analyze intersection operations. Most of the analysis was based on common queuing measures, using input-output flows to measure point queues on intersection approaches. To facilitate measuring the queues, a straightforward computer program was written and used to track the times when vehicles pass two points on each approach. These data allowed for the construction of vehicle arrival and departure curves. Queues, signal timings, arrival rates, and turning movements were all derived from these curves. Spatial queues were also examined, but are not presented in this paper.

Using the computer interface to record vehicle arrivals and departures on each approach at each intersection, the cumulative arrival and the cumulative departure curves were obtained. On a given approach the arrival curve, A(t), was measured at a point far enough upstream of the intersection that queues rarely overran the location but close enough to the intersection that the point was usually within the FOV. In the rare event that either of these assumptions was violated the analysis was suspended until the assumptions were restored. The departure curve, D(t), was measured at the stop bar and vehicle turning movements were recorded through different key presses using the computer interface as the vehicles passed. Following normal queuing theory the arrival curve was shifted forward in time by a constant free flow travel time to yield the virtual arrival curve at the stop bar, V(t). The time shift was estimated empirically on each approach by measuring several vehicles traversing the distance between the two locations under free flow conditions. In this way, V(t) reflects the time vehicles would have passed the stop bar if there were no delay between the two points. The point queue at any instant, then, is simply,

$$Q(t) = V(t) - D(t). \tag{4}$$

Fig. 3 shows two cycles from the point queue model of the eastbound approach to i1. The other approaches were similarly processed. Fig. 3A shows A(t), V(t) and D(t). On this approach the free flow travel time was determined to be 6 seconds. The traffic signal indications were not visible at the resolution of the UAV video. The signal phasing in Fig. 3B was extracted by watching vehicle movements on all four approaches and verified using concurrent video filmed on the ground. There are two periods where D(t) is nearly horizontal; corresponding to the red phase, the small number of departures were recorded from vehicles turning right on red. Fig. 3C shows Q(t) from Eq. 4. Queue growth after the signal turns red and subsequent decay after the signal turns green are clearly evident. The observed peaks correspond to queue lengths just before D(t) exhibits saturation flow at the start of the green phase.

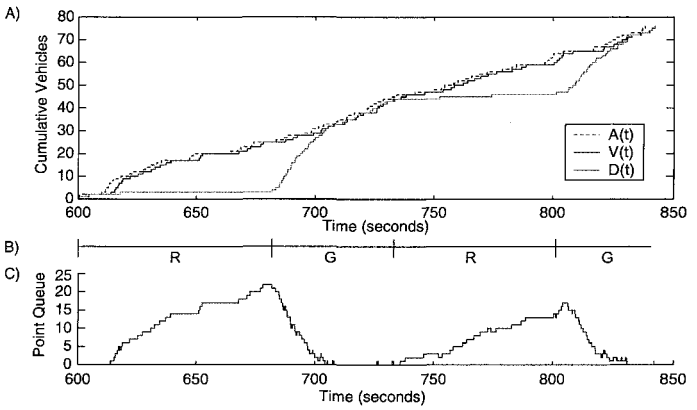


Figure 3, (A) Cumulative arrivals for two cycles from the eastbound approach to i1, (B) the signal phasing on this approach, (C) the measured point queue lengths.

Origin Destination Estimation

The UAV video while circling intersections i6 and i7 were used in conjunction with video taken from ground-based cameras at i1 to determine OD flows on this small network. The network is shown in Fig. 4A superimposed on an aerial photo. The origins were defined to be the approaches to the diamond interchange, i6 and i7, and the destinations were the three branches from i1. Fig. 4B shows a simplified schematic of the network where most of the roadway between i7 and i1 went unobserved. The primary goal of this effort is to match observations across this unobserved region and measure network OD flows. To this end a methodology was developed in which platoons of vehicles from the various origins were matched with those feeding into the downstream intersection. Subsequently, the vehicles are followed through to their respective destinations. The platoon-matching method avoids the need to match individual vehicles, since any unobserved reordering within a platoon from a given origin will not change the proportion of vehicles bound for a given destination.

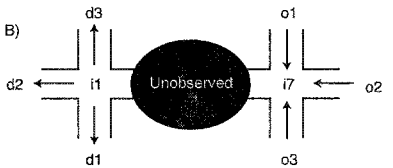
The computer interface developed to study intersection operating conditions was used here to record the origin, departure time, and lane from which vehicles depart the upstream intersection FOV. Similarly, the process was repeated for the vehicles at the downstream intersection, from arrival in the downstream FOV to their final destination. Within each intersection the vehicle movements were followed, yielding a map to the particular origin or destination. A few vehicles departed the network at the freeway on-ramps, which were visible from the UAV view. These vehicles were excluded from further analysis.

The number of vehicles at the upstream and downstream intersections along with their origins and destinations represent the marginal totals of the OD flows, i.e., the number of vehicles that begin at an origin or terminate at a destination. Measuring flows between origins and destinations could in principle be accomplished through exact one-to-one matching of all vehicles between the origins and destinations. However, this task would prove tedious over large samples, and in this case would be very difficult if not impossible given the resolution of the imagery and the non-overlapping fields of view obtained in the upstream UAV video and the downstream ground camera views.

Instead, distinct vehicle platoons were matched between the intersections. Key to the feasibility of this methodology is the fact that the headway between platoons is normally much larger than the headway between vehicles within platoons. Large headways mean overtaking

A)

B)



C)

	D1	D2	D3	Sum O
O1	18	32	1	51
O2	11	42	2	55
O3	5	8	3	16
Sum D	34	82	6	122

Figure 4, (A) OD network superimposed on an aerial photo, (B) schematic of network, most roadway between i7 and i1 went unobserved, (C) complete OD matrix.

between platoons is generally uncommon, and if it does occur, it will only be among the vehicles at the ends of adjacent platoons. Moreover, reordering of vehicles within platoons from overtaking will not impact OD flows, thus allowing the use of the much simpler First-In-First-Out (FIFO) assumption. As the upstream signal cycles through approaches, the clearance intervals and start up times introduce these larger headways between platoons from the various origins. Fig. 4C shows the resulting OD flows from the platoon-matching procedure applied to the empirical data. In the platoon matching method, measurement errors are inevitable during data reduction. Such errors could change the derived OD flows. The impact of such errors are investigated in Mishalani et al (in review), where it is shown that many of these errors can be detected using the observed data.

Conclusions

This paper presented methodological developments to exploit UAV imagery for multiple applications. The applications discussed herein were level of service, annual average daily traffic, intersection operations, and OD flows on a small network. All these applications and more were demonstrated from less than 2 hours of flight time. Most of the data reduction in this study was done manually using computers to simplify many of the tasks. If UAVs are to be used on a large scale for any of these applications, it is likely that additional aids would be developed to assist this process, e.g., software to keep the FOV on the road and hardware such as multiple cameras or specialized lenses to extend the FOV. In the long term, it is likely that many of the tasks could be automated.

Acknowledgements

This study was funded through a grant to the National Consortium for Remote Sensing in Transportation-Flows (NCRST-F) at the Ohio State University from the US Department of Transportation. The efforts of Steve Morris from MLB in providing the aircraft and operator for the field experiment are particularly appreciated, and Keith Redmill of OSU for his logistical support. The authors also acknowledge the help of the many agencies – including the Federal Aviation Administration and the Columbus Police Department – and their staff that enabled the UAV experiment.

References

Coifman, B., McCord, M., Mishalani, R., Redmill, K., (2004). "Surface Transportation Surveillance from Unmanned Aerial Vehicles" *Proc. of the 83rd Annual Meeting of the Transportation Research Board.*

Edie, L., (1963) "Discussion of Traffic Stream Measurements and Definitions," *Proc. 2nd International Symposium on the Theory of Traffic Flow*, pp 139-154.

GDS, (2005). "GeoDataSystems", http://www.geodatasystems.com, accessed on December 8, 2005.

Jiang, Z., McCord, M., Goel, P., (in press) Improved AADT Estimation by Combining Information in Image- and Ground-based Traffic Data, *ASCE Journal of Transportation Engineering.*

McCord, M., Yang, Y., Jiang, Z., Coifman, B., Goel, P., (2003) Estimating AADT from Satellite Imagery and Air Photos: Empirical Results, *Transportation Research Record 1855*, pp 136-142.

Mishalani, R., McCord, M., Coifman, B., Iswalt, M., Ji., (in review) "Platoon Based Origin Destination Estimation" submitted for publication in the *Proc of AATT 2006.*

MLB, (2005). "MLB Company," http://spyplanes.com, accessed on December 8, 2005.

OSU (2003). "Ohio State University Department of physical facilities," http://www.physfac.ohio-state.edu/mapping, accessed on July 23, 2003.

TRB, (2000). *Highway Capacity Manual*, Transportation Research Board.

Analyzing Freeway Service Patrol Data

Chengjun Zhan[1], Albert Gan[2], and Mohammed Hadi[3]

[1]Lehman Center for Transportation Research, Florida International University; PH (305) 348-4103; FAX (305) 348-2802; email: czhan003@fiu.edu
[2]Lehman Center for Transportation Research, Florida International University; PH (305) 348-3116; FAX (305) 348-2802; email: gana@fiu.edu
[3]Lehman Center for Transportation Research, Florida International University; PH (305) 348-0092; FAX (305) 348-2802; email: hadim@fiu.edu

Abstract

The Florida Department of Transportation (FDOT) has implemented freeway service patrol programs at major urban areas to ease traffic congestion due to non-recurring incidents. This paper describes an effort to use data mining techniques to extract information from the service patrol databases for Broward and Palm Beach counties, Florida. A pattern-based geocoding process was developed to geocode the databases. Geocoding was needed because the incident records do not contain specific geographic coordinates, but only a general description of the incident locations. An intelligent parser program was developed to scan and parse location information patterns to identify specific geographic locations. The program was able to achieve a high geocoding rate of over 95% for the service databases. A spatial analysis example making use of the geocoded data was demonstrated.

Introduction

As part of incident management strategies, service patrol programs have been in operations in the United States for over three decades (TxDOT, 1997; Latoski *et al.*, 1999). These programs utilize roving vehicles to patrol congested and/or high incident freeway sections to quickly detect and clear incidents and restore freeway capacities. In 2004, there were a total of 54 such programs in 23 states across the U.S. Florida's service patrol programs, better known as the Road Rangers programs, include over 100 patrol vehicles servicing over 1,400 centerline miles. In Broward and Palm Beach counties, FDOT District 4 deploys 22 patrol vehicles covering 111.3 centerline miles for segments of I-95, I-75 and I-595. Incident information from these programs is recorded in "service" databases. The current service databases include only descriptive location information that does not lend them to spatial analyses.

This paper describes an effort to apply data mining techniques (Chen *et al* 1996; Fayyad *et al* 1996) to mine service patrol databases for incident analyses. An efficient geocoding process that applies pattern recognition techniques was first developed to geocode the service databases. Geocoding is needed to provide geographic location information that is needed for spatial analyses such as adjusting the service area definitions and re-allocating service patrol vehicles. Following the geocoding process, as an application example, incident frequency and incident

location analyses were performed using the geocoded records. Geocoded incident records have been used in various applications, including evaluating the operation of service patrol programs, predicting the durations of future incidents (Ozbay and Kachroo 1999), and optimizing the service patrol dispatch operations.

Data Acquisition

The incident data used in this study were downloaded from the SMART SunGuide website of the FDOT District 4 (SMART 2005). Service patrol data for a total of 609 days from January 1st 2003 to August 31st 2004 were extracted from the Broward and Palm Beach service databases. During the period, the Palm Beach program completed 35,531 assists, or about 58.4 assists per day, while the Broward County program completed 59,277 assists, or about 97.33 assists per day.

In Broward County, the service area covers 57.73 centerline miles, which includes I-95 (25.29 centerline miles), I-75 (19.58 centerline miles), and I-595 (12.86 centerline miles). In Palm Beach County, the service area covers the entire I-95 corridor beginning at the Broward County line extending to Martin County line (53.57 centerline miles). Both service patrol programs operate 24 hours a day.

Traffic incidents are categorized into 12 different incident types, including "FHP", "Accident", "Flat Tire", "Fuel", "Cell Phone", "Jump Start", "Debris Removal", "Minor Repairs", "Overheated", "No Service", "Abandoned Vehicle", and "Tow to Shoulder". All the service patrol data are stored in the same table format with identical attribute definitions.

Geocoding with Pattern Recognition Technique

The service databases contain only a descriptive attribute ("Location") that describes where an incident occurred without the specific geographic coordinate for the location. To support spatial analysis of the service patrol data, geocoding is needed. Geocoding of the descriptive location information is a nontrivial task that is complicated by the following factors: 1) it is a text field with no constraint on the possible values. Hence, any kind of value may occur; 2) the data table has no attribute for the geographic coordinates of incidents; 3) different traffic operators have different abbreviation styles (patterns) for the same interchange/exit name; 4) there are no strict requirements for operators to enter detailed information.

Pattern recognition techniques have been studied extensively and have proved to be efficient in areas such as text and character processing, image processing, etc (Govindan and Shivaprasad, 1990). Based on pattern-based mining techniques, the following geocoding procedure was developed and implemented:

Step 1: Extract northbound/southbound approach information from the abbreviations such as "NB", "N/B", etc. This step is straightforward since the parser only needs to make exact pattern matches for the possible abbreviations.

Step 2: Extract directional information of an incident such as south/north of an interchange from abbreviations such as "S.", "S/", etc. This step is more complicated since it is very common for interchanges to have names with keywords "South", "North", etc. For example, one interchange is called "South 6th Ave", which may lead to wrong matches. The parser program handles this issue by first matching possible patterns for interchange names and then re-scanning the remaining location information for directional information.

Step 3: Scan the location information to extract reoccurrence patterns related to interchange and milepost information. Although different operators may use different abbreviation styles, they tend to stick to the same abbreviation styles themselves. A mapping table was manually created for each county to map the milepost, roadway section, exit number, interchange name, and possible interchange abbreviations. Table 1 displays some sample records from the mapping tables.

Step 4: This complex step consists of several scan and pattern matching sub-steps. First, the original records are scanned to retrieve abbreviations with numbers such as "MM 91" (MM means Mile Marker) to get a match for milepost information. If no match is found, the mapping table created earlier is used to find matches for interchange names. Since it is not practical to create mapping entries for each of the hundreds of possible abbreviations, a partial pattern matching technique is applied to identify mappings. For example, descriptive location information such as "NB N Palm Beach", "N N/B Palm B Blvd.", or "NB N Palm Bch L" should all be parsed to match "Northbound North of Palm Beach Lakes Blvd., Exit No. 71, Milepost 70.978". Some issues could arise in the mapping process. For example, as shown in Table 1, both "Palmetto Park Road" and "Palm Beach Lakes Blvd" include the possible pattern "Palm", which may cause ambiguity and result in wrong matches. The technique developed in this study carefully considers possible overlapping of patterns to resolve such problems.

Table 1. Samples of the Mapping Table for Palm Beach County.

Exit	Interchange Name	Section	Milepost	Possible Pattern
41	SW 10th St.	86070000	40.894	SW 10th ST
44	Palmetto Park Road	93220000	44.085	Palme
71	Palm Beach Lakes Blvd.	93220000	70.978	PBL

Manual checking on a small testing dataset was used to validate the correctness of the parser program. By using the parser program, a total of 34,800 out of 35,531 records for the Palm Beach County were successfully geocoded, giving it a high 97.94% success rate. For Broward County, a total of 56,647 out of 59,247 records were successfully geocoded, resulting in a 95.57% success rate, which is also considered high.

Spatial Incident Analysis – An Example

Geocoded service data could be useful for various analyses, such as secondary crash rate analysis, spatial analysis using GIS tools, etc. Incident analysis based on spatial

information can be used to identify locations with high incident rates for the patrolling freeway corridors. This may lead to rescheduling of service patrolling vehicles or other measures to improve service. This section describes one example of such applications.

Owing to the successful geocoding of the service databases, incident frequency distribution based on interchange information can be performed and the result is shown in Figure 1. As displayed in Figure 1, there are far more interchanges with high incident frequency in Broward County than in Palm Beach County. The possible reasons for this difference include a much higher AADT for the freeway sections in Broward County and more patrolling vehicles in Broward County. To focus on the evaluation of the service patrol program performance for the freeway segments with high incident frequency, two service areas with high incident frequency, as shown in Figure 1, were identified for each county:

- Area 1 (Broward County): an 11-mile segment along I-95 between Griffin Road and Atlantic Blvd. It is also the segment with the highest average incident occurrence frequency among the service areas for the study period.
- Area 2 (Broward County): a 6-mile segment on I-75 stretching from Griffin Road to Sawgrass Expressway.
- Area 3 (Palm Beach County): an I-95 segment from Hillsborough Blvd. to Palmetto Park Road with a length of about three miles.
- Area 4 (Palm Beach County): a 4-mile I-95 segment goes from Atlantic Ave. to Woolbright Road.

Incident analysis on service areas with the highest incident frequency is useful since such areas usually have higher traffic delays than other service areas. A preliminary study found that for all the service areas, the five most frequently occurring incident categories were "Flat Tire", "Accident", "Fuel", "No Service", and "Abandoned Vehicle". Table 2 shows the incident frequency as well as the mean, 90^{th}, and 95^{th} percentile incident duration values for these five major incident categories. The following observations could be made from Table 2:

- Incident belongings to the five major incident categories accounted for more than 70% of the total number of incidents for each study area.
- "Flat Tire" incidents have the highest occurrence frequency for all the areas.
- The mean durations for "Abandoned Vehicle" and "No Service" incident categories are much smaller than other categories, which is expected since the two categories require less labor work and can generally be cleared quickly.
- Data for study area 2 on I-75 shows that the "Accident" incident category has a low frequency but a very high mean duration. This indicates that the current service quality for the area is poor.
- The mean incident durations for area 4, especially for the "Accident" category, is generally lower than the average values for other areas in the table.

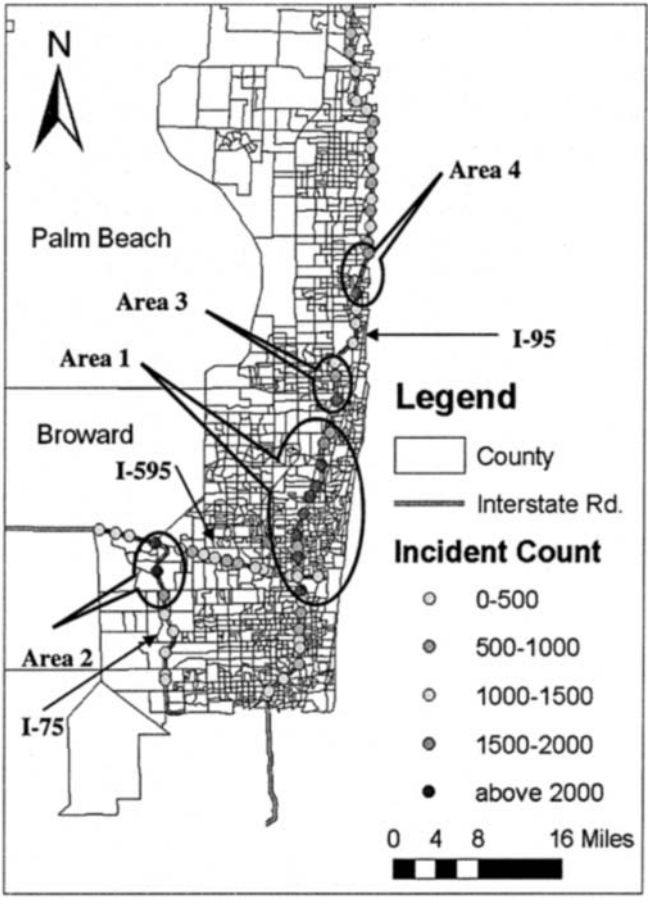


Figure 1. Incident Counts for Broward and Palm Beach Counties.

Conclusions

This paper presented a data mining technique to geocode the service patrol databases from two major counties in Florida. Geocoding was needed because of the lack of spatial definition in the service patrol data. A pattern-based method was developed to scan and parse discovered location patterns. The results show that the method was able to geocode over 95% of the incident records. One application example was given to show the use of the geocoded databases to assess the service quality at different service areas and to identify and target areas for service improvements. Additional applications of such databases may include prediction of incident durations based on local traffic and geometric conditions and optimization of service dispatch operations.

Table 2. Analysis of Major Incident Categories for High Incident Areas.

Area No.	Total Count	Incident Category	Number of Records	Percent in Total	Mean (min)	90th Percent (min)	95th Percent (min)
1	21385	Flat Tire	6070	28.38%	17.09	26	35
		Accident	2436	11.39%	39.92	80	101
		Fuel	2490	11.64%	11.38	18	25
		Abandoned Vehicle	2727	12.75%	6.19	10	15
		No Service	2049	9.58%	6.96	12	17
2	5948	Flat Tire	1799	30.25%	16.34	25	32
		Accident	372	6.25%	53.54	122	158
		Fuel	825	13.87%	11.62	16	25
		Abandoned Vehicle	648	10.89%	5.37	7	10
		No Service	628	10.56%	6.38	12	15
3	3991	Flat Tire	1142	28.61%	17.50	26	35
		Accident	452	11.33%	40.05	75	102
		Fuel	407	10.20%	10.09	20	25
		Abandoned Vehicle	541	13.56%	6.49	8	10
		No Service	423	10.60%	5.78	10	15
4	3657	Flat Tire	933	25.51%	14.98	25	30
		Accident	432	11.81%	31.90	62	80
		Fuel	440	12.03%	8.29	15	20
		Abandoned Vehicle	443	12.11%	5.63	7	11
		No Service	414	11.32%	5.79	10	15

References

Chen, M. S., Han, J. W., and Yu, P. S. (1996). "Data Mining: An Overview from a Database Perspective." *IEEE Trans. on Know. and Data Engr.*, 8(6), 866-883.

Fayyad, U. M., Shapiro, G. P., Smyth, P., and Uthurusamy, R. (1996). *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, Boston, MA.

Garib, A., Radwan, A. E., and Al-Deek, H. (1997). "Estimating Magnitude and Duration of Incident Delays." *Journal of Transportation Engineering*, 123(6), 459-466.

Govindan, V. K., and Shivaprasad, A. P. (1990). "Character Recognition – A Review." *Journal of Pattern Recognition*, 23(7), 671-683.

Latoski, S. P., Pal, R., and Sinha, K. C. (1999). "Cost-Effectiveness of Hoosier Helper Freeway Service Patrol." *ITE Journal*, 125(5), 429-438.

Ozbay, K., and Kachroo, P. (1999). *Incident Management in Intelligent Transportation Systems*. Artech House, Boston, MA.

SMART SunGuide Website (2005). Florida Department of Transportation District 4. <http://www.smartsunguide.com > (Aug. 10th, 2005).

Texas Department of Transportation (1997). *Texas Courtesy Patrol*. Austin, Texas.

Using Data Mining to Analyze Archived Traffic Related Data

Vanessa Amado, MSCE[1] and Mark R. Virkler, PhD, P.E.[2]

[1]PhD Candidate, University of Missouri, E2509 Engineering Building East, Columbia, MO 65211; PH (787) 587-9928; email: vae59@mizzou.edu
[2]Chair and Professor, University of Missouri, E2509 Laferre Hall Engineering Building East, Columbia, MO 65211; PH (573) 882-7434; email: virklerm@missouri.edu

Abstract

Little is known to date about the value of applying data mining, especially association mining, to transportation-generated data. The rapid development of intelligent transportation systems (ITS) has generated large amounts of data for transportation professionals. Some of these data are being used in data mining investigations following on the success researchers on areas like business and computer sciences have had, yet there is still much to be learned.

The methodology used in this study consisted of using association methods to obtain new information from pairs of data such as time and level of service. In this paper, how traditional views of traffic data can be complemented with data mining techniques is explored and described, thereby demonstrating the opportunities that data mining could offer in better understanding the situation and, furthermore, help yield better decisions.

The results present graphical illustrations of the associations found between the different pairs of data and the corresponding discussions.

Introduction

While ITS has and will keep providing benefits for motorists as they take advantage of the better service provided, some of the gains may be offset by the expected growth in highway travel by 2025 (Bureau of Transportation Statistics (BTS) 2000). Trends of embedding new technology into operations and management of the transportation systems will keep accelerating as new and more advanced technology is developed, however congestion will remain an issue. Thus, not only should new technology be developed to address these problems, but ways to interpret the data that are archived by means of the many ITS interlinked systems should also be created. It would be a huge task to try to understand all of a system's data at once, given that the amount of data archived in some traffic management centers (TMCs) is immense. ITS generated data is not archived for every ITS system, even in cities like Los Angeles where the ITS infrastructure is extremely advanced, however large banks of data are rapidly emerging as this technology keeps growing.

The main objective of the research was to obtain more information out of the archived traffic generated data from an expressway facility and to learn how to make better use of such data by means of data mining tools.

Background Information

Data Mining

Even though the application of data mining algorithms in the transportation field is still new, the amount of transportation related data being collected and archived nowadays provide the basis for such algorithms to be used as a means to gain additional information that could improve the decision making process for the professionals in this field.

Data mining algorithms provide the tools needed for one to increase the amount of information that a database holds. As the collection and archival of data increases and computers get faster and cheaper, the interest of researchers and data analysts of discovering new patterns in the data increase as well. While this research only deals with the association function applied to traffic generated data from an expressway in the San Juan Metropolitan Region (SJMR) in Puerto Rico, note there are many more applications available within the civil engineering domain. For example, a study conducted using clustering and regression to identify outliers in weigh-in-motion data from Mn/ROAD data (Buchheit et al. 2002), the application of knowledge discovery in databases and data mining to large amounts of construction project data to identify the novel patterns in construction fields (Soibelman and Kim 2000), and the application of classification analysis to large sets of pavement condition data to predict the present serviceability rating (PSR) of pavements from the state of Missouri (Amado 2001). Many more examples of data mining applications within the field of civil engineering can be found by conducting a search in most of the engineering libraries of higher education institutions in the US.

Case Study

Las Américas Expressway (PR-18) was built over 30 years ago to serve two functions: to provide a corridor between the Old San Juan and the suburbs and to serve as an expressway connector between the SJMR and PR-52. PR-52 is one of the most important expressways in Puerto Rico given that it provides access to the south part of the island. PR-18 also connects PR-22, which is another important expressway, being the only designated Interstate highway in Puerto Rico. PR-22 provides access to and from the SJMR to the west side of the island (Figure 1).

The Department of Transportation and Public Works of Puerto Rico (DTOPPR for its Spanish acronym) has conducted a series of improvements that have increased the capacity of PR-18. Given the rapid residential and commercial development in the SJMR, there is no room for expansion of this highway. Additional improvements are being made to the transportation system of SJMR, such as the "Tren Urbano," which is a huge project consisting of a combination of heavy and light rail system that will connect the five major sectors (Bayamón, Guaynabo, Río Piedras, Carolina, Old San Juan, and Hato Rey) within the SJMR (Pesquera and González 1996). With the construction and effective operation of "Tren Urbano," the congestion in SJMR is expected to be reduced (Pesquera and González 1996). However it is well known

that more people are expected to be driving vehicles in the years to come, thus new measures are needed to describe and improve the quality of flow.

Methodology

Association Mining

Association mining is mostly known for its consumer-oriented intended for grocery stores, phone companies, credit card companies, and banking applications. However, it is aimed at finding relationships between different attributes, often in large databases (Evans 2003). The most common example of association mining would be whether customer who buys item A is also likely to buy item B. The idea is to figure out how to determine the presence of some set of items given the presence of other items in a transaction.

Knowing the basic theory of association mining permits the application of the concept to traffic generated data. If instead of transactions one considers the presence of events, such as the events that occur in a highway at certain time of the day given the presence of other events, one might be able to find the relationship between events. For example, what accident types are related to certain average speeds? what type(s) of accidents are occurring more often? And what are the average vehicle speeds for the most common accident types? Questions of this nature are important to highway engineers to broaden the knowledge of driver's behavior under specific external conditions, whether these are weather related or caused by human factor. These are events that affect traffic flow conditions affecting, as a consequence, the capacity.

The data was cleaned and examined using spreadsheets; however, the data mining applications from the IBM® DB2 Intelligent Miner for Data (IBM 1996) were used as the main force tool for the analysis.

Association mining brings the unique ability to find the combinations of events that are not evident by simply scanning the data. Furthermore, it offers a different means of studying traffic-generated data.

Archived Data

Historical data for this research were obtained from various divisions within the DTOPPR and the National Climatic Data Center (NCDC). In Puerto Rico there is no Traffic Management Center (TMC), thus traffic-generated data is collected, archived, and analyzed by the Office of Data Acquisition and Analysis of Transit within DTOPPR. The data collected by the latter division is used in the Highway Performance Monitoring Systems (HPMS) study as required by the Federal Highway Administration (FHWA) and Transportation of Metropolitan Areas study, which is an internal study used in DTOPPR (Burgos 2005). The data and information are also used for the design of projects coordinated by the Office of Environmental Studies and Pavement Management (Pérez 2005).

Data were collected at point locations using two types of automatic traffic counters, single and double inductance loops, and accumulated in roadside controllers. These field measurements were collected for each direction in PR-18. Quality control checks on data were performed once the data was taken to DTOPPR offices.

Weather data are collected each hour in the San Juan Luis Muñoz Marín International Airport (SJU) station using Automated Surface Observing Systems (ASOS) equipment for measurements (McCown 2005). Hourly precipitation totals for the entire month for the last 50 years is available with free access for ".edu" domain in either text version or ASCII comma-delimited files within the National Climatic Data Center website (NCDC 2005).

Lessons Learned

When dealing with traffic generated data, especially archived data from different sources, one will always be confronted with the fact that data will be missing and data will have different formats. State and/or local agencies that have more advanced transportation management systems (TMS) probably have the most organized databases, yet that is not likely to be the case for most state nor local transportation agencies. The fact that people who collect the data are not the ones using the data for research, presents the biggest problem because the sense of purpose for such activity is mostly missing.

Up to this point, whether one is to use the gathered data to perform conventional methods for the analysis of traffic generated data or use the data to perform, say, data mining applications, there is no difference in the challenges one could confront when trying to create a working database of traffic generated data from different sources. However, once the working database is fully developed and cleaned, the advantages of using either method (conventional or data mining) will depend primarily on the objectives of the project. The major benefit learned from applying data mining to traffic generated data is that it allows the analysis of multiple variables, and, if all done right, the analysis is completed in a matter of minutes. A list of general lessons learned are as follows:

- A single case study can contribute to local and national knowledge. In an ideal world, every state or local agency would have complete sets of historical data for each highway within their highway network. Everyone knows how unrealistic that is due in part to the high costs of data collection procedures and data management, yet probably every state agency has at least one highway in which historical data has been collected for some period of time within the last five (5) years. The initial exercise of applying data mining will yield the needs of data when it is needed but not available.
- The more use of the data collected will improve the quality of performance. Some state agencies are reluctant to collect data of vital highways within their network because, aside from the operating costs involved, they do not know what to do with the data other than storing it. However, during the mining of data one learns really quickly what other type of data are needed and in what format should the data be collected and stored. The value of "good" data is learned as soon as the amount of missing data surpasses the amount of complete or clean data. There will always be missing data and erroneous data given that machines malfunction every now and then and/or people misread numbers or values, but the more use of data local and state agencies carry out the more complete sets of data they will eventually have in their databases.

- Engineers are likely to understand data in its rawest form, yet the same thing will not be likely for the average person. Thus, the presentation of data has to be attractive and understandable to the average person in order to receive acceptance. One of the benefits of using data mining is the amount of visuals in which data can be displayed (for example: bar charts, graphs, and tables).
- Data mining allows one to work with the data in its rawest form and presented in its most fancy form. In fact it is during the mining process itself when information is gathered and the new knowledge is reached when the process is completed. The multiple information like statistics, mining tools, and patterns found in tables of data and graphical views provide a great deal of benefit for the data-mining user. The more information is obtained about a certain highway the more understanding one has about the operational performance of that facility.

More specific lessons were learned after conducting the association mining on traffic data from weekdays. Models were created for time and level of service for every weekday and compared to traditional views of traffic analysis (density vs time, flow vs time, speed vs time, and speed vs flow). Nonetheless for the purpose of this paper and due to length limitations, only the models for Monday will be presented and discussed. The findings were as follows:

- More associations were found for the northbound direction (towards the SJMR), indicating that far more traffic is moving in that direction. From looking at the traditional graphs on the left side of Figures 2 and 3 one cannot clearly tell in which direction there is more traffic moving. Association mining allowed the fast identification of the majority traffic movement.
- The levels of service (LOS) with the most support, with more occurrences, could be identified easily from the associations graph as opposed to from the speed versus flow graph (top right corner on Figures 2 and 3). The LOS with most support for the San Juan direction (northbound) were LOS C and D and for the Caguas direction (southbound) were LOS B and C.
- From the speed versus flow graphs (upper right corner on Figures 2 and 3) it appears as though much data points lay within LOS A, B, C, and some on D, however the associations graph presents the LOS that have the most associations between one another. These are LOS B, C, and D on both directions. Even though it appears like there are many instances in which the traffic operations on both directions have a LOS A, it quickly falls into lower LOS and operation conditions remain between LOS B, C, and D for most of the day.
- It was also learned that the closer traffic flow data lay within the graph (second graph on left side of Figures 2 and 3) the fewer associations the mining tool will find. That is, when traffic is traveling at similar speeds not much change will occur between levels of services and, thus, the mining tool will find fewer associations. This can be seen on Figure 3.

Conclusions

The application of association mining to complement the traditional view of traffic analysis was presented in this paper. From the study, it appears that association mining does provide the opportunity to better understand the operational conditions of traffic on any given weekday. Using associations in combination with

the traditional means of analyzing traffic data allows the user to understand the relationship between two variables, in this case Time and LOS in a different graphical form. The information represented via the data mining visualization aid can assist traffic engineers to understand the real relationship between two variables as it illustrates the support of occurrences. Likewise, the visualization tool allows for the user to identify any hidden relationships between two variables.

General or specific findings, a contribution is provided due first, by the application of a methodology that until recently was only applied to consumer oriented databases, and second, by the fact that data from an agency which is still developing its own TMC was used. The latter can be a reflect on state or local agencies in the US that are currently developing their first TMC.

Acknowledgements

This research was funded by the Department of Civil & Environmental Engineering at the University of Missouri-Columbia, the Midwest Transportation Consortium (MTC), and the MAGEP fellowship. The authors wish to thank the PRHTA and the IBM® Corporation for their collaboration.

References

Amado, V. (2001). Expanding the use of pavement management data. Department of Civil & Environmental Engineering. Columbia, University of Missouri: 83.

Buchheit, R. B., Garrett, J.H.Jr., McNeil, S., and Chalkline, M.H. (2002). Automated procedures for improving the accuracy of sensor-based monitoring data. Applications Of Advanced Technology in Transportation Proceedings of the 7th International Conference, Cambridge, Massachusetts, American Society of Civil Engineers.

Bureau of Transportation Statistics (BTS) (2000). The changing face of transportation. Washington, D.C., Bureau of Transportation Statistics and U.S. Department of Transportation.

Burgos, J. (2005). Department of Transportation and Public Works of Puerto Rico (DTOPPR). April 2005, electronic communication.

Evans, R. (2003). Mining your warranty data using ibm db2 intelligent miner for data association methos. www-106.ibm.com/developerworks/db2/library/technicalarticle/dm-0311evans/ (April 13, 2005), IBM Corp. 2005: 40.

IBM Corporation (1996). Using the intelligent miner for data, version 6 release 1, SH12-6394-00.

McCown, S. (2005). National Climatic Data Center (NCDC). February 2005, electronic communication.

NCDC (2005). Local climatological data, san juan luis munoz marin international airport (SJU) station. http:www.ncdc.noaa.gov/oa/ncdc.html (February 1, 2005), National Climatic Data Center. 2005.

Pérez, I. A. (2004). Department of Transportation and Public Works of Puerto Rico (DTOPPR). February 2004, personal communication.

Pesquera, C. I., and Gonzalez, S.L. (1996). "Tren urbano on track to relieve congestion." Railway Gazette International 152(1).

Soibelman, L., and Kim, H. (2000). Generating construction knowledge with knowledge discovery in databases. Proceedings of the 8th International Conference on: Computing in Civil and Building Eingineering, Stanford, California.

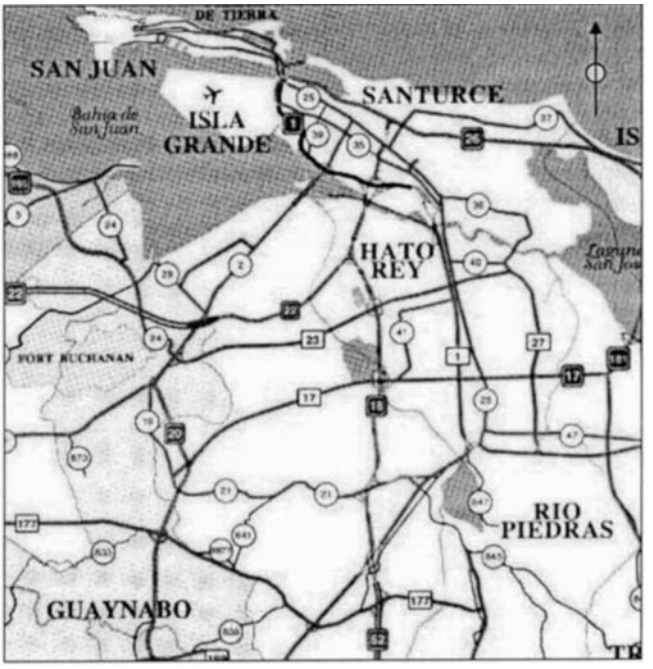


Figure 1. Location of PR-18 in the san juan metropolitan region. (Source: Guía Urbana del Area Metropolitana, 2000 San Juan Metropolitan Area Metro Data. CD-ROM)

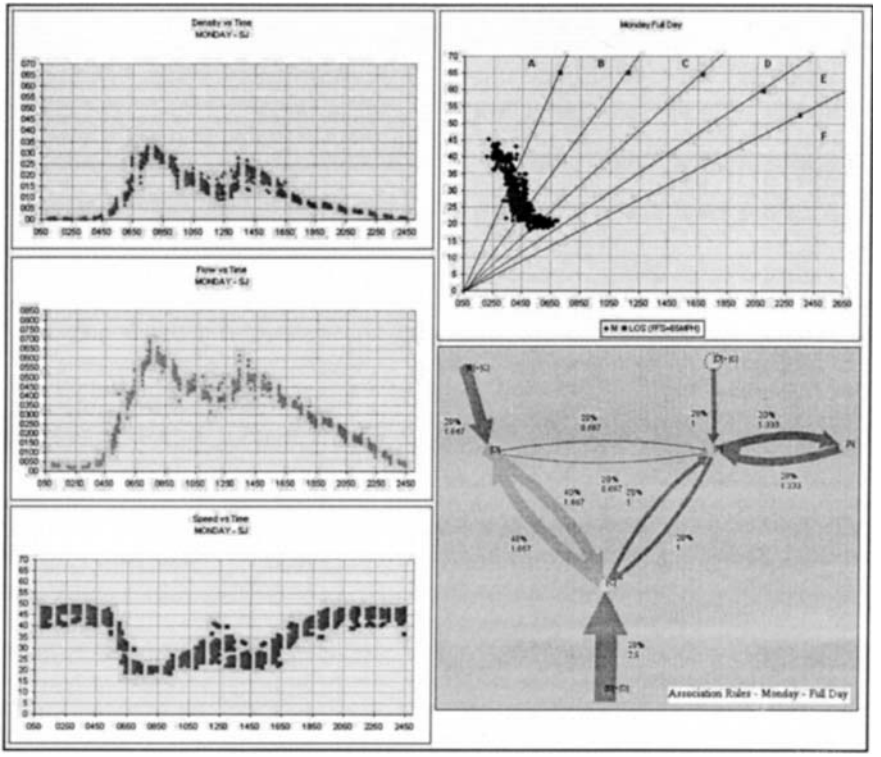


Figure 2. Graphs for monday model – san juan direction (northbound)

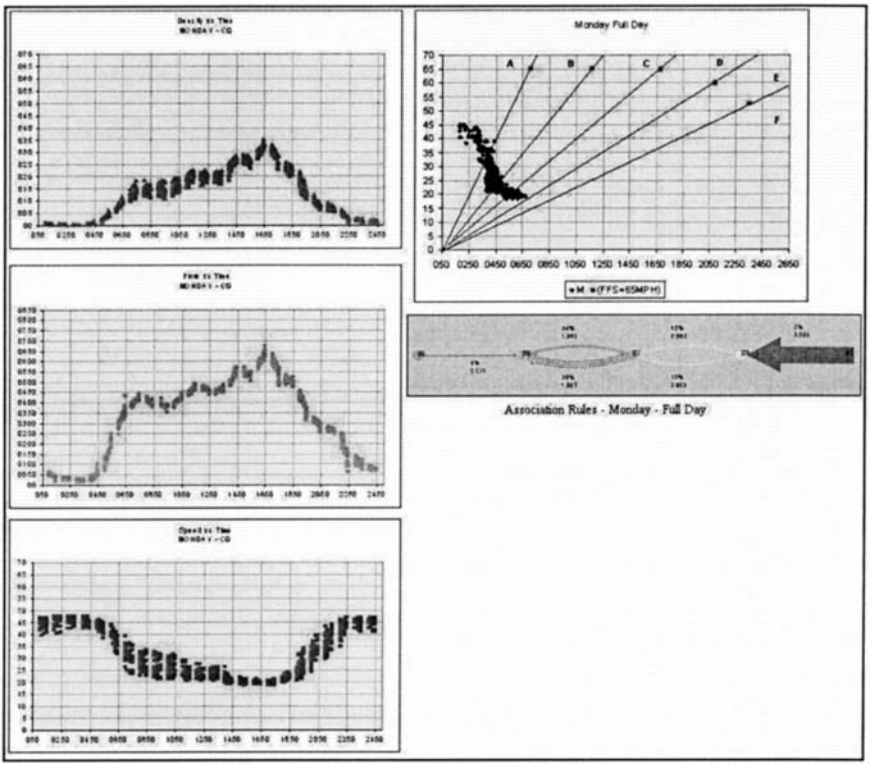


Figure 3. Graphs for monday model - caguas direction (southbound)

Dynamic Activity-Travel Diary Data Collection Using a GPS-enabled Personal Digital Assistant

B. Kochan[1], T. Bellemans[1], D. Janssens[1] and G. Wets[1*]

[1]Transportation Research Institute (IMOB), Faculty of Applied Economics, Hasselt University, Wetenschapspark 5 bus 6, B-3590 Diepenbeek, Belgium; PH (32) 11 26 91 58; FAX (32) 11 26 91 99 ; email: {Bruno.Kochan; Tom.Bellemans; Davy.Janssens; Geert.Wets}@UHasselt.be
[*]Corresponding author

Abstract

Activity-based transportation models have set the standard for modelling travel demand for the last decade. It seems common practice nowadays to collect the data to estimate these activity-based transportation models by means of activity-travel diaries. This paper presents a general functional framework of an advanced activity-travel diary data collection application to be deployed on a GPS-enabled personal digital assistant (PDA). The different modules, which are the building blocks of the application, will be scrutinized as well.

Introduction

In the past, four-step models have been developed in order to predict travel demand in the long run. The predicted travel demand, as outcome of the four-step models, can be used to support different kinds of decisions such as e.g. investments in new road infrastructure. In these four-step models, travel is assumed to be the result of four subsequent decisions which are modelled separately. More recently, especially in the eighties and early nineties, several researchers claimed that very limited insight was offered into the relationship between travel and non-travel aspects in the widely adopted four-step models. Indeed, travel has an isolated existence in these models and the question *why* people undertake trips is completely neglected. This is where activity-based transportation models come into play. The major idea behind activity-based models is that travel demand is derived from the activities that individuals and households need or wish to perform. The main difference between traditional (i.e. four-step) transportation forecasting methodologies and activity-based transportation models is that the latter attempts to predict interdependencies between several facets of activity profiles. These facets are

often identified as *which* activities are conducted *where*, *when* and for how *long*, with *whom*, and *which transport* modes are used.

As activity-based transportation models mature, they incorporate increasing levels of detail. An evolution towards dynamic activity-based models, which incorporates learning effects, can be observed in the literature (Joh 2004). The dynamics of travel-behaviour are driven by learning over time and short-term adaptation based on within-day rescheduling. As opposed to static models, dynamic models try to capture these dynamics using enhanced activity-travel data. In order to accommodate the growing data requirements for calibration and validation of the dynamic activity-based models, more detailed activity-travel diary data need to be collected. As the collection of basic activity-travel diary data puts a heavy burden on the respondents already, new techniques need to be developed to allow for the collection of even more detailed scheduling behaviour data. In this paper, a general functional framework of an advanced activity-travel diary data collection application to be deployed on a GPS enabled personal digital assistant (PDA) is presented. This tool must allow for the collection of detailed activity-travel diary data while limiting the burden for the respondents.

The remainder of this paper is organized as follows. The second section gives an overview of the state-of-the-art regarding computerized activity diary data collection tools. In the third section, the advantages and disadvantages of a GPS-enabled personal digital assistant will be discussed. In the fourth section, the functional description of the data collection tool will be given further consideration. In the last section, we will draw a conclusion.

State-of-the-art in Computerized Activity-Travel Data Collection Tools

CHASE (Computerized Household Activity Scheduling Elicitor) was the first computer aided self interview of activity scheduling behaviour (Doherty and Miller 2000). The purpose was to work out a survey which was able to track down the preceding scheduling process resulting into the definitive execution of an individual's schedule, along with the observed activity-travel patterns as the outcome. In the past, traditional survey techniques using diaries (e.g. paper-and-pencil techniques) were limited almost exclusively to observed patterns, providing little insights into decision processes. This shortcoming was dealt with through the development of a multi-day computerized scheduling interface. The users' task consisted of keeping track of their scheduling decisions by adding, modifying and deleting activities to their schedule as they occurred, during a multi-day period. The application made notes of each of these scheduling decisions, along with prompting for additional information (e.g. the reasons for these decisions, the exact timing of these decisions). The latter would be much more complex using paper-and-pencil techniques. Initial testing results indicated that this computerized approach revealed a considerable amount of information on the scheduling process and observed patterns, whilst minimizing respondent burden (Doherty and Miller 2000). Many adjustments and sophistications of this method have followed the original approach, including

applications on the Internet (Lee et al. 2000), development of a geographical information system interface for location and route tracking (Kreitz, Doherty and Rindsfüser 2002), and integration of GPS in a personal digital assistant application (Doherty et al. 2001).

In contributing to this line of research, the authors suggest an application which runs on a GPS enabled personal digital assistant (PDA). Several key development issues are: (i) desire to capture the dynamic activity-travel scheduling processes, (ii) desire to reduce respondent burden and (iii) to improve activity-travel data quality.

The application that is described in this paper captures the dynamic activity-travel scheduling process by first collecting information on the activities the respondent plans to execute and by collecting information on the activities that the respondent did execute in reality (diary) afterwards. Next, the planning and the diary are compared and additional information about the differences is gathered if required.

Advantages and Disadvantages of a GPS-enabled Personal Digital Assistant

In the past, desktop computer-assisted data collection tools were used for filling out scheduling surveys which provided activity-travel diary data. However, these systems are not able to trace the actual activity-travel execution due to their mobility constraints. In order to solve this problem, one might think of a personal digital assistant (PDA) with GPS technology for enhancing the data collection tool's mobility. The potential advantages of using a personal digital assistant with GPS to supplement travel survey data collection are numerous: (i) when using a desktop computer-assisted data collection tool, the respondents have to remember the exact locations of their start and end positions, whereas with a PDA with GPS, trip origin, destination, and route data are automatically collected without burdening the respondent for the data; (ii) as the respondent may forget to report an activity trip, another advantage exists in recovery of unreported trips, as all routes are recorded; (iii) accurate trip start and end times are automatically determined, as well as trip lengths; (iv) the GPS data can be used to verify reported data; (v) both the data entry cost and the cost of post-processing the data, constitute a significant share of the total data collection cost (Zhou 2004). Both costs can be reduced to a minimum with computer-assisted forms of data collection.

One of the most important shortcomings of GPS technology is the fact that the system is not always reliable throughout the entire trip-recording period. Indeed, civilian GPS receivers have potential position errors resulting from e.g. multipath, selective availability,... However, by combining the GPS data with other data sources such as e.g. location information reported by the respondent and geographical information system (GIS) maps, these errors can in general be overcome. Another issue associated with the use of a hand-held device is the storage capacity available to save the data that is collected. However, with the ever decreasing storage capacity prices, PDA's can readily be fitted with sufficient memory to conduct the surveys at a reasonable price. As a PDA is powered by a

battery, it has to be recharged regularly, which is an extra burden for the respondent. In order to reduce the number of times the PDA needs to be recharged, an energy conserving battery management system was integrated into the data collection application. This way the autonomy of the data collection tool can significantly be improved.

In the next section the functional description of the GPS-enabled activity-travel diary data collection tool is outlined.

Functional Description of the GPS-enabled Activity-Travel Diary Data Collection Tool

The central theme of the data collection tool revolves around a personal digital assistant (PDA). Compared to a typical computerized activity-travel diary data collection system (e.g. CHASE) a mobile system does not restrict the location for data collection and is easy to carry around by survey respondents for in-situation data input. Moreover, as the PDA is equipped with a GPS receiver, GPS data can be collected as well.

The system conceptually consists of two Graphical User Interfaces (GUI) (Household Survey and Activity Based Survey), a GPS logger, a data structure (Activity Diary & Household Data and GPS Data), a data quality control module (Data Integrity Checks), a Trip Identification module, a GIS module and a Communication module (see Figure 1). The modular structure of the application allows for customization. The implementation of modules that are less important for the research at hand can be omitted.

The GPS logger is used to trace the physical travel paths and the travel times. If the GPS logger is active, it receives the GPS data from the GPS chip and stores it using the GPS Data module. The GPS logger collects data continuously and therefore it needs to operate in the background. This automatic feature has two advantages. In the first place, it facilitates data capturing and secondly, although the survey respondent may forget to register a new activity, the GPS logger captures the user's position during the travel period. This way, the system can prevent the loss of activity-travel data. Indeed, once the system detects a change in location that is not reported as travel by the survey respondent, the system prompts the respondent for additional information.

The GPS data, stored in the GPS Data module, can also be used for trip identification. Once the performed trip is identified, it can be used to verify whether the information about activities, reported by the respondent, is consistent with the actually recorded trip. In case there are inconsistencies, the respondent will be prompted for further clarification.

The Household Survey GUI inquires for personal demographic and activity/travel-related information. These data are collected at the beginning of the survey period and are stored in the Activity Diary & Household Data module.

During the survey period, the respondent interacts frequently with the Activity Based Survey GUI, which is the major interface of the application. This GUI is used to register the activity-travel diary data during the survey period. It is used to enter, to modify or to delete an activity or a trip but it is also triggered if the Data Integrity Checks module detects an inconsistency (e.g. a city name that does not exist) and the Activity Diary Data needs to be altered.

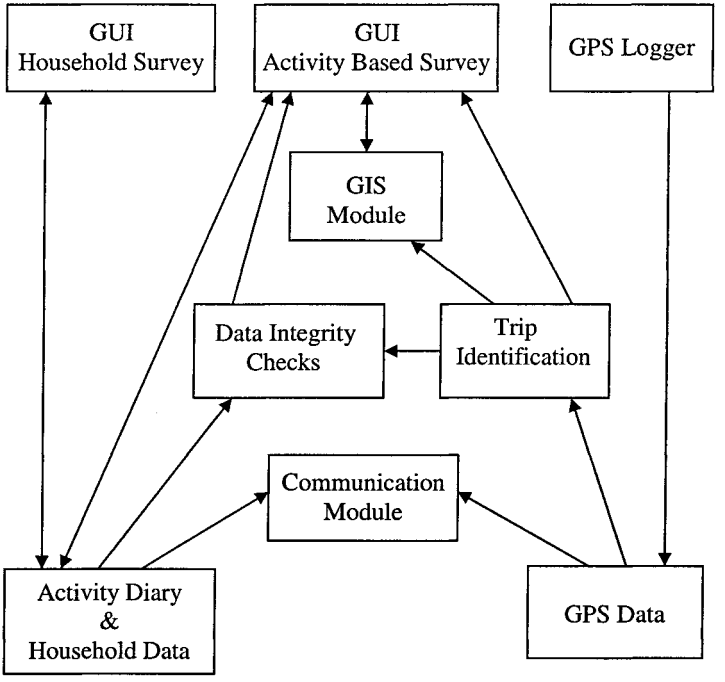


Figure 1. Schematic representation of the functional description modules composing the activity-travel diary data collection tool. The arrows denote the flow of information between modules.

The information stored on the hand-held devices can be downloaded through the Communication module. Depending on the implementation of the Communication module and on the available hardware, the data can be collected and stored on a data server either during or after the survey period.

The spatial dimension or the "where"-facet is the most difficult to collect in traditional paper-and-pencil diaries. People often do not precisely recall the exact

location or the street name where a particular activity was carried out. Hence, traditional diaries are often restricted by limitations about the detail of information that is collected. The computer-assisted data collection tools can make a significant contribution here by integrating a geographical information system (GIS) module, which enables the user to either pinpoint a location on a map or to manually enter a location.

Computer-assisted data collection tools allow for data quality control. Indeed, a computer system can check for anomalies and prompt the respondent for additional information. Errors that report activities where the start hour of an activity is later than the end hour, activity locations that do not exist and many others are detected by the Data Integrity Checks module of the PDA application.

Conclusion

In this paper, a data collection tool, which is able to capture dynamic activity-travel scheduling behaviour was presented. The detailed data collected by this tool will be used to develop a dynamic activity-based transport model.

In the functional description of the application, a modular approach towards a general data collection application was presented. Next, the importance of each of these modules was described. Currently, the presented data collection tool is deployed in a large scale activity-travel survey in Flanders, Belgium.

References

Doherty, S.T., and Miller, E.J. (2000). "A computerized household activity scheduling survey." *Transportation*, 27(1), 75-97.

Doherty, S.T. et al. (2001). "Moving beyond Observed Outcomes: Integrating Global Positioning Systems and Interactive Computer-Based Travel Behaviour Surveys." *Proc., Transportation Research Board Conference on Personal Travel: The Long and Short of It,* Washington D.C., No. E-C026, 449-466.

Joh, C.H. (2004). *Measuring and predicting adaptation in multidimensional activity-travel patterns,* CIP-Data Koninklijke Bibliotheek, Den Haag, The Netherlands.

Kreitz, M., Doherty, S.T., and Rindsfüser, G. (2002). "Collection of Spatial Behavioural Data and their Use in Activity Scheduling Models." *Proc., 81st Annual Meeting of the Transportation Research Board,* Washington D.C.

Lee, M.S. et al. (2000). "Conducting an Interactive Survey of Household Weekly Activities via Internet: Preliminary Results from a Pilot Study." *Proc., 9th International Association for Travel Behaviour Research Conference*, Gold Coast, Queensland, Australia.

Zhou, J. (2004). "Real-time Tracking of Activity Scheduling/Schedule Execution Within Unified Data Collection Framework." *Proc., 83rd Annual Meeting of the Transportation Research Board,* Washington D.C.

EXPERIMENTING WITH REAL-TIME ATIS: STEPPING FORWARD FROM *ADVANCE*

Alixandra Demers[1], George F. List[1], Jeffrey Wojtowicz[2,4], Alain Kornhauser[5], Al Wallace[3,4], Earl E. Lee[3,4], and Paul Salasznyk[3,4]

[1]Department of Civil, Construction, and Environmental Engineering, North Carolina State University, Campus Box 7908, Raleigh, NC 27695-7908; PH Demers (919) 515-4455, List (919) 515-2331; emails: ademers@unity.ncsu.edu, gflist@ncsu.edu
[2]Department of Civil and Environmental Engineering; PH (518) 276-8306; email: wojtoj@rpi.edu
[3]Decision Sciences & Engineering Systems; PH: Wallace (518) 276-6854, Lee & Salasznyk (518) 276-2759; emails: Wallace: wallaw@rpi.edu, Lee: leee7@rpi.edu, Salasznyk: salasp@rpi.edu
[4]Rensselaer Polytechnic Institute, 110 Eighth St., Troy, NY 12180;
[5]ALK Technologies, Inc., 1000 Herrontown Rd., Princeton, NJ 08540; PH (609) 683-0220; email: alaink@princeton.edu

ABSTRACT

In the early 1990's, an in-vehicle navigation and route guidance project called *ADVANCE* was conducted in the northeastern suburbs of Chicago. It proved that travel time data could be updated on in-vehicle devices (albeit not in real-time) to assist drivers in choosing faster routes to their destinations.

This past spring, about a decade later, a more progressive but similar 3-month field experiment was conducted in upstate New York. Nearly 200 participants used state-of-the-art, in-vehicle navigation and route guidance technology in conjunction with GPS tracking and broadband wireless to share travel time data and pick the shortest paths through a congested network. The route guidance devices observed travel times, uploaded them to a central server that updated a travel time database, and then downloaded every minute to each of the probe vehicles to ensure the latest travel time information was being used while enroute. The experiment resulted in a total of 4,111,210 latitude-longitude position/ speed/ time points. The largest number of location points per user was 98,018 while the smallest was 117; the average per user was just over 26,000 location points, or 325.5 points per trip. There were 12,629 probe trips for a traveled distance of 147,316 miles over a duration of 3,945.8 hours.

This paper presents a discussion of the Capital District ATIS project including the parallels and differences with the *ADVANCE* effort. Areas covered are: travel time data, project background, description of the study area, participant statistics, experiment design, sample results, and a summary with future research directions.

INTRODUCTION

In the early 1990's, the in-vehicle navigation and route guidance project called *ADVANCE* (**A**dvanced **D**river and **V**ehicle **A**dvisory **N**avigation **C**onc**E**pt) was conducted in the northeastern suburbs of Chicago. Planned as a large field project with hopes of 3,000 to 5,000 volunteer drivers testing equipment over 4 years, a host of public and private agencies worked as a team led by the Illinois Institute of Technology. Although the final field experiment included only 70 households and 110 drivers for approximately 2 weeks each between June and December 1995, it proved that travel time data could be updated on in-vehicle devices (albeit not in real-time) to assist drivers in choosing faster routes to their destinations (De Leuw 1997).

In Spring 2004, about a decade later, approximately 200 volunteers participated in the Capital District Advanced Traveler Information System (CD-ATIS). For three months, the participants tested a wireless, GPS-based, route guidance system. The travel times observed by them were uploaded to a central server, updating a travel time database, and then downloaded by the participants' devices every minute to ensure the latest travel time information was always available while enroute.

In this paper, the authors describe the project and how it compares to the *ADVANCE* project to give readers a sense of context. The study area, design, and participant selection of the experiment are described. Next, extractable results are mentioned followed by some sample results at the macro level. Finally, the paper is closed with a discussion of the researchers' future directions of inquiry.

PROJECT BACKGROUND

The CD-ATIS project was a joint public-private venture. Fourteen agencies in all joined together, led by Rensselaer Polytechnic Institute (RPI) and funded by the Federal Highway Administration and the New York State Department of Transportation, to make this experiment happen. The project budget was $1.4 million, significantly less than *ADVANCE*'s estimated $35 to 42 million (Boyce 2002).

The primary goal of the project was to test the viability of a stand-alone wireless, GPS-based, route guidance system where the probe vehicles themselves were the source of real-time, travel time observations. The *ADVANCE* project was based on a similar concept; the feedback was intended to occur instantly, yet they transferred data weekly since the outbound messaging did not work well without overloading the radio frequency. Another distinction was that the CD-ATIS project aimed for 200 participants while *ADVANCE* was planned for 3,000 to 5,000 participants. Although the main goal of both was to conduct an operational test, *ADVANCE* also had a broader purpose of fusing data from multiple sources whereas CD-ATIS intended to focus solely on probe data. Moreover, the *ADVANCE* study area included the entire northeastern section of Chicago whereas the CD-ATIS project focused on a few square miles.

Test Bed Area. The test bed was a small region of Troy and North Greenbush, New York as shown in the inset of Figure 1. The project participants were bound to two main traffic generators, the Rensselaer Technology Park (TechPark) and Hudson Valley Community College (HVCC). Plus, there were limited routes to access these

two locations so our team could easily monitor changes in traffic patterns, such as those due to incidents.

To enable the participants of the experiment to reach the test bed area and travel within the region for other trips, an approximate 40-mile radius study area (Figure 1) was programmed into the navigation software, a beta version of CoPilot Live provided by ALK Technologies, Inc. In Figure 1, the small black squares indicate "monuments" (think of these as virtual toll gates) programmed into the software as checkpoints for travel time estimation. With the study area established, the focus turned to experiment design and participant selection.

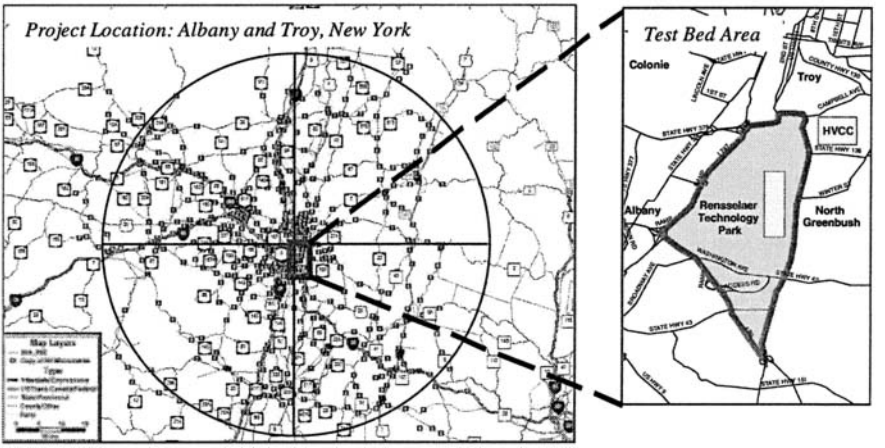


Figure 1 Study Area and Test Bed Location

Experiment Design. CD-ATIS experimental design focused on three issues (other than the technology to employ): the number of probes to use, the geographic area to cover, and the duration of the experiment. Funding drove the answer to two of these: only about 200 participants could be engaged for three months; and the limited number of participants dictated the answer to the third: a small geographic area needed to be the experiment's focus so that high enough probe densities would be reached, providing useful travel time information. Moreover, a further decision was made to focus on the morning commute, from 7 to 9a.m., so that the probe density would be as high as possible. To further improve expected probe densities on the arcs of interest, the team simulated the operation of the network, under various scenarios, to assist in selecting the best possible participants based on their home location.

Participant Selection. With the test bed area selected, the experiment was designed to capitalize on the heavy morning commute travel patterns in the Capital District area of New York (Albany, Troy area). As such, participants were selected because they made early morning (8 a.m.) trips to either HVCC or the Tech Park. The notion was that these travelers would benefit from the travel times of their fellow participants and the quality of information would increase as the morning progressed.

Three cohorts for a total of 201 participants were included in the experiment. The majority of participants (112 or 56 percent) were HVCC students with 8 a.m. classes while HVCC faculty and staff comprised only 4 percent (9) of them. Tech Park employees were 38 percent (76) of the participants. The remaining 4 participants (2 percent) were not among the main cohorts.

All participants were selected based on an online survey (using the Zoomerang website) and underwent uniform training by RPI researchers. Once selected, both a pre- and post-survey (again online) were conducted to collect the thoughts and driving perceptions of the participants.

System Selection. In concert with participant selection, the route guidance system was chosen based on cutting-edge two-way communication technologies and in-vehicle navigation systems (onboard vs. portable). Much had changed in available technology since *ADVANCE* where radio frequency communication protocol and dead-reckoning were the tools (Argonne National Laboratory 1997; De Leuw 1997).

As Cetin *et al.* (Cetin, Demers et al. 2004), have shown, there are three key aspects to the challenge of offering travel time information on a budget. One is determining how to instrument the system for observability by installing technologies that provide data about where the delays arise. The second is determining how collected data should be transformed into useful information (applies to all systems so not discussed here). The third is determining how the information should be disseminated to travelers.

Conceptually, tradeoffs exist among the sensing systems available in terms of their effectiveness and cost. Classically, the main sensing option is inductive loops. The cost of loops is moderate, whereas the data detail and maneuverability is low. Video data is the most expensive, the level of data detail is low as is maneuverability. AVI-based transponders provide much higher data detail and more maneuverability but at significant cost. Probes provide the greatest degree of data detail, and greatest maneuverability, at the lowest cost. However, AVL systems are not presently feasible to implement so very few vehicles are AVL-equipped today.

Dissemination options are basically five-fold. None of them are new. Radio and television are the two oldest and well known. Radio can send information to the travelers throughout their trip while television only helps during pre-trip decision making. Three more options have arisen from ITS efforts: highway advisory radio, variable message signs, and 511 services. Like radio and TV, these options disseminate the same information to all travelers; they are not tailored. Highway advisory radio can be focused on a particular market segment, and variable message signs can be focused on people traveling on a specific link, but it is difficult to create messages in such a way that each individual person's information needs are met. The traveler has to listen selectively to the broadcast, extract useful pieces of information, then integrate them into a mental picture of the network's condition that is helpful from his or her perspective.

The system chosen for the CD-ATIS involved four portable components: a Pocket PC loaded with RPI CoPilot Live, an SD card to hold map and trip data, a GPS receiver

to locate the vehicle, and a wireless communication card (3G card from Sprint). With these items, any car could be turned into a probe. ALK Technologies, Inc. partnered with the team to develop a beta version of their CoPilot navigation software with two-way communication and travel time estimation using real-time data. To support the server and assist in the detection of incidents, a command center was run by RPI personnel. It was fully staffed during the morning commute for system observing plus radar guns at two key locations in mornings and a 24-hour Help Center (via e-mail and on-site visits) for participant issues.

SAMPLE RESULTS

The experiment provided a rich data set offering insight on a host of topics at both macro and micro levels. For example, the resulting dataset can be used to examine trip making patterns across days, by specific groups of participants, on days with incidents, and so forth. A handful of examples are presented in this section; however, space constraints preclude detailed discussion.

Trip Characteristics – Macro Level. Macro level characteristics were mined from the raw GPS data for 162 of the 200 drivers. The experiment resulted in a total of 4,111,210 latitude-longitude position/ speed/ date/ time points. The largest number of location points per user was 98,018 while the smallest was 117 points. The average per user was just over 26,000 location points, or 325.5 points per trip. In terms of probe trips, there were 12,629 trips for a traveled distance of 147,316 miles over a duration of 3,945.8 hours. The average travel time, distance, and speed was 0.312 hours, 11.7 miles, and 32.3 miles per hour, respectively. Of note and as hoped, the majority of trips were quite short, less than approximately 9 miles.

Link Characteristics – Micro Level. Figure 2 includes a popular parallel pair of routes. It shows Mill Street and Morrison Street which are at the north end of the test bed. Note that Morrison St. is the favored route most days but that on days such as 50 and 70 there were higher probe traffic volumes on Mill St. and lower numbers on Morrison St. Studying these links, it is evident that travel times vary from day-to-day and during different time periods with long travel times on Morrison St. on 4/5/2005 (day 50) and 4/25/2005 (day 70), corresponding with the traffic volume shifts seen.

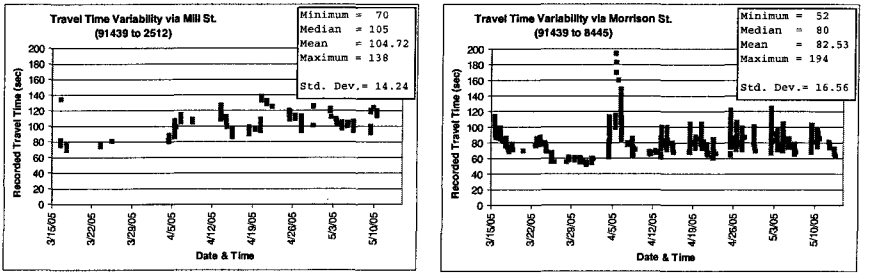


Figure 2 Travel Time Variability on a Sample Link

Traffic Flow – Micro Level. The final sample result (in Figure 3) is focused on analyzing traffic flow characteristics, namely, the travel time intensities of left,

through, and right turn movements by the probes. Figure 4 includes a graph with a small map showing the associated location and movements. Substantiating typical assumptions, this particular intersection has right turn movements with much shorter travel time intensities than the left turn movements.

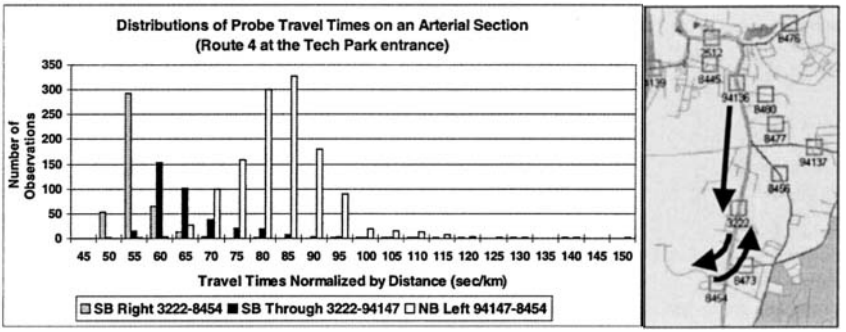


Figure 3 Travel Time Intensities along the Route 4 Arterial

FUTURE RESEARCH DIRECTIONS

The area of advanced traveler information systems is undergoing rapid growth, of which this field experiment (showing the real-time, two-way communication idea from *ADVANCE* is possible) was one step along the avenue of probe research. Based on our experiment, members of the team plan to explore this avenue further in the areas of: route guidance software usage levels and penetration rates, user compliance, path development, trip and link speeds and travel times, and estimated time of arrival forecasting. We hope this experiment inspires others to push their ATIS and ATMS research to extend the edges of ITS knowledge.

ACKNOWLEDGEMENT

The authors wish to thank all the research partners involved - ALK Technologies, Inc. (especially Alain Kornhauser); Cornell University; Polytechnic University; CUBRC; and Annese & Associates - for helping to make the project happen. Special thanks go to NYSDOT and the USDOT for providing the funding that turned the team's ideas into reality. More thanks go to Mecit Cetin (now an assistant professor at the University of South Carolina) for studying probe density issues. Finally, the authors are indebted to the team of twenty undergraduates from Rensselaer that manned the experiment every weekday starting in the wee hours of the morning.

REFERENCES

Argonne National Laboratory (1997). Formal Evaluation of the Targeted ATIS.

Boyce, D. (2002). "A memoir of the ADVANCE Project." Its Journal 7(2): 105-130.

Cetin, M., A. Demers, et al. (2004). Advanced Traveler Information System Traffic Surveillance Technologies. Troy, Rensselaer Polytechnic Institute Center for Infrastructure and Transportation Studies.

De Leuw, C. C. (1997). "ADVANCE Insights and Achievements Compendium."

De Leuw, C. C. (1997). Insights and Achievements Compendium, De Leuw, Cather & Company.

REAL-TIME BACKGROUND GENERATION AND UPDATE USING ONLINE CCTV VIDEOS

Woochul Lee[1] and Bin Ran[2]

ABSTRACT

This paper presents a robust background image generation and update. The foreground (i.e., moving objects) is typically identified by the difference between two consecutive images or between a background (i.e., non-moving parts) and an input image. The former can detect moving objects faster while the latter can detect more accurately. The accuracy of the background image is very important for detecting the foreground image, especially in the background subtraction method. The speed of background generation and update is also important in real-time based operation. The proposed method can generate the background image well enough to detect the foreground objects and update the image fast enough to reflect the environmental changes in illumination and weather and scene changes due to camera motion. The proposed method includes two steps, the background generation and the background update. In the background generation step, a differential image of two consecutive images was created to separate moving and non-moving parts. That differential image was compared with the previous background image to generate an updated one. The experimental results using online CCTV videos at several different locations and traffic situations show that our method can perform robustly and effectively.

INTRODUCTION

Traffic surveillance technologies play an important role in incident detection, traffic monitoring (management), violated vehicle detection, and travel information. Loop detectors are a typical sensor to have done this role. Video image detection, a substitute of loops, has been extensively used in recent few decades because it has more advantage features than conventional loop detectors. For example, it can cover more area, can extract more information, and is less intrusive to traffic than the loops. However, the performance of the video detection has been issued when the environmental changes or scene change due to camera motion have happened. The capabilities of video detection at night decreases dramatically. It is necessary to consider those changes for the video detection.

[1] Graduate Student, Department of Civil and Environmental Engineering, University of Wisconsin-Madison, 53706; Phone:608-262-3967; woochullee@wisc.edu

[2] Professor, Department of Civil and Environmental Engineering, University of Wisconsin-Madison, 53706; Phone:608-262-0052; bran@engr.wisc.edu

Typical traffic scenes contain vehicles, roadways, traffic facilities, and peripheral objects, such as trees, and sidewalks. In a sequence of traffic scenes, some vehicles are moving while others are stationary. Within each image, moving vehicles and objects related to traffic flow are called foreground and the remainder is referred to as background. Moving vehicles can be detected in a traffic scene by a temporal difference or background subtraction method. The temporal difference method detects objects by examining the intensity change between two consecutive traffic images. The background subtraction method detects moving objects by subtracting the previously created background image from the new input images. The former can detect moving objects faster while the latter can detect them more accurately. The main objective of this research is the development of a background generation method that will detect moving vehicles accurately and quickly in real-time operation.

A background image is also called a reference image in the image processing field. It typically contains uninteresting stationary parts of a scene. The background image is created to detect moving objects and identify changes in a traffic scene. An accurate background image is necessary to detect the foreground image. The speed of background generation and update is also important in real-time based operations.

A background image can be generated in several ways, mainly with statistical approaches which range from simple methods (i.e., average, mode, and voting etc) to complicated ones (i.e., Bayesian decision and kalman filter etc). The simplest method is to average each pixel value for a given time period. This works well if moving objects are relatively small and fast in a scene, but this doesn't occur all the time in many real traffic conditions. Because the quality of a background image is generally proportional to the number of stored images, this method requires a large amount of computer memory to store all the pixel values. Therefore, the average method is inappropriate for real-time image processing.

The score board method of background generation (Lai, 1998) assigns a positive score for small differences in pixel intensity between two frames, and a negative score for large ones. Whether the sum of the score until current frame is either negative or positive, the mode method or average method is selected, respectively. Accurate results using the mode method depend on the length of the stored images and require additional computing time.

Chung (2003) proposed the progressive method. This approach uses partial background images to eliminate the unstable effects of the physical vibrations of the camera. Histograms can record and trace changes in intensity to determine the best background values. In addition, this method can consider weather and illumination changes to generate a background image. The progressive method requires less memory to store the image sequences than the voting method, but it is slower than the average method.

Because environmental factors such as lighting and weather condition change, it is important to update the background image at predetermined time intervals. Use of the Kalman filter method to update the background increased the system's robustness to lighting changes (Gao, 2001) (Kilger, 1992).

BACKGROUND GENERATION

The proposed method generates the background accurately enough to detect the foreground and updates it fast enough to reflect environmental changes and scene changes due to the CCTV camera's motion. In the background generation step, the differential image of two consecutive images is created to separate moving and non-moving parts. The differential image is compared with the previous background image and an updated background image is created. The intensity difference of $(i\text{-}1)^{th}$ frame (F_{i-1}) and $(i)^{th}$ frame (F_i) is compared for each red, green, and blue (RGB) value. If the difference of each RGB value is less than the threshold (T_g) of 0.005, then the pixel is treated as a candidate pixel for the background and is interpreted as no change occurred at that pixel in the given time period. Since α is set to 0.5, the average RGB value of two images is assigned to the frame difference image (D_i) at that pixel. If the difference is greater than the threshold of 0.1, the pixel RGB value of the difference image (D_i) is set to zero, which represents a black color.

$$D_i(x,y) = \begin{cases} bg^C & \left|F_{i-1}^C(x,y) - F_i^C(x,y)\right| < T_g \\ 0 & otherwise \end{cases}$$

$$bg^C = \alpha F_{i-1}^C(x,y) + (1-\alpha)F_i^C(x,y)$$

where $C \in (R, G, B)$ is one of the RGB color components, i is a time, T_g is the threshold for the background generation step, α is a weight factor ranged from 0 to 1, $F_{i-1}{}^c(x,y)$ is the pixel color value of the $(i\text{-}1)^{th}$ frame, $F_i^c(x,y)$ is the pixel color value of the $(i)^{th}$ frame, x,y is the coordinate of an image on the X-axis and Y-axis, respectively. The α value of corresponding threshold of 0.005, 0.05, or 0.1 is fixed at 0.5, 0.6, or 0.95, respectively.

In the background update step, the difference image (D_i) and previous background image (B_{i-1}) are compared to create a current background image (B_i). Another threshold (T_u) is employed to decide an update scheme for each pixel. The absolute color intensity difference of background (B_{i-1}) and difference image (D_i) is a factor which determines a weight factor (α) for the update scheme. Based on the absolute difference and its corresponding threshold (T_u), α value is selected. For example, if the absolute difference at a pixel is less than the threshold of 0.005, then α value is set to 0.5 which means that a current background image pixel value is the average of two images for that pixel. If the absolute difference for a pixel is greater than or equal to the threshold of 0.1, a current background image pixel value is the sum of 95 percent of the current background and 5 percent of the difference image at that pixel value.

The current background pixel values are determined by the following equations.

$$B_i^C(x,y) = \alpha B_{i-1}^C(x,y) + (1-\alpha)D_i^C(x,y)$$

$$if \quad \left|B_{i-1}^C(x,y) - D_i^C(x,y)\right| < T_u$$

where $C \in (R, G, B)$ is one of the RGB color components, $B_i(x,y)$ is the current background image pixel value, $B_{i-1}(x,y)$ is the previous background image pixel value,

$D_i(x,y)$ is the difference image pixel value, α is a value between 0 to 1, i is a time, T_u is the threshold for the update step, x,y is the coordinate of an image on the X-axis and Y-axis, respectively. The α value of corresponding threshold of 0.005 or 0.1is fixed at 0.5 or 0.6, respectively.

Figure 1 shows a sample result of each procedure of the background image generation, update, and vehicle detection. The moving vehicle mask was created by simply subtracting the background image (B_i) from the input image (F_i) following by a noise removal. Most of the black parts on the difference image (D_i) could belong to the foreground (i.e., moving vehicles). Some of the black spots could be part of the roadway or trees. This ambiguity is a limitation of using thresholds to detect moving vehicles in the temporal difference method. However, this limitation does not generally affect the creation of a background image because the black color parts of the difference image (D_i) are usually small portion and could be made up in the following difference image. To generate a current background (B_i), only those non-black pixels of the difference image are examined with the previous background (B_{i-1}).

EXPERIMENTAL RESULTS

In our experiments, online CCTV videos were acquired from the WSDOT CCTV cameras installed on major freeways in the Puget Sound area in Seattle. Video clips of several different locations and twenty-four hours were collected and examined by using the proposed methods. Each video clip has five seconds of playing time and clips are typically updated every 90 seconds. Each image in the sequence is 320 by 240 pixels in joint photographic experts group (JPEG) format at a sampling rate of ten-frames per second. Every odd frame, however, was taken in this research not only to save computing power but also to provide enough information for experiments. The detection area of each scene was fixed at two thirds of the image height from bottom because of increasing the effectiveness of our method.

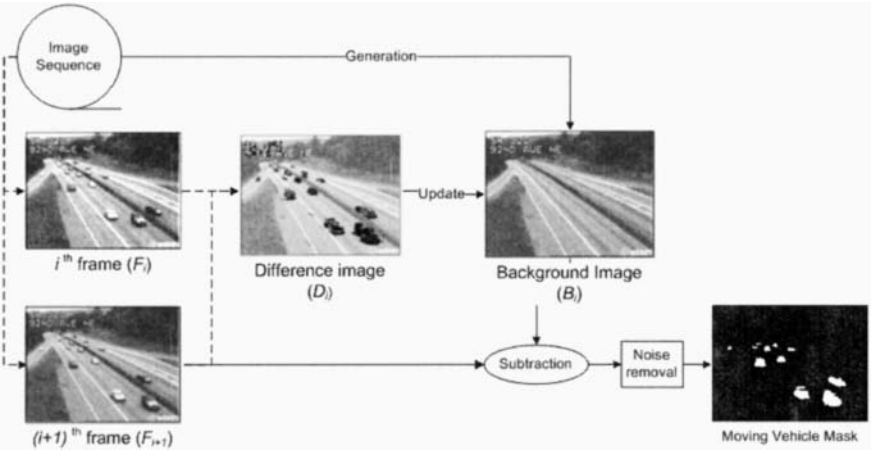


Figure 1. Sample results of each step in the proposed background image generation method.

One hundred of frames were used to create an initial background image, and then the background image was kept updating using a sequential input frame. If the difference of two input frames was too much (caused by sudden illumination change or scene change by camera moving), then the background image was regenerated by using another hundred of frames. Figure 2 shows background regeneration results when the CCTV camera's viewpoint was changed from east to west. Our method creates a new background image of west viewpoint in quickly and accurately.

Sample results are shown in Figure 3. The first case was taken during daytime with normal traffic flow. The background image was very good quality without noises and could separate moving vehicles accurately. The second case was taken daytime with high traffic flow. Since vehicles were moving very slow, certain parts of vehicle could be treated as a background. Vehicle blobs on mask image were collapsed a little bit due to noise of the background image and occlusion, but vehicle detection was still possible. The last case was a normal traffic flow at night. The reflection of headlights on the roadway caused a trouble for creating background image and detecting moving vehicles. It was like shadows under daylight which is always an issue of the video image detection field.

CONCLUSION

The proposed method performed robustly in normal traffic situation during daytime, and could be extended to apply at high traffic situation and/or nighttime. The gradual or sudden of illumination/weather change and scene change due to camera moving were both considered in our background image generation and update. For the further experiments, considering a dynamic threshold method may increase detection accuracy and reduce noise due to sudden changes. In addition, the shadow removal (i.e., reflection of headlight beam) should also be considered to reduce detection errors.

REFERENCES

Chung, Y. C., J. M. Wang, and S. W. Chen, "Progressive Background Images Generation," Journal of Taiwan Normal University, Vol. 47, No. 22. Oct 2002.

Gao, D.-s., J. Zhou, and L.-p. Xin, "A novel algorithm of adaptive background estimation," presented at Image Processing, 2001. Proceedings. 2001 International Conference on, 2001.

Kilger, M., "A shadow handler in a video-based real-time traffic monitoring system," presented at Applications of Computer Vision, Proceedings, 1992., IEEE Workshop on, 1992.

Lai, A. H. S. and N. H. C. Yung, "A fast and accurate scoreboard algorithm for estimating stationary backgrounds in an image sequence," presented at Circuits and Systems, 1998. ISCAS '98. Proceedings of the 1998 IEEE International Symposium on, 1998.

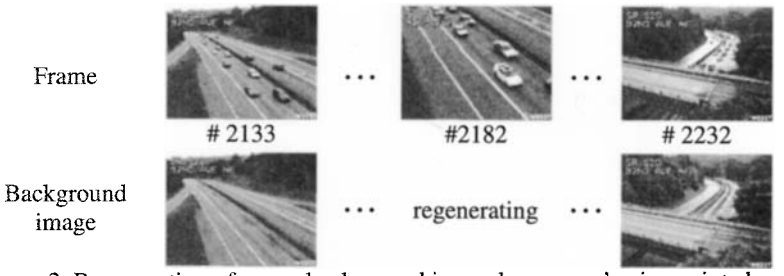


Figure 2. Regeneration of a new background image by camera's viewpoint change.

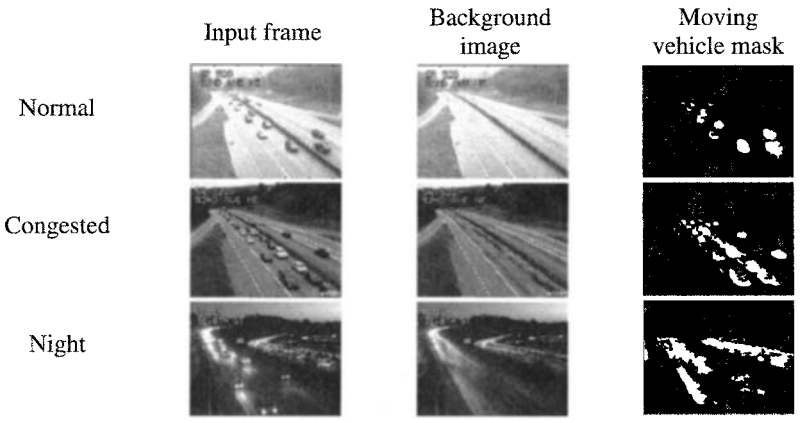


Figure 3. Experimental results with different situations.

Software-Based Traffic Management Information System for Urban Road Networks: An Empirical Study

R Gyampoh-Vidogah, R. Moreton
University of Wolverhampton, UK

University of Wolverhampton, Wulfruna Street, Wolverhampton. United Kingdom WV1 1SB. Fax +44 (0)1902 321478 email: r.gyampoh-vidogah@wlv.ac.uk and r.moreton@wlv.ac.uk.

Abstract

This paper presents a study as part of a growing body of research on traffic information systems implementation and their use within transport and traffic units. The paper starts by discussing traffic management systems in improving the competitive edge of transportation sectors and traffic divisions in particular. The paper then discusses the process of building an information management system and technology to improve the transportation and traffic units. The paper concludes with lessons learned from the experience of building traffic information systems and the steps needed to improve it. These steps will enable traffic managers to adapt their decisions; deliver efficient traffic management; make decisions that suites changing traffic patterns associated with incidents and special events on the road network. Consequently, traffic departments will no longer need to rely on snapshot views of traffic conditions. Traffic planning will be revolutionised because the system will incorporate traffic simulation models that will enable traffic departments to analyse congested networks.

Keywords: information systems implementation, traffic management, construction, and engineering, transport infrastructure

Introduction

Vehicular traffic flow over the transportation infrastructure affects almost every facet of life in the UK and has a primary impact on the economy. Faced with an increasing demand for road space, urban road network systems are reaching their existing capacity (Ullman et al 2001). In the UK for instance, statistics from the Department of Transport, (2004) show that, (i) traffic has increased by 79 per cent between 1980 and 2003, from 277 to 495 billion vehicle kilometres. Most of this growth occurred between 1980 and 1990 with traffic increase of 19 per cent; (ii) the majority of the growth has been in car traffic, which has gone up by 83 per cent since 1980, from 215 to 393 billion vehicle kilometres; (iii) light vans' traffic has more than doubled since 1980, from 26 to 58 billion vehicle kilometres and the distance travelled by heavy goods vehicles has also increased from 20 to 28 billion vehicle kilometres, a rise of 45 per cent since 1980; (iv) bus and coach traffic has also increased, from 3.5 to 5.4 billion vehicle kilometres. All indications are that, these increases in traffic are expected to continue. It has been estimated that, under pre-integrated Transport White Paper policies, road traffic in the UK will increase by 24% and 51% from 1996 to 2016. Journey times are expected to increase significantly in some areas very substantially, with times spent on journeys on urban motorways predicted to double in the peak period by 2031. Already one-fifth of the time taken in making a journey is spent stationary and on the busiest roads in many cities, journey times in the rush hour could lengthen by as much as 70% over the next twenty years. In many parts of the country, reducing traffic congestion is one of the main policy issues highlighted in recent reports (National Road Traffic, 1997).

However, as consequence of public resistance to uncontrolled road construction, changes in transport planning policy and programmes have been made in recent years. The new policy imperative is to ensure the best possible use of existing network infrastructure and resources. This policy shift manifests itself in a variety of ways including traffic management (Bannister, 2002). Consequently, the inability of existing road network to cope with increased demand has been identified as one of the most pressing infrastructure issues globally. Indeed, traffic problems have become social and economic embarrassment in the form of heavy congestion, deteriorating road safety, regression of mobility and the environmental effects of traffic (Carretero et al, 2003). Furthermore, traffic congestion does not just affect journey times it also exacerbates the environmental impact of vehicle use. In this respect, cautious estimates suggest that traffic related emissions of carbon dioxide are 25% higher in the centres of large cities, as a result of congestion (National Road Transport, ibid). To address this problem some researchers and practitioners have proposed traffic development and control theories combined with the traffic flow using decision support system (DSS). However, the support systems are not robust enough to aid urban road intersections (Zhu and Wu 2003; Ghaman *et al*; 2004; Davidson *et al*, 2005; Karim and Adeli, 2005). Sullivan *et al,* (2004) and Wu *et al*, (2005) recommended an intelligent decision-support tool to traffic departments to optimise road networks capacity by making changes to signal cycles. The limitation therefore is that, the system is unable to help traffic managers supervise, control and monitor traffic flow.

Based on the above statistics and problems identified, the primary aim and objectives of the research reported in this paper is to develop a software-based traffic information management system that will present a pictorial and realistic view of traffic conditions that will; (i) allow traffic departments optimise road networks capacity by making changes to signal cycles; (ii) implement diversion schemes and traffic information to

reduce to acceptable levels the congestion on the road networks. This research will lead to the provision of services and tools that can deliver advance urban road traffic management using real-time data gathering techniques and rendering algorithms in a geographical information (GIS) based system. This will encourage efficient and collaborative working to ensure the widest business benefit from the techniques, applications and technologies. The software will eventually accept any network grid as input and thus become an essential part of advanced traffic management system in the 21st century. This paper presents the outcome of the results of the preliminary study carried out to establish the architecture of the software based traffic management system.

Methodology

A preliminary case study based on interviews and observation technique was considered to investigate and address the problems. The case study method was selected because the goal of the research was not to achieve statistical generalisation at this point rather analytical generalisation (Yin, 1994). As such, employees of a public organisation were interviewed. This case study approach was more revealing because the aim was to establish current practical and technical problems with traffic management in line with approaches for identifying different paradigms for managing information systems and technology Achterberg, *et al*, (1993) and provide a detailed view of system requirement Denzin and Lincoln (2003). Finally, this study broadened the view of the information systems management within the wider context of transportation systems, which is of great relevance. In this study, interviews were used at various times throughout for different purposes to help: (i) gather facts about the procedures used for traffic management in the unit studied; (ii) ascertain levels of understanding of traffic managers and employees of the current system processes for implementing traffic management information system; and (iii) validate aspects of proposed system design to enable the new traffic management system to be implemented with confidence.

The Traffic Unit

The traffic unit studied is a division is a metropolitan city council with over forty-three (43) employees led by the head of services and traffic unit manager. At present the traffic control unit is unable to take advantage of existing infrastructure using state of the art traffic control systems, bus priority schemes and for the regulation of public parking space. Traffic managers are unable to adapt their decisions, deliver traffic management that responds to rapidly changing traffic patterns associated with incidents and special events on the road network. In addition, the department relies on snapshot views of traffic conditions using CCTV with banks of television screens and does not at present incorporate traffic simulation models that will enable traffic departments analyse congested networks.

Results of the Study

Studying Reports

Transport reports were studied to obtain a clear idea of the strategic direction of the traffic management unit. These reports stated any improvements that had been made by key decision-makers prior to the study. By studying the annual reports, interviews were tailored to the unit's traffic information systems. The report also touched on the need to improve efficiency by collaboration with external partners and other consultants, continued cost reduction and reliable diversion schemes of traffic flow.

These findings helped guide the range and depths of follow up questions after the initial interviews with the Traffic Manager and Head of Services.

Interviews

Current Problems with Traffic Management

According to the Head of Services, the current system does not provide robust information to commuters about the state of the transport or traffic. Additionally it does not identify and seek to ease local bottlenecks to the free movement of traffic. The current software/systems available is: (i) unable to accept any network grid because the current system is fragmented; (ii) there is no state-of-the art software technologies with the capacity to assist managers to determine traffic condition; (iii) traffic managers are not able to deliver efficient traffic management to reduce congestion because the system cannot handle changing traffic patterns associated with incidents on the road network; (iv) unable to optimise the road network capacity by making changes to signal cycles, implement diversion schemes and traffic information to reduce to acceptable levels the congestion on the road networks; and (v) the most cost-effective methods for obtaining real time traffic data cannot be quantified. The overall effect in his view confirming the national trend (Department of Transport, ibid.) is that as vehicle numbers grew, congestion has worsened.

Procedure for developing traffic systems

According to the head of services and the traffic manager, the success of this aspect of transport policy has been hampered by the unavailability of better tools for traffic managers to make timely and informed decisions.

Expertise and state of infrastructure

According to the traffic manager, the division often employs external software and traffic engineering consultants to assist in finding solutions specific problems. This include redesign of signal cycles, intersection design and data collection and analysis. In terms of the traffic infrastructure, there are limited number of traffic sensors supported by cameras that cover some roads. In their opinion, overcoming technical challenges for the simulation and realistic modeling of traffic conditions using software is a hurdle. Therefore there is a need for research on traffic information system implementation to enable the process of developing information management system and technology for traffic management.

Discussion of Traffic Management Systems

Literature review and the results of the preliminary studies suggest that, (i) the current software is unable to handle a complete network grid because of its fragmented infrastructure. (ii) There are no software technologies that have the capacity to simulate and build street models that will present a realistic view of traffic conditions; (iii) the most cost-effective methods for obtaining real time traffic data cannot be quantified and ways in which theoretical traffic modeling and simulation techniques can be incorporated within the software are not in place; (iv) inability to display complete pictorial view of traffic flow in the city. The consequence is increased and higher travel times for those commuting within the metropolitan area.

Recommendation for Traffic Information Systems

In view of the fact that interfacing and interoperability is necessary for software based traffic management system a modular approach to implementation is required in order to attain economy of scale. Consequently, a structured way of describing the system

with a view to ensure interoperability between its components is a priority. This will allow many different specific system implementation by grouping a number of common views that identify and describe the necessary commonalties for interoperability across these implementation. The main architecture view of the system will include: (i) A functional and control architecture showing the key processes that are to be performed by the traffic management system as well as any information and relationships between these processes; and (ii) Physical and communication architecture, showing the allocation of processes to physical entities (hardware/software) and the characteristics of the information and control flows should be established.

Generic information and management architecture

Figure 1 is a generic process illustrating the traffic interface comprises of driver task management system, traffic control and monitoring systems and vehicle control systems.

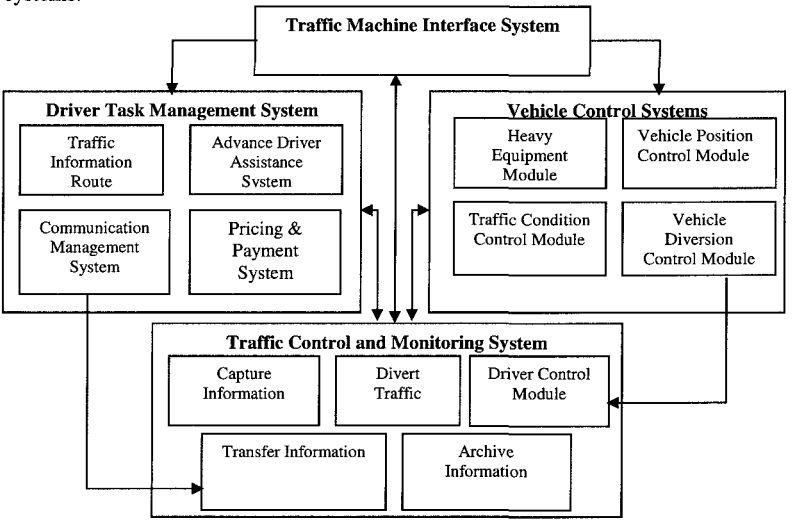


Figure 1: Generic information processes and architecture of traffic information

The driver task management system will provide the infrastructure for network centric information systems and computing. This module will provide the means to receive and process diversion schemes for traffic information to manage traffic on the road network. The system is able to communicate and assist drivers including the ability to automate the price of parking spaces. The traffic control and monitoring system have the capability of capturing data, transferring the data from sensors for processing. This module will be developed using electronic data management system and or knowledge based system to demonstrate that an integrated dynamic urban traffic management and control system can adapt and respond to traffic conditions in real-time using GIS technology. The vehicle control systems have the capability of optimising road networks. A new sensor grid will be implemented to include ground and vehicular-based and dedicated sensors, based on vehicular platforms, sensors employed by motorists, and embedded logistics sensors. The sensor grid will provide traffic managers with a high degree of awareness of heavy vehicles and buses across

intersections. The operational architecture of the grid will enable the planning and traffic operations in a manner that will achieve an overwhelming effect at precise places and time.

Physical Communication Architecture
Development of software-based traffic management information system for machine interface design will include the simulation of the behaviour of transport and traffic machine interface (TMI) within the traffic control station. In order to produce a high quality design of TMI, a method is required to analyse operators' monitoring in using the design. A recommended generic system that can simulate, analyse and evaluate the TMI design information from the viewpoint of traffic information should be considered. It is expected that advance driver-assistance systems will contribute to the reduction of traffic incidents. Moreover, if congestion is taken into consideration, this feature of the system will become even more important in the future. The key factors in realising that these systems are environmental recognition technology and high system reliability. Environmental-recognition devices such as millimetre wave radar and vision sensors should be put at certain points. These devices will receive information from the surrounding infrastructure and navigation systems which can then be used to divert vehicles and signal cycles.

Recommendation of steps needed to improve Traffic Information Systems
This research recommended that, software-based traffic management system should be reviewed in detail in relation to the current trends affecting: (i) traffic conditions and congestion (ii) traffic incident management by examining ways in which theoretical traffic modelling and simulation techniques can be incorporated within current software (iii) road network design, adaptive and integrated traffic signal management to enable traffic managers initiate decisions to clear traffic incidents faster and more efficiently. Overall, the system will ease the strain on the road network and reduce the fuel-wasted emissions associated with congestion and delays through the application of information technology. The system will use well-established geographical information system (GIS) and modelling technologies to provide traffic management system.

Advantages of the System
The design of this system will lead to the following benefits: (i) encourage traffic engineers to consider redesign of junctions and lanes leading to the optimisation of the existing road network capacity; (ii) enable traffic engineers to re-time their traffic signals, then simulate the results on the system and thus optimise their signal systems. This will result in reduced stops and delays and thereby decrease vehicular emissions in urban centres; (iii) enable managers to make informed decisions that will help reduce congestion to suite the rapidly changing traffic patterns associated with incidents and special events on the road network. Apart from the potential to assist the traffic division to make significant changes to traffic planning and management, the system should potentially extend to other cities and towns in much the same way. This is because by design this system can accept and use any network grid provided the traffic data is available.

Limitation
The limitation of the proposed system reported in this paper at this stage of the research is in the area of data collection and analysis. Consequently, further work needs to be undertaken in this area. The authors are unable to model traffic

management fully to the specification and implementation levels. Another area for development is the testing of the generic model. Preliminary testing through interviewers suggest that the concepts behind this research are capable of helping the development team make decisions on how to implement the system but further research work is needed. This process is ongoing.

Summary and Conclusion

This paper has presented the preliminary findings of research in the development of a software based traffic management system of a large metropolitan area in the UK. This described the design of generic architecture of software-based traffic management system with robust communication architecture. The design presented in this paper based on case study requires the best use of existing infrastructure through traffic control, bus priority schemes and the regulation of the price of public parking space. The experience of developing traffic information systems indicated that the transport policy depends on the availability of better tools needed for traffic managers to make timely and informed decisions. This study suggests that there the need for research on traffic information systems implementation and their use within transport and traffic units should be explored further to improve the transportation and traffic units in general.

References

Achterberg, J., Fischer, S. J. and Vinig, T. G. (1993). *Identifying different paradigms for managing information technology*, Proceedings of the Computer personnel research, ACM Press, St Louis, Missouri, United States: 37 - 55.

Bannister, D. (2002). *Transport Planning*, E&FN Spon, London.

Carretero, J.; Pérez, J.M.; García-Carballeira, F.; Calderón, A.; Fernández, J. José García, J.D.; Lozano, A.; Cardona, L.; Cotaina, N.; and Prete, P. (2003). Applying RCM in large-scale systems: a case study with railway networks. *Reliability Engineering and System Safety*, 82(3): 257– 273.

Davidson, J, Bouchart, F., Cavill, S and Jowitt, P (2005) Real-Time Connectivity Modelling of Water Distribution Networks to Predict Contamination Spread. *Computing and Civil Engineering* (19) 377-386.

Denzin, N. K., and Lincoln, Y. S. (Eds.). (2003). *Strategies of qualitative inquiry* (2nd ed.). Thousand Oaks, CA: Sage.

Department of Transport (2004). Roads, vehicles and congestion. http://www.dft.org.UK. > (February 2005).

Ghaman, R., Zhang, L., Li, Ling and Liu, Ke. Application of the Hardware-in-the-Loop Simulation for Updating the Traffic Signal Timing Plans. 8th International Conference on Applications of Advanced Technologies in Transportation Engineering 2004

Karim A. and Adeli, H. (2005). CBR Model for Freeway Work Zone Traffic Management. *Journal of Transport Engineering*. 129(2) 134-145 .

National Road Traffic Forecasts, (1997). Central estimates of Traffic Growth. All Vehicles.ww.dft.gov.uk/stellent/groups/dft_econappr/documents/page/dft_Econappr 610548.pdf.> (February 2005).

Sullivan, E.C, Gerfen, J. Recker W., and Yazdan, F. Upgrading the Technology of Simulation-Based Training for Traffic Management in California. 8th International Conference on Applications of Advanced Technologies in Transportation Engineering 2004

Ullman, G.L.; Fontaine M.D; Schrock S.D.; Wiles P.B.; *A Review of Traffic Management and Enforcement Problems and Improvement Options at High-Volume, High-Speed Work Zones in Texas*. Report No. FHWA/TX-01/2137-1. Texas Transportation Institute, College Station, Texas, February 2001.

Wu, Z; Yang, S and Gao J (2005). Research and Development of the Traffic Design Computer Aided System for Urban Road Grade Intersections in China. Proceedings of the 2005 ASCE International Conference on Computing in Civil Engineering. July 12–15, 2005, Cancun, Mexico.

Yin R.K. (1994). Case Study Research: Design and Methods (2nd ed): Thousand Oaks, Sage Publications, CA.

Zhu, Z. and Wu T. (2003). Two-Phase Fluids Model for Freeway Traffic Flow and Its Application to Simulate Evolution of Solutions in Traffic. *Journal of Transport Engineering*, 129(1) 51-56.

Statistical Models for Preferred Time Headway and Time Headway of Drivers in Steady State Car-Following

Ghulam H. Bham[1] and Siva Rama Prasad Ancha[2]
[1]Assistant Professor, Univ. of Missouri, Rolla, (ghbham@umr.edu)
[2]Graduate Student, North Dakota State University, Fargo (siva.ancha@ndsu.edu)

ABSTRACT

This paper proposes two shifted continuous distribution models for preferred time headway (Tp) and time headway of drivers in steady state car-following (Tws), i.e. the lognormal and gamma models. Lognormal distribution with a shift of 0.21 seconds provides a good fit to Tp, whereas a shift of 0.26 is proposed for Tws. Similarly, shifted Gamma distribution is proposed for Tp and Tws with shifts of 0.21 seconds and 0.26 seconds, respectively. Detailed data sets (FHWA Data) observed at one-second intervals collected from aerial photography for four sites have been used in this paper. The data sites are a basic freeway section, a ramp merge, a lane drop and a ramp weave section. Tp and Tws were found to be the shortest (mean, median and st. dev.) on the ramp weave section compared to all other sites. It was also found that the shifted Lognormal distribution provided a better fit to all the observed sites compared to the shifted Gamma distribution.

INTRODUCTION

Primary interest in preferred time headway (Tp) of drivers stems from its use in microscopic traffic simulation models. Simulation models like CELLSIM (Bham and Benkohal, 2004) and INTELSIM (Aycin and Benekohal, 1998) utilize Tp for computing the desired space gap (ds) of drivers (ds equals product of speed of follower and Tp), which represents the space gap desired by drivers during steady state car-following. Models such as NETSIM (Halati, et. al. 1997), INTRAS/FRESIM (Halati, 1991), and CARSIM (Benekohal and Treiterer, 1988) use reaction time of drivers to model the space gap that following vehicles maintain with a leading vehicle. Although, reaction time is difficult to obtain from field data, Tp can be found from field data by averaging the time gap of drivers during steady state car-following. However, the distribution of Tp is not known, neither the relationship between Tp and reaction time or time headway (Tw) of drivers is known. To use Tp in traffic simulation models, its distribution i.e. its mean and standard deviation are required. This paper determines the distribution of Tp and Tws in steady state car-following. The distribution of Tp is not different from the distribution of Tw, but due to the variation in vehicle length and difference in traffic composition, a shift exists between the two variables (Bham, Tanatamatorn and Benekohal, 2004).

DEFINITIONS

Tp is defined as the time difference between the rear bumper of a leader and the front bumper of a follower when the follower is in steady-state car-following with its leader. Tw is defined as the time difference between the front bumper of a vehicle and the front bumper of the next vehicle at a point on the highway. The Tp of drivers, therefore, does not take into account the length of the vehicles, but on the contrary, Tw and Tws take that into account. A driver is considered to be in steady state car-following when separated by less than or equal to 76 m (250 feet) and the speed and acceleration of leader and follower are equalized, i.e. the difference in speed between the leader and follower is less than or equal to ±1.52 m/sec (± 5 ft/s) (Aycin and Benekohal, 1998).

LITERATURE REVIEW

Tw and its distribution in the traffic stream has long been a subject of many studies. May and Wagner (1960) examined the headway characteristics on major urban highways and found that the mean, median and mode of Tw decreased as traffic volume increased. Mean Tw was found to be higher than the median, and the median was higher than the mode. May (1965) also examined the normal, negative exponential, shifted negative exponential, Pearson type III, and composite distributions for Tw of vehicles. The composite distribution was the combination of normal distribution of vehicles that were interacting and a shifted negative exponential distribution for vehicles that were not interacting. The shifted negative exponential distribution was tested using the Chi-square test, which failed at the 5% level of significance. Buckley (1968) examined several distribution functions and proposed a composite semi-Poisson model for Tw distribution as it provided the best evaluation. Data was collected on the median lane from a six-lane freeway in Sydney, Australia. Tolle (1971, 1976) examined the lognormal, composite exponential and the Pearson Type III distributions for traffic volume ranging from 800 to 1800 vehicles. He concluded that the lognormal distribution provided better fit than the other two models. He recommended the use of Kolmogorov-Smirnov (KS) test to overcome the problem of small variance, which may invalidate the results using the chi-square test. Cowan (1975) proposed four headway models i.e. M1, M2, M3 and M4. M1 used the Poisson process, M2 the shifted exponential distribution, M3 used the mixed and exponential distributions, and M4 used a generalized M3 and a general distribution. The M3 and M4 models were more realistic and provided excellent fit. The models were tested using the data from a two-way two-lane road in Sydney, Australia. Branston (1976) compared three models; the shifted lognormal model with a shift of 0.3, and two composite models; the queuing model proposed by Miller (1961) and the semi Poisson (Buckely, 1968) model. The composite models had headways for non-following and following vehicles. He found that the queuing model with a lognormal distribution for the following vehicle's headway provided the best fit for the data sites and the model could be used for a wide range of traffic flow and site conditions. Data was collected on a motorway in London (M4) and two-way roads in Indiana. Khasnabis and Heimbach (1980) examined negative exponential, Pearson Type III, Schuhl (Greco Parameter), Schuhl and Pearson Type III, Schuhl and negative exponential, and the Schuhl model. They found that the Schuhl model provides a good fit for the two-lane two-way roadway data in low volume conditions.

From the literature it was found that time headway distribution can be represented by the shifted lognormal distribution. Preferred headway represents driver behavior in steady state car following. No studies have been carried out to determine the distribution of Tp. This paper proposes the shifted lognormal and shifted gamma distributions for Tp and Tws. In the future, methods need to be devised to obtain the distribution of Tp given the distribution of time headway for different traffic and site conditions.

FIELD DATA SETS

Four data sets (Smith, 1985) collected using aerial photography for 1-second intervals are used in the study. The data sets provided lateral and longitudinal position, speed, and length of individual vehicles. The chosen sites are a) a ramp weave, b) ramp merge, c) lane drop and d) a basic freeway section of highway. The sites are shown in Figure 1. Data for the median lane are used in the study to minimize the effect of lane changing on headway of drivers. The data sets were collected at p.m. peak period and are accurate within 0.9 to 1.5 m.

Data Extraction and Analysis of Data Sets

Two types of headway were obtained from the data sets. First, Tp of following vehicles was calculated for pair of vehicles in steady state car-following. To calculate Tp, vehicles at speeds greater than 72 km/h (45 mph) were observed, as drivers driving at speeds lower than

72 km/hr may be forced to follow leader at shorter space gaps compared to their preferred space gaps. Second, time headway data for vehicles in steady state condition, indicated by Tws in the paper, was reduced similar to Tp.

For Tp and Tws, data for pair of passenger cars was extracted from the median lane only. Data from the pair was discarded when either the leader or the follower changed a lane during the observation. Data was also discarded when a third vehicle entered between a pair of vehicles. Headway data for passenger cars was used only, as the number of long vehicles was very low. In computing Tp and Tws, only vehicles in steady state car-following were considered. This does not imply that the entire traffic stream is in steady state condition, but it means that the pairs of vehicles which are in steady state condition are considered for analysis.

For the four data sets, minimum, maximum, mean, median and the standard deviation of headways are presented in Table 1. As observed in other studies (May and Wagner, 1960; Sadeghhosseini and Benekohal, 2002), the mean was found to be greater than the median, which indicates a large concentration of shorter headways. The mean and median for the weaving section is the lowest compared to all other sites.

Table 1. Time Headway Characteristics

	Preferred time headway (Tp)				Time headway, steady state condition (Tws)			
Data Site	Ramp Weave	Ramp Merge	Lane Drop	Basic Freeway	Ramp Weave	Ramp Merge	Lane Drop	Basic Freeway
St. Dev	0.532	0.605	0.568	0.567	0.531	0.604	0.568	0.568
Mean	1.050	1.313	1.210	1.253	1.258	1.499	1.396	1.445
Median	0.881	1.176	1.047	1.117	1.098	1.353	1.238	1.305
Max	3.253	3.561	3.027	3.240	3.452	3.786	3.254	3.426
Min	0.345	0.332	0.316	0.349	0.545	0.469	0.540	0.562
Count	245	595	611	375	245	595	611	375

MODELING PREFERRED TIME HEADWAY AND TIME HEADWAY

To model the time headway distributions, headway data was grouped into 0.2-second intervals. The lognormal model was fitted to the plots, which didn't provide a good fit to the data sets. The shifted lognormal distribution was then tried, which provided a good fit to the data sets. The shifted lognormal distribution was selected, as several studies described in the literature showed a good fit to the field data. Shifted Gamma distribution was also fitted to the data which provided a good fit for Tws.

Shifted Lognormal Distribution

The probability density function of a shifted lognormal distribution is expressed as (Law and Kelton, 2000):

$$f(x)=\begin{cases}\dfrac{1}{(x-\gamma)\sqrt{2\pi\sigma^2}}\exp\dfrac{-[\ln(x-\gamma)-\mu]^2}{2\sigma^2}, & \text{if } x>\gamma\\ 0 & \text{otherwise}\end{cases}$$

where γ is shift, σ^2 is variance, and μ represents mean for $\sigma > 0$ and any real numbers γ and μ. Mean (μ) and the standard deviation (σ) were estimated using the maximum likelihood estimators as (Law and Kelton, 2000):

$$\hat{u}=\frac{\sum_{n=1}^{N}\ln(X_n-\gamma)}{N},\ \hat{\sigma}=\left\{\frac{\sum_{n=1}^{N}[\ln(X_n-\gamma)-\hat{\mu}]^2}{N}\right\}^{1/2}$$

where X_n represents time headways, and N represents total number of observations.

The lognormal models with shifts ranging from 0.15 to 0.5 seconds were examined. Tp and Tws were found to provide a good fit with a shift of 0.21 and 0.26 seconds, respectively. Table 2 shows the values of shift for the different headways and the data sets, and the results of the KS test.

Shifted Gamma Distribution

The probability density function of a shifted gamma distribution is expressed as:

$$f(x)=\begin{cases}\dfrac{\beta^{-x}(x-\gamma)^{\alpha-1}e^{-(x+\gamma)/\beta}}{\Gamma(\alpha)}, & if\ x>0, x>\gamma\\ 0 & , \text{otherwise}\end{cases}$$

where α is the shape parameter, and β is the scale parameter. Approximations to α and β were obtained using the following equation:

$$T=\left[\ln\sum_{n=1}^{N}\frac{(X_n-\gamma)}{N}-\sum_{n=1}^{N}\frac{\ln(X_n-\gamma)}{N}\right]^{-1}$$

α was then obtained as a function of T using Table 6.20 (Law and Kelton, 2000) and β using the following equation:

$$\hat{\beta}=\sum_{n=1}^{N}\frac{(X_n-\gamma)}{N}\cdot\frac{1}{\hat{\alpha}}$$

Shifted gamma model was also tested using the K-S test with shifts that ranged from 0 to 0.47 sec for both Tp and Tws. The results of the goodness of fit tests for both distributions are discussed next.

Table 2. KS Test for Headway of Vehicles: Shifted Lognormal & Gamma Distributions

	preferred time headway (Tp)				time headway, steady state condition (Tws)			
Data Site	Ramp Weave	Ramp Merge	Lane Drop	Basic Freeway	Ramp Weave	Ramp Merge	Lane Drop	Basic Freeway
Shift	0.21	0.21	0.21	0.21	0.26	0.26	0.26	0.26
N	245	595	611	375	245	595	611	375
Dn LogNor	7.42%	5.43%	3.11%	4.7%	7.81%	4.18%	2.73%	6.55%
Dn Gamma	7.47%	3.90%	2.64%	6.07%	7.65%	2.76%	3.11%	7.92%
dn,0.01	10.41%	6.68%	6.59%	8.42%	10.41%	3.11	6.59%	8.42%
dn,0.05	8.69%	5.58%	5.50%	7.02%	8.69%	5.58%	5.50%	7.02%

Kolmogrov-Smirnov (KS) Goodness of Fit Test

KS test was used to test the difference between the shifted lognormal distribution and the field data. KS test is used specifically ahead of Chi-square test because Chi-square test has the ability to invalidate the test because of the small variances in the data. On the contrary, KS test overcomes this drawback and produce statistically good fits to the data. Also, the necessary calculations are simple in the KS test. Hence, KS test is used, as it is more powerful than the chi-square test (Banks J, 2001). First, the distribution functions for field data F(x) and model G(x) were calculated. Then, D_n, the maximum absolute difference between the two distributions was computed, where 'n' denotes the number of samples from the two distributions. The null and alternative hypotheses were: H_o: F(x) = G(x) versus H_1: F(x) ≠ G(x), for all x. H_o is rejected at level α, if $D_n > d_{n,\alpha}$,. The models were tested at 0.01 and 0.05 levels of significance. The results are presented in Table 2. For the Gamma

distribution, the KS test failed at the 0.05 level of significance for the basic freeway segment. This can be attributed to the fact that the Gamma distribution predicts low frequencies at small headways when compared to lognormal distributions, which made the test invalid. It is therefore inferred that the shifted lognormal distribution provides better fit to Tp and Tws data compared to shifted Gamma distribution. For the statistical tests, it can be inferred that the shifted lognormal model provides good fit to Tp and Tws. It can also be inferred that the shifted lognormal model can be used to generate Tp of vehicles for use in simulation models. This also supports the work by Mei and Bullen (1993) in which shifted lognormal distribution was used to fit the headway of vehicles at high volume conditions in which many vehicles are in car-following. The distribution of Tp is not necessarily different from the distribution of Tw, but due to the variation in the lengths of vehicles and the vehicle composition, the shift between the two distributions is different. More detailed discussion is out of the scope of this paper.

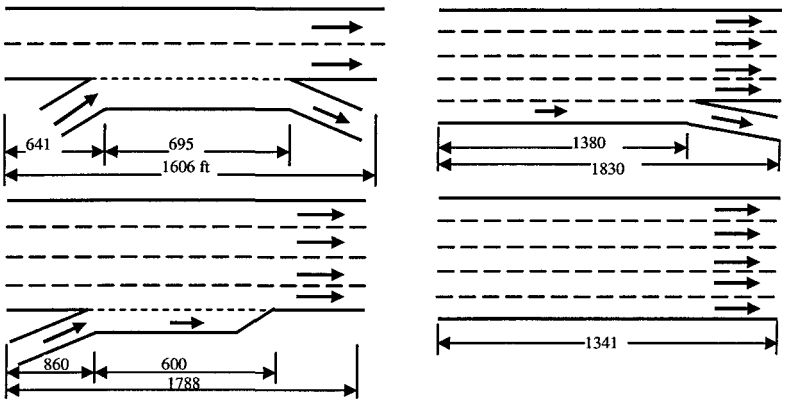


Figure 1. Ramp Weave: Baltimore - Washington Parkway I-95 N (top left), Ramp Merge: Roscoe Blvd. I-405 NB, Van Nuys, CA, (bottom left), Lane Drop: Ventura Freeway at White Oak, US101 WB, Van Nuys, CA (top right), Basic Freeway Segment: Mulholland Drive I-405, Los Angeles, CA (bottom right)

CONCLUSIONS AND RECOMMENDATIONS

The paper determines the headway distribution of car-following drivers in steady state at their preferred time headway (preferred time gap) 'Tp' and time headway 'Tws'. The headway data was observed from four sites i.e. a ramp weave, lane drop, a ramp merge and a basic freeway section. These four types of sites were selected in order to investigate and represent the driver behavior at different sites with different traffic characteristics. Shifted lognormal distribution was found to be representative of Tp and Tws. Tp with a shift of 0.21, a standard deviation of 0.57, and a mean of 1.2 seconds was found to be representative of the distribution for the four sites studied. Similarly, the Tws distribution for the four sites can be modeled with a shift of 0.26, standard deviation of 0.57 and a mean of 1.4 seconds. The proposed distributions may be used to generate Tp for similar sites and traffic flow conditions for use in microscopic traffic simulation models.

The shape of density functions for Tp and Tws were found to be similar. Figure 2 shows the shifted lognormal and shifted gamma distribution for preferred time headway and time headway (steady-state condition) for lane drop section. Comparison of Tp for the four sites showed that the weaving section has the highest number of shorter values of Tp compared to other sections. It is recommended that a relationship between Tp and Tws, and

between Tp and the reaction time of drivers be determined for different sites and traffic flow conditions. The interrelationship between Tp and traffic flow stream characteristics also needs to be determined.

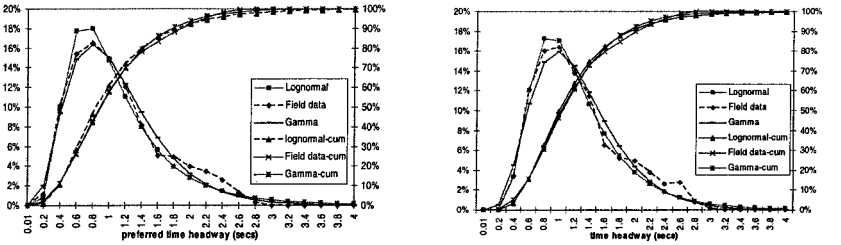


Figure 2. Shifted Lognormal and Shifted Gamma Distribution for Preferred Time Headway and Time Headway (steady-state condition) for Lane drop section

REFERENCES

Aycin, M. F. and R. F. Benekohal. (1998). "Linear Acceleration Car-Following Model Development and Validation." *Transportation Research Record,* 1644, NRC, Washington, D.C., 10-19.

Banks, J., et. al. (2001). "Discrete-Event System Simulation." *3rd Edition*, Prentice Hall, 269.

Benekohal, R. F., and J. Treiterer. (1988). "CARSIM: Car Following Model for Simulation of Traffic in Normal and Stop-and-Go Conditions." *Transportation Research Record,* 1194, NRC, Washington, D.C., 99-111.

Bham, G. H. and R. F. Benekohal. (2004). "A High Fidelity Traffic Simulation Model based on Car-Following and Cellular Automata Concepts." *Transportation Research C,* Vol. 12, No.1, 1-32.

Bham, G. H., N. Tanatamatorn and R. F. Benekohal. (2004). "Models for Preferred Time Headway of Drivers and Interrelationship with Fundamental Characteristics of Traffic Flow." *Transportation Research Board Proc.,* NRC, Washington, D.C.,

Branston, D. (1976). "Models of Single Lane Time Headway Distribution." *Transportation Science*, Vol. 10, No.2, 125-148.

Buckley, D. J. (1968). "A Semi-Poisson Model of Traffic Flow." *Trans Science*, Vol.2, No.2, 107-133.

Cowan, R. J. (1975). "Useful Headway Models." *Transportation Research*, Vol.9, 371-375.

Halati, A. (1991). "FRESIM - Freeway Simulation Model." *Paper #910202, 70th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Halati, A., H. Lieu, and S. Walker. (1997). "CORSIM - Corridor Traffic Simulation Model." *Traffic Congestion and Traffic Safety in the 21st century*, American Society of Civil Engineers, 570-576.

Khasnabis, S. and C. L. Heimbach, (1980). "Headway-Distribution Models for Two-Lane Rural Highways." *Transportation Research Record* 772, NRC, Washington, D.C., 44-50.

Law, A. M. and W. D. Kelton. (2000). "Simulation Modeling and Analysis." *2nd Edition,* McGraw Hill.

May, A.D. and Wagner, F. A. (1960). "Headway Characteristics and Interrelationships of Fundamental Characteristics of Traffic Flow." *Highway Research Board,* Vol. 39, 524-547.

May, A.D. (1965). "Gap Availability Studies." *Highway Research Record* 72, National Research Council, Washington, D.C., 105-136.

Mei, M. and Bullen, A. G. (1993). "Lognormal Distribution for Highway Traffic Flows." *Transportation Research Record* 1398, National Research Council, Washington, D.C., 125-128.

Miller, A. J. (1961). "A Queuing Model for Road Traffic Flow." *J. Royal Stat. Soc.* Series B, Vol. 23, 64-75.

Sadeghhosseini, S., and Benekohal, R. F. (2002). "Headway Models for Low to High Volume Highway Traffic." *Transportation Research Board Proc.,* NRC, Washington, D.C.

Smith, S. A. (1985). "Freeway Data Collection for Studying Vehicle Interaction-Technical Report." *Report No. FHWA/RD-85/108,* U.S., Department of Transportation.

Tolle, J. E. (1971). "The Lognormal Headway Distribution Model." *Traffic Engineering and Control,* Vol.13, No.1, 22-24.

Tolle, J. E. (1976). "Vehicular Headway Distributions: Testing and Results." *Transportation Research Record* 567, National Research Council, Washington, D.C., 56-64.

A Case Study on Measuring Travel Time, Speed, and Delay Using GPS-Instrumented Test Vehicles

Hoe Kyoung Kim[1], Seung Kook Wu[1], and Michael Hunter[2]

[1] Graduate Research Assistant, School of Civil & Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Drive, Atlanta, GA 30332-0355; PH (404-894-2236); FAX (404-894-2278); email: gtg199j@mail.gatech.edu, gtg169n@mail.gatech.edu

[2] Assistant Professor, School of Civil & Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Drive, Atlanta, GA 30332-0355; PH (404-385-1243); FAX (404-894-2278); email: michael.hunter@ce.gatech.edu

ABSTRACT

The Georgia Institute of Technology research team has recently conducted an arterial performance monitoring study on a 15-intersection signal control system in Cobb County, Georgia, using GPS-instrumented test vehicles. This paper describes the defined GPS error detecting criteria and the data reduction procedure developed by the research team. Error detecting criteria are developed based on distance between two consecutive GPS points, acceleration rate, deceleration rate, horizontal dilution of precision (HDOP), and number of satellites. The impact on percentage of GPS data eliminated as a result of these varying criteria is given and performance measures are calculated for several different criteria values, allowing for a determination of the sensitivity of the performance measures to the error checking criteria set.

1. INTRODUCTION

The performance of urban arterial streets is affected by a variety of factors such as roadside activities, traffic control devices, arterial geometry, and so on. A dominant factor in urban arterial street operations is the presence of traffic signals, which govern the flow of vehicles that may enter and exit an arterial segment. Consequently, the performance of an arterial street is predominantly influenced by delays incurred at signalized intersections. Travel speed and intersection approach delay, two of the most frequently utilized measures of effectiveness on an arterial street, may be calculated as a function of arterial segment travel time, resulting in travel time as a primary characteristic utilized in field measurement of arterial performance.

The Georgia Institute of Technology research team has conducted a performance measure study with GPS-instrumented test vehicles as a part of a before and after study to evaluate a 15 intersection traffic signal system in a suburban area in Cobb County, Georgia. This paper describes the defined GPS error detecting criteria and the data reduction and calculation procedure developed by the research team. Error detecting criteria are developed based on distance between two consecutive GPS points, acceleration rate, deceleration rate, horizontal dilution of precision (HDOP), and number of satellites. The impact of these criteria on the percentage of GPS data eliminated and arterial performance measures is provided, allowing for a

determination of the sensitivity of the performance measures to the error checking criteria set.

2. GPS DATA COLLECTION PROCESS

The typical GPS system mounted in a test vehicle consists of a PDA (iPAQ h2215) or laptop, a GPS receiver (HAiCOM 303E Multi-Mode Foldable GPS Receiver), an external antenna, and a power cable. Satellite signals are received by the antenna, processed by the receiver, and stored in the laptop or PDA. With this system, time, speed, latitude, longitude, HDOP, and number of satellites are recorded at one-second intervals. The research team determined three fixed routes through the traffic system with each intersection in the system covered by at least two of the fixed routes. A log sheet was provided to each driver to record start time, end time, route ID, and any comments (i.e., incidents, driver mistakes, roadside construction, etc.) for each run.

3. GPS DATA REDUCTION PROCESS

The GPS data reduction process has two steps: initial error checking and data processing, followed by the Travel Run Intersection Passing Time Identification (TRIPTI) algorithm. The initial error checking and data processing procedure flags any GPS points identified by the five criteria (distance between two consecutive GPS points, acceleration rate, deceleration rate, HDOP, and number of satellites), interpolates missing data points, removes runs in which driver errors occurred (driver errors are primarily identified by reviewing driver logs), and attaches run identifiers to the GPS data. Data points are interpolated where only a few seconds of data are missing. Runs with large sections of missing data, or a high number of GPS points flagged as potential errors, are superimposed on a digital map and inspected visually. Unless the visual inspection is able to confirm the accuracy of these points, or there exists a reasonable ability to interpolate points, the entire run is eliminated for further analysis.

After the initial error checking and data processing, each run is processed through the TRIPTI algorithm. The TRIPTI algorithm is utilized to identify the time at which a test vehicle completes its pass through an intersection, enabling an intersection-to-intersection travel time calculation. Then, for each run, the sequence of intersections through which the test vehicle passes and the time at which the test vehicle crosses the exit boundary of each intersection are determined. Only crossing times that do not contain an error flag from the error checking process are retained for Measure of Effectiveness (MOE) calculations. Routes and intersection crossing time identification requires that the intersection and intersection boundary coordinates be set as the initial step of the algorithm. Currently this process is completed through visual inspection. Further recalibration of these values is also undertaken during the first few runs of the TRIPTI algorithm for each route.

Based on the output from the TRIPTI algorithm, travel time, average delay, and average speed for *any* two intersections included in the travel runs, not just consecutive intersections, can be calculated. This allows an analyst to consider different corridor segment lengths for performance measures, from between consecutive intersections to corridor sections incorporating several intersections or to the entire corridor over different time periods (Hunter, Wu, and Kim, 2006).

4. CRITERIA FOR DETECTING GPS ERROR POINT

This study utilized four primary criteria to identify potentially erroneous GPS data points. GPS data violating any of these criteria were labeled as potential GPS errors points.

Distance Error. The critical distance criterion value is based on the maximum speed reported by the travel time run drivers or another known reasonable maximum. For example, in the Georgia Tech study the highest reported speed from drivers was 50 mph, thus a distance of 80.67 feet (distance traveled at 55 mph in one second, which allows for some driver perception error) was utilized as the maximum allowable distance between consecutive GPS points.

Speed Error. The speed error criterion is defined by maximum acceleration and deceleration rates as these represent the possible speed change in a given time period (one second). Where acceleration and deceleration violate the set criteria, the GPS points should be regarded as potential error points. For the following analysis a maximum acceleration rate of 5.37 mph/s and maximum deceleration rate of -7.64 mph/s values were utilized (ITE 1999, AASHTO 2001).

HDOP Error. Dilution of precision (DOP) values are commonly used to indicate the quality of the geometric configuration of satellites. This study utilized the horizontal dilution of precision (HDOP) value to evaluate horizontal accuracy (2D-coordinates) of GPS data points. Desirable HDOP values are below four, and values above eight should not be used (Kowama, 2005). The HDOP cutoff value utilized in a study requires a balance between desired data accuracy and loss of potentially accurate data.

Number of Satellite Error. Unlike the above criteria, the number of satellite criterion is relatively inflexible. It must be set to four or more to satisfy the minimum requirement for GPS technology to solve the four simultaneous equations required to determine positional information (x, y, and z coordinates) and clock bias (Logsdon, 1995).

5. SENSITIVITY ANALYSIS

The primary intent of the sensitivity analysis is to develop an understanding of the characteristics of the data to aid in the development of bounds specific to the given data for the five error criteria. The initial sensitivity analysis is conducted on all of the signal system travel run data, with no data removed by the above mentioned error checking process. The data set is comprised of the 715 runs collected on the three fixed routes over six weekdays and four weekend days from November 9, 2004 to December 8, 2004. The 715 travel runs contain a total of 324,289 GPS points.

Sensitivity of GPS Data to Criteria. The proportional change of GPS data loss observed when increasing or decreasing a criterion value is a useful indicator to help in determining the criteria error flag values. Figure 1 shows sensitivity analyses of criterion values to the GPS data loss. Since the criterion value of number of satellites must be greater than or equal to four, the effect of the varying number of satellites on GPS data loss is not considered.

For the distance criterion, the number GPS points dramatically decrease until 70 to 80 ft level, corresponding to 48 or 55 mph, respectively. Considering that the maximum speed reported by the travel time run drivers is approximately 50 mph, this

is reasonable outcome. A GPS distance over 80 ft is thus likely to indicate a potential error. Less than one percent of the data gathered exceeds this 80 ft criterion.

For the maximum acceleration and deceleration rate criteria, the number of GPS points tapers off until 5 to 7 mph/s for acceleration rate and -6 to -7 mph/s for deceleration rate. These values match reasonably well with those expected from the literature (ITE 1999 and AASHTO 2001).

The impact of HDOP on the GPS data is seen until a value of four to five, with only approximately 1% of the total data points exceeding these values. Thus, for this data set, there is little difference in data loss between the more conservative value of four and more inclusive value of eight. Since the study area is located in a typical open suburban area, the potential GPS data loss resulting from HDOP criterion in this study may be relatively lower than that experienced in a downtown or rural area that has tall buildings or significant roadway canopy.

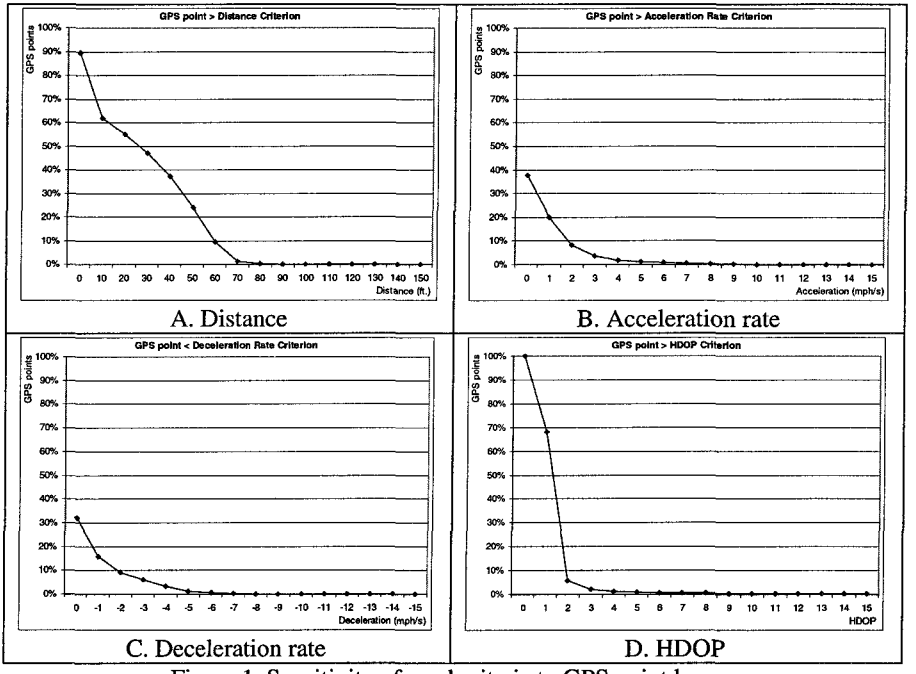


Figure 1. Sensitivity of used criteria to GPS point loss.

Effect of the Criteria on Arterial Performance Measures. To determine arterial performance measures, the error checking process and TRIPTI algorithm were employed. As part of this process the total number of travel runs was reduced by 19 to 696 runs. The 19 runs were eliminated due to driver errors (14 runs) and potential significant GPS errors (five runs). The remaining 696 runs have been employed to observe the effect of varying the distance, HDOP, acceleration, and deceleration criteria on calculated arterial performance measures. Table 1 shows 15 criterion

scenarios. In each scenario only one criterion is varied, and all other criteria are at default values of 80.67 ft, 5.37 mph/s, -7.64 mph/s, and 4 for distance, acceleration rate, deceleration rate, and HDOP, respectively.

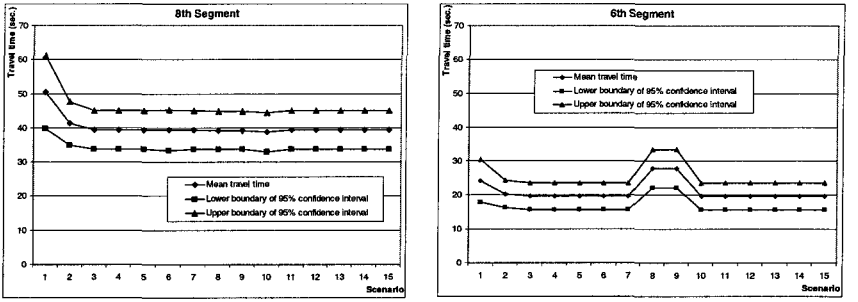
Table 1. 15 Criterion scenarios used in TRIPTI algorithm

Scenario	Criterion	Values
1 ~ 5	Distance (ft)	51.33 / 66.00 / 80.67 / 95.33 / 110.00
6 ~ 9	Acceleration rate (mph/s)	3 / 5 / 10 / 15
10 ~ 12	Deceleration rate (mph/s)	-5 / -10 / -15
13 ~ 15	HDOP	4 / 6 / 8

For this analysis travel time is considered at the link level. Figure 2 gives travel time comparisons for two sample segments along the study corridor. For each of the two segments, the calculated travel time is shown for each of the 15 criterion scenarios. Whenever a GPS point was seen to violate any of the error criteria, the TRIPTI algorithm excludes the data for that vehicle in the average link performance measure.

Scenarios 1 through 5 capture the impact of the varying the distance criterion. At the shorter distance boundaries the average travel time increases as the analysis is likely to be biased to a higher travel time; that is, higher speed vehicles are eliminated as potential errors. For example, in Figure 2A the difference between the mean travel time of scenarios 1 and 2 is approximately 10 seconds. Scenario 1 screens out vehicles with a speed greater than 35 mph (i.e., a 51.33 ft distance between GPS points) as they exit the segment. This is clearly an unrealistic distance criteria as speeds approaching 50 mph were reported by drivers. At the longer distance criteria (scenarios 3, 4, and 5, equate to 55 mph, 65 mph, and 75 mph, respectively), the variation in calculated travel time is insignificant. This highlights that a criteria boundary should not be selected such that the likelihood of eliminating true data is high. The impact on the analysis of incorporating data with potentially small errors is significantly less than that of eliminating good data.

This type of sensitivity analysis is a critical step in analyzing any GPS data sets. For example, for the corridor under study it was determined that the 5.37 mph/s maximum acceleration rate based on available literature was likely biasing the calculated travel times by eliminating numerous vehicles that were stopped or slowed by the traffic signal from the analysis. The sensitivity analysis clearly demonstrates that the 5.37 mph/s value significantly impacted the calculated travel times (Figure 2B). When considering the entire set of GPS point data, acceleration rates exceeding the 5.37 mph/s value are rare, however, closer inspection of the data revealed that the higher accelerations tended to be concentrated at the exit boundaries of the roadway segments; that is, where vehicles are likely to be accelerating from a traffic signal. Thus, the 5.37 mph/s value was introducing a potential bias in the analysis by eliminating vehicles that stopped at the signal. Based on the sensitivity analysis and an inspection of a sample of data points, where higher acceleration rates were seen, the error bound on acceleration was approximately doubled for the corridor analysis. It was also of interest to note that the difference in calculated travel time that resulted from an HDOP of four versus eight was insignificant. Thus, the lost data was minimal when utilizing the more conservative HDOP of four.



A. Typical segment travel time pattern (EB)

B. Criterion dependent segment travel time pattern (EB)

Figure 2. Effect of varying criteria on segment performance measures.

6. CONCLUSIONS

This paper presented a case study on processing GPS-instrumented test vehicle travel run data. Prior to processing the collected data, the research team examined GPS data to detect erroneous data points. Five different criteria were adopted to identify the potential GPS errors: distance between two consecutive points, acceleration rate, deceleration rate, HDOP, and number of satellites.

Data should not be eliminated from an analysis unless it is known to be incorrect, regardless of the impact on analysis results. The challenge becomes knowing when a data point is in error. The presented travel time case study data sensitivity analysis showed that arterial performance measures could be significantly affected by the varying GPS error detecting criteria, with the potential bias introduced by eliminating good data potentially higher than that introduced by limited erroneous data. Final determination of a data point's reliability relies on the analyst's judgment, considering tradeoffs between the likelihood of a data point being erroneous, impact of the potential error on the analysis, and the impact of potentially eliminating accurate data.

REFERENCES

AASHTO (2001). "A policy on design of highway and streets." Washington D.C., 111.

Hunter, M., Wu, S. K., and Kim, H. K. (accepted for publication 2006). "A practical procedure to collect arterial travel time data using GPS instrumented test vehicles." Transportation Research Record, National Academy of Sciences, Washington D.C.

ITE (1999). "Traffic engineering handbook." Washington D.C., 62-65.

Kowama (2005). "GPS-Explained." www.kowoma.de/en/gps/errors.htm (July 2005).

Logsdon, T. (1995). "Understanding the NAVSTAR: GPS, GIS, and IVHS." Van Nostrand Reinhold, New York.

Turner, S.M., Eisele, W. L., Benz, R. J., and Holdener, D. J. (1998). "Travel time data collection handbook." FHWA, Washington D.C.

Instrumented Vehicle Measured Speed Variation and Freeway Traffic Congestion

Joonho Ko[1], Randall Guensler[2], Michael Hunter[3], and Hainan Li[4]

[1,2,3,4]School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Dr. NW. Atlanta, Georgia, 30332-0355
[1]Graduate Research Assistant; PH (404) 385-2376; email: joonho.ko@ce.gatech.edu
[2]Professor; PH (404) 894-0405; email: randall.guensler@ce.gatech.edu
[3]Assistant Professor; PH (404) 385-1243; email: michael.hunter@ce.gatech.edu
[4]Research Engineer II; PH (404) 385-0694; email: hainan.li@ce.gatech.edu

Abstract

This paper investigates the characteristics of speed variation obtained from speed profiles of GPS-equipped instrumented vehicles driven by the participants of the Commute Atlanta Project, an ongoing instrumented vehicle research program deployed in Atlanta, Georgia. In particular, this research effort examines the relationships between acceleration noise and the level of traffic congestion on freeway segments. The results of this study may provide another prospective use of Vehicle Infrastructure Integration (VII) data which are likely to play a significant role in monitoring traffic congestion.

Introduction

An advantage of vehicles instrumented with global positioning system (GPS) devices is that they can provide high resolution speed data on a second-by-second basis. Unlike traditional traffic data collection techniques which only capture average speed, GPS instrumented vehicles allow researchers to capture the microscopic speed variations experienced by individual drivers. This allows researchers to study how the degree of speed variation, together with average speed, is related to traffic congestion, driver comfort, fuel consumption, and emissions.

The instrumented vehicle data utilized in this paper is a subset of the Commute Atlanta data set. The Commute Atlanta project, an instrumented vehicle research program funded by the Federal Highway Administration (FHWA) Value Pricing Program and the Georgia Department of Transportation (GDOT), provides an opportunity for investigating the relationship of speed variation and traffic congestion with a large data set collected in a real-world context. The Commute Atlanta project monitors second-by-second activity of approximately 500 instrumented vehicles using GPS devices, enabling researchers to examine the microscopic traffic flow conditions that each participant experiences.

This paper investigates the characteristics of speed variation observed on four GA400 freeway segments in Atlanta, Georgia. Of particular interest in this paper is the examination of the relationship between acceleration noise and the level of traffic congestion on freeway segments. In this effort acceleration noise is defined as the root mean square of the instrumented vehicle's accelerations and traffic congestion is determined utilizing traffic density obtained from video detection system (VDS) cameras managed by the GDOT Transportation Management Center (TMC). The authors also analyze potential correlations between several speed variation measures.

Measures of Speed Variation

The degree of speed variation can be quantified by several measures such as the standard deviation of speed, the coefficient of variation, total absolute speed difference, and acceleration noise. These measures can be readily computed from the second-by-second speed profiles obtained from GPS-equipped vehicles. The standard deviation of speed is the measure of average deviation from the mean speed, and the coefficient of variation is computed by dividing the standard deviation of speed by the average speed. Total absolute speed difference is computed by dividing the sum of the absolute accelerations by travel distance. Previous research efforts have utilized these measures to quantify the level of traffic congestion as a function of speed fluctuations (Babu and Pattnaik 1997). The underlying assumption of these efforts is that drivers tend to maintain their desired speed, however, as traffic increases the drivers are forced to change their speed to adjust to the surrounding traffic conditions.

Acceleration noise, defined as the root mean square deviation of the vehicle acceleration (i.e. standard deviation), was first proposed to characterize the driver-car-road complex nearly five decades ago (Herman et al. 1959). In practice, acceleration noise has been computed as either the standard deviation (i.e., the deviation around the mean acceleration) or the root mean square of acceleration (i.e., the deviation around an acceleration of zero). When the average acceleration is zero (i.e., the initial and final speeds are the same) these two measures are approximately equal. However, the acceleration noise calculation using standard deviation has a shortcoming in that zero acceleration noise can be obtained when speed changes are made at a constant rate, even though the gradual speed changes occur due to congestion. To avoid this shortcoming, the authors computed acceleration noise from the root mean square of acceleration. Among the speed variation measures, acceleration noise has been most widely studied to describe traffic conditions. For example, an attempt was made to relate acceleration noise to freeway level of service using the energy-momentum concept (Drew et al. 1967), and a more recent study utilized acceleration noise as a traffic congestion measure (Taylor et al. 2000).

Analysis

Data. The data employed in this study consists of second-by-second speed data obtained from GPS-equipped instrumented vehicles and traffic density data obtained from the GDOT TMC. The TMC data were used to capture the macro-level traffic conditions associated with the instrumented vehicle trips. Four GA400 freeway segments (northbound only) in Atlanta, Georgia were selected for data collection as high resolution (aggregated in 20-second bins) of archived traffic data were available for this corridor. The four segments were selected as they represent four different freeway lane and speed limit configurations, as shown in Table 1. These segments are at least one mile apart, and Segment 3 (3 lanes) is located between Segment 2 (4 lanes) and Segment 4 (2 lanes), which indicates a lane reduction from 4 lanes to 3 lanes and then from 3 lanes to 2 lanes occurs. The selected segments do not contain ramp influence areas, which are defined as areas within 1,500 ft of ramp-freeway merge or diverge points (TRB 2000).

A total of 6,707 instrumented vehicles trips were recorded from March to August, 2004. However, only a subset of the vehicle trip data were usable as some instrumented vehicle trips could not be matched with corresponding TMC data due to temporary TMC data outages and other trips were eliminated due to potential data quality issues. To ensure the quality of the GPS data those data points where the number of satellites was less than 4 or the positional dilution of precision (PDOP) was outside the range of 1 and 8 were identified as potentially unreliable (Ogle 2005) and those trips with an unreliable data rate (number of unreliable data points/total number of data points) of 0.5 or greater were discarded from the data set. Since the original speed data were filtered using Kalman filters this approach should be reasonable (Jun et al. 2006). Consequently, 4,889 trips, 73% of the initial instrumented vehicle trips, were used in this analysis. The distribution of vehicles and vehicle trips is shown in Table 1.

Table 1. Characteristics of Selected Freeway Segments and Collected Data Size

Segment	Length (mile)	Speed Limit (mph)	Number of lanes	Number of Trips	Number of Vehicles
1	0.35	55	4	1,444	173
2	0.25	65	4	1,540	161
3	0.22	65	3	930	146
4	0.32	65	2	975	126

Correlation of Speed Variation Measures. The relationships among speed variation measures were analyzed using correlation coefficients, i.e., the measure of the linear association between two variables. From Table 2, it may be shown that speed variation measures are generally correlated with each other, with all correlation coefficients higher than 0.69. The highest correlation, 0.95, is found between total absolute speed difference and the coefficient of variation. Of all the measures acceleration noise has the highest correlation with the standard deviation of speed and the lowest with GPS average speed. This phenomenon is intuitively reasonable since acceleration noise represents the average second-by-second speed variation over a roadway segment, and thus, it is less likely to be strongly influenced by the average speed than the other measures which have stronger associations with average speed. For example, the definition of the coefficient of variation directly utilizes average speed. Acceleration noise seems to provide additional information about a trip that may not be revealed by speed alone.

Table 2. Correlation Coefficients of Speed Measures

	GPS Speed	Acceleration Noise	Standard Deviation of speed	Coefficient of Variation	Total Absolute Speed Difference
GPS Speed	1.00	-0.62	-0.71	-0.78	-0.80
Acceleration Noise		1.00	0.89	0.69	0.76
Standard Deviation of speed			1.00	0.84	0.80
Coefficient of Variation				1.00	0.95
Total Absolute Speed Difference					1.00

Acceleration Noise and Traffic Congestion. Acceleration noise was related to the Highway Capacity Manual (HCM) traffic density-based congestion measure, which is the primary measure for evaluation of level of service (LOS) on freeways (TRB 2000). Figure 1 illustrates scatter plots for each segment with letter marks indicating the average acceleration noise for each HCM LOS A-to-F. From the scatter plots it is seen that acceleration noise, in general, increases as density increases. However, a high degree of variability is found in the plots, with R^2 values for linear models falling within the range of 0.27 (Segment 1) to 0.43 (Segment 3). This variability may be due to various factors associated with driver/vehicle characteristics and localized traffic/roadway conditions such as traveled lane, lane width, vehicle position within a platoon, truck volumes, and so on. Unlike previous research efforts which adopted test car methods utilizing designated drivers/routes and pre-specified driving rules the instrumented vehicle data employed in this study were not collected under controlled conditions. Therefore, the variations found in this study are likely to be greater than those in previous research efforts.

Variability in acceleration noise characteristics among segments is may also be observed in Figure 1. For example, the average acceleration noise of LOS E for Segment 3 is notably larger than on the other segments, implying a relatively high degree of traffic conflicts in this segment. The average values of acceleration noise under LOS E are 0.59, 0.50, 1.42, and 0.7 for Segments 1, 2, 3, and 4, respectively. A potential reason for this situation is that the lane reduction from 3 lanes to 2 lanes occurs 0.27 mile downstream of Segment 3. This lane reduction is likely most detrimental to traffic conditions at capacity, i.e., LOS E. Another example of variability between segments is that Segment 4 has generally higher average acceleration noise under LOS A-to-D ranges than the other segments. This phenomenon is likely due to the lower capacity of Segment 4 (2 lanes), providing relatively lower quality of service than the other segments. These findings further support the potential for acceleration noise to supplement traffic flow quality measures as they suggest acceleration noise has the capability to reflect the impact of roadway characteristics. However, more detailed analysis is required to support this conclusion.

Acceleration noise, instrumented vehicle speed and traffic congestion were also analyzed in unison using confidence regions by assuming that acceleration noise and speed are normally distributed for each LOS range. Figure 2 illustrates 90% confidence regions for each segment and for each LOS range. In general, the confidence ellipses indicate that worse traffic conditions have lower speeds and higher acceleration noise, as expected. In addition, the confidence ellipses imply that acceleration noise is more sensitive to traffic conditions than speed, in particular for LOS A to C ranges, which is supported by the observation that the width (speed) of confidence ellipses changes marginally, but the height (acceleration noise) changes notably when traffic conditions move from LOS A to C. This finding further implies that an application of acceleration noise combined with speed can enhance the capability of probe vehicle system for traffic congestion monitoring.

Another notable feature in Figure 2 is the large variation in traffic flow quality observed between and within density-based LOS ranges. This variability is suggested by the significant overlaps of differing LOS ellipses and the wide range of speed and acceleration noise covered by individual ellipses. This phenomenon is of interest as it suggests that the macroscopic measure density likely does not appropriately represent microscopic traffic flow quality.

Conclusions

The authors investigated the relationship between GPS-instrumented vehicle measured speed variation and traffic congestion. A correlation analysis was performed to examine the relationships among various speed variation measures. It was seen that speed variation measures are generally highly correlated with each other and that acceleration noise has the highest correlation with the standard deviation of speed and the lowest with the average speed. This analysis led to the initial observation that acceleration noise seems to provide additional information about a trip that may not be revealed by speed alone.

Also examined was the relationship between acceleration noise and the level of traffic congestion on freeway segments. Speed profiles obtained from real-world vehicles and TMC data were utilized in this analysis. The results showed that in general, acceleration noise increases as traffic becomes congested. In addition, it was also found that acceleration noise may have the capability to capture the impact of roadway characteristics on traffic flow quality. Finally, it was observed that acceleration noise is more sensitive to traffic microscopic traffic conditions than speed under LOS A-to-C ranges and that significant variations in traffic flow quality seem to exist within each HCM-defined LOS.

These findings suggest that a combined application of average speed with speed variation may provide an improved description of traffic flow quality than the mere use of average speed, which has been the generally accepted approach. The results of this study may also provide another prospective use of Vehicle Infrastructure Integration data which may someday play a significant role in monitoring traffic congestion.

References

Babu, Y., and Pattnaik, S. (1997). "Acceleration Noise and Level of Service of Urban Roads - A Case Study." *Journal of Advanced Transportation*, Vol. 31(No. 3), pp. 325-342.

Drew, D., Dudek, C., and Keese, C. (1967). "Freeway Level of Service as Described by an Energy-Acceleration Noise Model." *Highway Research Record 162*, pp. 30-85.

Herman, R., Montroll, E., Potts, R., and Rothery, R. (1959). "Traffic Dynamics: Analysis of Stability in Car Following." *Operations Research*, Vol. 7, pp. 86-106.

Jun, J., Guensler, R., and Ogle, J. (2006). "Smoothing Methods Designed to Minimize the Impact of GPS Random Error on Travel Distance, Speed, and Acceleration Profile Estimates." Accepted for publication in *Transportation Research Record, Journal of the Transportation Research Board.*

Ogle, J. (2005). "Quantitative Assessment of Driver Speeding Behavior Using Instrumented Vehicles," Georgia Institute of Technology, Atlanta.

Taylor, M. A. P., Woolley, J. E., and Zito, R. (2000). "Integration of the Global Positioning System and Geographical Information Systems for Traffic Congestion Studies." *Transportation Research Part C*, Vol. 8, pp. 257-285.

TRB. (2000). *Highway Capacity Manual.*

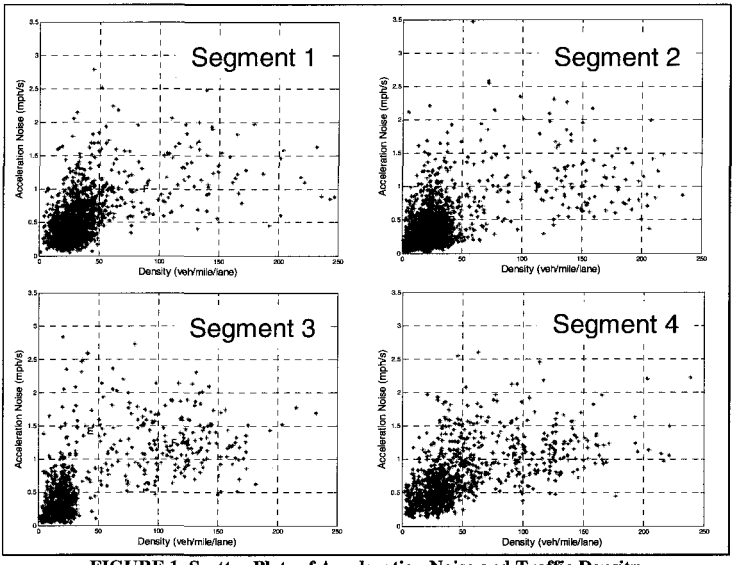


FIGURE 1. Scatter Plots of Acceleration Noise and Traffic Density

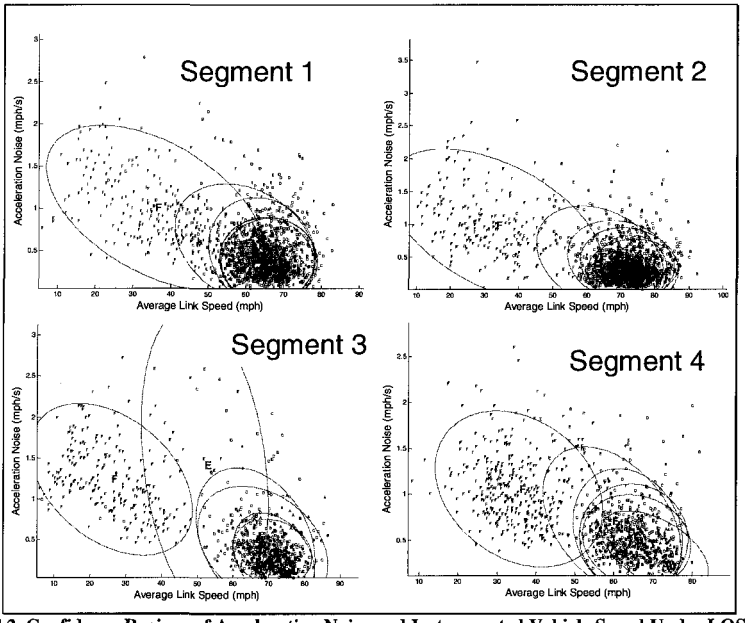


FIGURE 2. Confidence Regions of Acceleration Noise and Instrumented Vehicle Speed Under LOS Ranges

Enhancing Transit Facility Design Using Pedestrian Simulation

Steven P. Scalici[1], P.E., Patrick J. O'Mara[2], P.E., and Raymond Dominguez[3]

[1, 2, 3] STV Inc., 225 Park Avenue South, NY, NY 10003; (212) 777-4400; FAX (212) 529-5237; email: scalicsp@stvinc.com; omarap@stvinc.com; domingr@stvinc.com

Abstract

The use of simulation as an integral element of transit facility design has become increasingly popular with the development of new software in a transportation engineer's toolbox. Beyond the most basic operational assessments, these tools have the ability to demonstrate people movements, frequently traveled paths, and separately, levels of service. Of particular interest to architects and designers, the simulations provide improved and, at times, more realistic results that better clarify and illustrate the "effectiveness" of their work in a dynamic manner.

Furthermore, gone are the days of only relying on spreadsheets and bringing to the public a few rudimentary tables listing analysis results. It is now commonplace for public interest groups to frequently expect consultants to have such a dynamic graphical tool at the ready for them to examine.

STV is involved in a number of projects that require input as to how transportation terminals and stations will function well before design alternatives have been finalized. In fact, the provision for simulation efforts has been a required evaluation criterion for the acceptance of specific conceptual and preliminary design schemes.

Given today's heighten security concerns, accounting for rapid emergency egress is of paramount interest to designers. Pedestrian simulation modeling allows for the relatively quick analysis of multiple pedestrian scenarios including emergency evacuation and normal circulation possibilities, which realistically assigns people to the nearest exits, measures egress times, identifies points of congestion, and requires minimum preparation time.

The capabilities and advantages of pedestrian simulation modeling will be demonstrated based on analyses performed for projects such as the design of the Port Authority of New York and New Jersey's World Trade Center Transportation Hub and the reconstruction of New York City Transit's Cortlandt Street 1 Train station within WTC Site. Both projects are intermodal in nature, and all involved the complex "mixing" of people from many origins and destinations.

The images and videos provided by the simulation, and included with this paper, have proven to serve as powerful and conclusive input to designs that provide for the smooth and efficient movement of people through stations and terminals.

Introduction

Almost two decades ago, the onset of Long Island Rail Road's attempt to bring commuter rail service into Manhattan's Upper East Side brought focus to the agency's desire to provide more-than-adequate operations for its customers as they were to ascend from their deep-tunnel platforms. Of course, earlier strides were made to detail pedestrian movements with Fruin's groundbreaking studies in the 1960s (Fruin 1971, 1987), and later with Benz' time-space methodologies (Benz 1985), but these were static analyses requiring reviewers and the public to look at abstract results as being adequately representative. Such "leaps of faith" were largely dependant on analysts' communicative skills.

Complicating this work was the advent of the National Fire Protection Association's 1982 guidelines (NFPA 1987) to assess transit station pedestrian clearance abilities

from emergency scenarios. These guidelines, too, presented more calculation arcana for engineers to potentially mystify others with.

Then, basic analyses were acceptable to STV clients since these assessments were well conducted and detailed in reports. And now, such work with the basic building blocks of pedestrian analyses is still well received. Yet, tools have become available to make our analyses and products better.

As transit systems were planned and constructed in foreign cities in the late 1990s, so grew the appetites of architects and engineers to examine the effects of their designs on pedestrian flows *before* such plans were constructed. And designers could ill afford to construct inadequately designed terminals and stations. These concerns, in turn, was the genesis to using powerful computers and mathematical advances to create visual simulations of people movements as an overlay onto proposed layouts.

Today, our latest marquee project, one of the country's most visible, involves planning and designing the new transportation hub at New York City's World Trade Center site. Our use of static *and* dynamic tools is a requirement of our client to demonstrate that the agreed upon basis of designs will be achieved.

The paper focuses on three concerns. First, how is a model developed? Second, once functional, what will the analyst see and is it telling a "realistic story"? Finally, what benefits accrue from the work involved to construct such a model?

Model Development

STV uses the STEPS (***S***imulation of ***T***ransient ***E***vacuation and ***P***edestrian movement***S***) (Mott MacDonald 2005) simulation software modeling program that simulates pedestrian movements under both normal and emergency conditions, and provides easily understandable, real-time 3D simulations that graphically presents pedestrian movements, level of service, and usage.

The development of the pedestrian model can be organized into two categories, the building of the physical background elements and the creation of pedestrian movements. The physical background elements of the model consist of floor levels, walls, escalators, stairs, turnstiles, elevators, and train movements in to and out of a station/terminal – in essence, all the physical elements that compose a "working structure." Throughout the modeling process, various physical background items are adjusted in order to realistically simulate pedestrian conditions and control/shape pedestrian maneuvers, such as walking speeds on stairs/escalators, turnstile/doorway processing rates, and pedestrian movement restrictions adjacent to platform edges.

The pedestrian events identify the characteristics of the people who will be modeled in the simulation; specifically, the number of people (i.e., groupings), their origins and destinations, walking speeds, composition (i.e., tourist vs. commuter), assigned routes, physical size, and patience levels.

Model Outputs

Following is a sample of three of the advantageous simulation outputs provided by the STEPS model that could assist an analyst examining a station. First, a real-time simulation output of pedestrians walking and queuing within corridors, ascending/descending VCEs, boarding/alighting trains, and traversing through turnstiles is provided to give the viewer a sense of scale for the project in terms of the size of the facility and the volume of people. For example, a stakeholder can visually understand the difference between 500 versus 5,000 people per hour walking through a 10-foot-wide corridor.

Figure 1 is a sample simulation image of a station platform as alighting train passengers surge toward the vertical circulation elements. The simulation realistically models pedestrian behaviors on a platform as people move to the nearest VCE bank, as compared to walking on the platform, in order to alight a platform sooner.

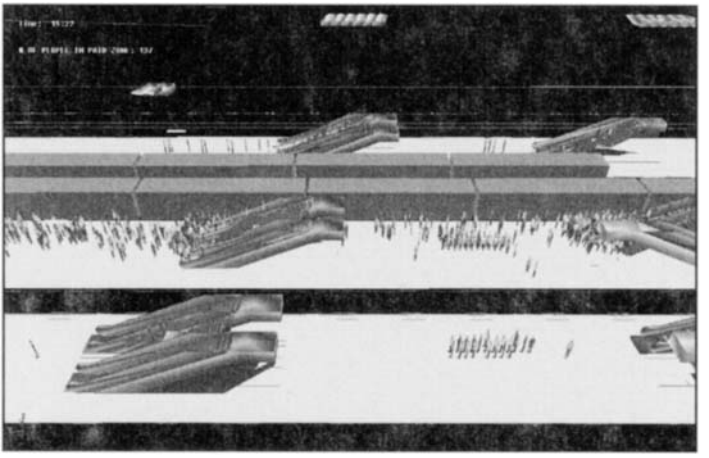


Figure 1: Platform Clearance Conditions during Train Alighting Surge

The sample image presented in Figure 2 focuses on pedestrian queuing near the base of a VCE bank. Again, the model appropriately simulates pedestrian flow behaviors as the passengers tend to walk along the shortest path between origin and destination and, consequently, tend to select a vertical circulation element closest to their origin.

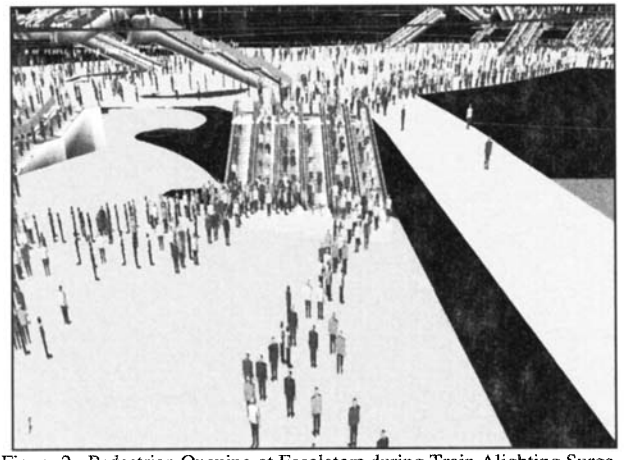


Figure 2: Pedestrian Queuing at Escalators during Train Alighting Surge

Second, the simulation identifies "hot spots," locations that exhibit a high pedestrian density and possibly operate at poor levels of service. These images are similar to Doppler radar pictures, which identify locations of intense precipitation; in STEPS, deep red signifies dense pedestrian activity operating at a poor level of service.

The procedures for estimating and evaluating pedestrian capacity and level of service (LOS) are based on criteria established by Fruin and recommended within the Transportation Research Board's *Transit Capacity and Quality of Service Manual* (TRB 2003). Pedestrian LOS thresholds related to walking are based on the freedom to select desired walking speeds and the ability to bypass slower-moving pedestrians.

The images presented in Figures 3 and 4 highlight the use of the level-of-service simulation for the images presented in Figures 1 and 2, respectively. Notice the deep red colors at the base of the VCE banks, indicating pedestrian queuing and poor LOS conditions. The green shades indicate moderate pedestrian flow at an acceptable LOS condition, and the cool blue colors highlight areas of low pedestrian activity.

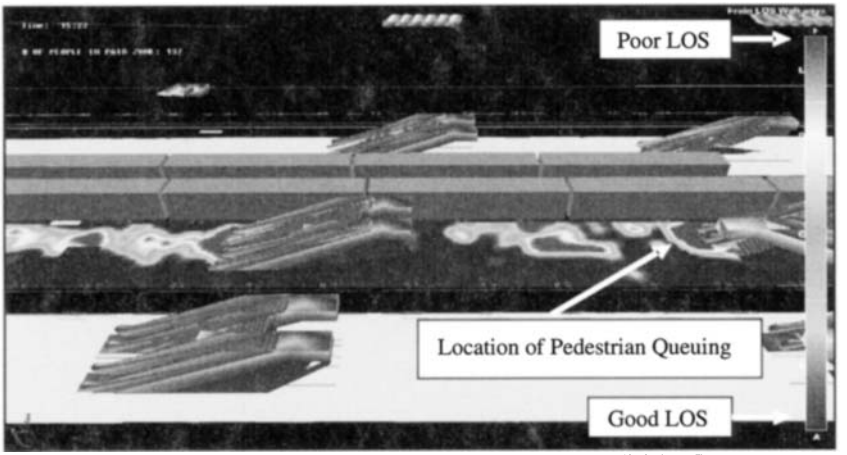


Figure 3: Platform Clearance LOS Conditions during Train Alighting Surge

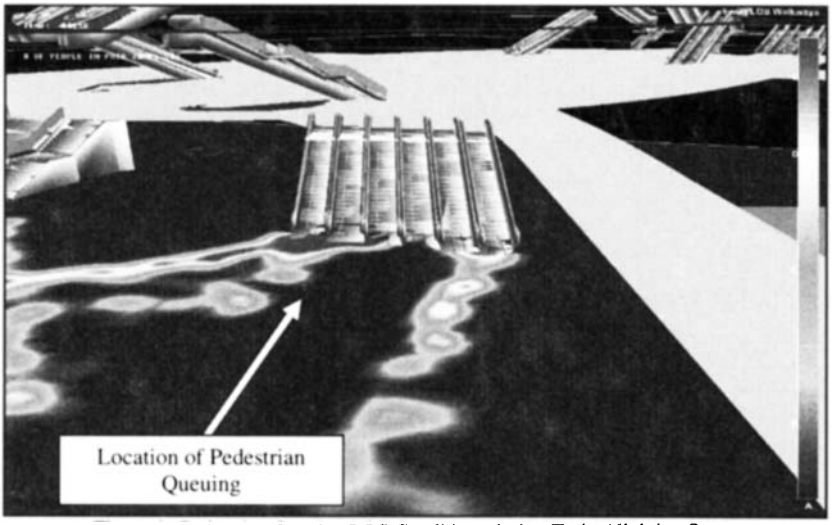


Figure 4: Pedestrian Queuing LOS Conditions during Train Alighting Surge

Third, the simulation can highlight high-usage walking paths through a terminal/station (see Figure 5). The easy identification of desired pedestrian routes can guide designers to properly locate wayfinding signage, VCEs, and retail kiosks so as not to obstruct preferred pedestrian paths. Obstructions within these desire lines create pedestrian turbulence, a reduction in pedestrian walking speeds, and a deterioration in pedestrian LOS.

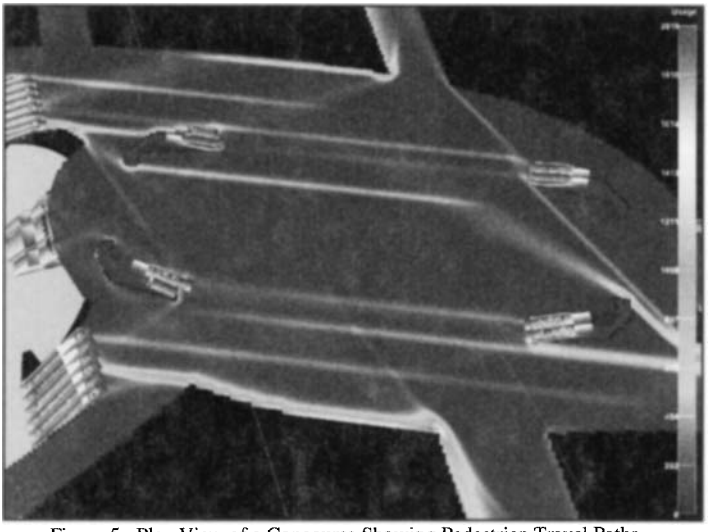


Figure 5: Plan View of a Concourse Showing Pedestrian Travel Paths

Here is where the value of dynamic pedestrian modeling bears fruit in that as different designs are created, we can see their effectiveness relative to each other and thus help winnow down multiple design choices to a select few.

In addition to the simulation outputs, the model can provide statistical, spreadsheet-type outputs, such as the number of people in the model per time increment, number of people in a particular area (i.e., pedestrian occupancy/density), pedestrian egress times, and pedestrian flow volumes at specific checkpoints.

Model Benefits

Several pedestrian simulation benefits have already been highlighted, but more detailed discussions of specific model advantages worth noting follow. One valued feature of pedestrian simulation is that transit-surge effects can be seen, rather than simply read as shown right in the sample output table listing waiting times and person densities over time to pass through individual turnstiles. Such a table is informative to the analyst for comparison with independent static analyses, but can be visually complicated to the public.

Time Interval	Density at Turnstile Entrance (m^2/person)	Exit Wait Time at Specific Turnstiles (seconds)					
		1	2	3	4	5	6
1:25	10.0	0.3	0.5	1.1	0.0	0.5	0.1
1:30	13.3	0.3	0.6	1.0	0.3	0.4	0.1
1:35	10.0	0.2	0.7	1.4	0.8	0.4	0.1
1:40	10.0	0.3	0.6	1.2	1.0	0.4	0.1
1:45	13.3	0.4	0.7	1.1	0.8	0.3	0.2
1:50	13.3	0.4	0.7	0.8	0.9	0.3	0.2
1:55	5.7			0.8	0.8	0.3	0.2

More information and tools are needed for our work because standard planning procedures for calculating the number of turnstiles needed to process passengers at a station are based on average peak period passenger arrival rates. Passengers truly exit a station in groups based on the train arrivals. A dynamic simulation model allows users to see the effects of transit surges at the turnstile array, measure the maximum queue length, and calculate passenger egress times, certainly more comprehendible than such a table shown above. This feature becomes increasingly more useful at terminal stations where the transit surge volume could be highest and at stations with several transit lines where multiple transit surges may converge.

The model has been used to demonstrate the influence of various design alternatives on pedestrian movements and flows. Alternative station support-column configurations have been examined to evaluate their pedestrian impedance. Similarly, a convex versus concave configuration of a set of doors has been modeled to determine the pedestrian door-use distribution across an entry/exit.

Given today's heighten security concerns, accounting for rapid emergency egress during chance events such as escalator failures is of paramount interest to designers. Once a base model has been created, relatively quick analysis of multiple pedestrian scenarios including emergency evacuation and normal circulation possibilities can be examined. Conditions such as "downed" escalators or closed/untenable exits can be developed to measure egress times and identify congestion.

Conclusions

The use of simulation as an integral element of transit facility design has become increasingly popular with the development of new software in the transportation engineer's toolbox. Beyond the most basic operational assessments, these tools have the ability to demonstrate people movements, frequently traveled paths, and separately, levels of service. Of particular interest to architects and designers, the simulations provide improved and, at times, more realistic results that better clarify and illustrate the "effectiveness" of their work in a dynamic manner. That is, know where bottlenecks could happen during the virtual assessment, and modify plans to prevent such design shortcomings.

The sampling of the benefits and features of pedestrian simulation modeling highlighted in this paper provide a glimpse into the resources and capabilities of using this tool to serve as powerful and conclusive input to designs that provide for the smooth and efficient movement of people through stations and terminals.

References

Benz, Gregory (1985). *Pedestrian Time-Space Concept: A New Approach to the Planning and Design of Pedestrian Facilities*, New York, NY

Fruin, John, J. (1971, 1987). *Pedestrian Planning and Design, Revised Edition*, Elevator World, Inc. Mobile, AL.

NFPA 130 Standard for Fixed Guideway Transit and Passenger Rail Systems. National Fire Protection Association, Quincy, MA

STEPS software version 2.1 was designed by Mott MacDonald, and is used under a license agreement dated May 18, 2005.

Transportation Research Board (2003). *Transit Capacity and Quality of Service Manual*, 2nd Edition, TRB, Washington, DC

A Wireless Mesh Network for Real-Time Vehicle Guidance

Alan Wolff[1] and C. C. Lee[1]

[1]Department of EECS, Northwestern University, Evanston, IL 60208; PH: (847) 491-4443, (847)491-7375; emails: awolff@northwestern.edu, cclee@ece.northwestern.edu

Abstract- We present a large-scale vehicle routing architecture and the associated protocols that aim at delivering real time vehicle guidance. In our architecture, communication functions and vehicle route processing are separated into distinct subsystems, where the former deals with wireless communication through a roadside mesh-like network and the latter involves a design framework for low-complexity real-time vehicle routing table computation. We demonstrate that the bandwidth requirement is modest, the processing requirement within the vehicle is minimal, and vehicle routing complexity can be drastically reduced.

INTRODUCTION

The objective of a real-time vehicle route guidance system is to provide each driver with optimal real-time turn-by-turn directions based on system-wide traffic data including traffic loads and travel times [13]. Loop detectors, probe vehicles, license plate matching, and combinations of different techniques can be used to collect roadway travel times [10]. For real-time vehicle guidance to be achieved it is necessary to efficiently transmit and process the vast amount of system-wide data, however it is collected, to produce optimum vehicle routing tables. To address the communication challenges, we present a solution that involves a dense roadside mesh-like sensor network and a hierarchical routing protocol for collecting real-time traffic data from block-level segments of the road system. Routing tables are distributed to active vehicles in a manner that requires minimal sensor node complexity and minimal onboard processing. To address the computing challenges, we present a scalable multi-tier route computation approach that involves hierarchical addressing for sensor nodes, associates vehicle locations and destinations with sensor locations and addresses, and accordingly distinguishes "local" routing from "non-local" routing, and uses route aggregation techniques for non-local vehicle routing to keep route computational complexity at a manageable level, regardless of system size.

A TWO-SUBSYSTEM ARCHITECTURE

The main objectives and key features of our system are: low-cost sensor nodes, minimal onboard vehicle functions, and low-complexity vehicle route computation. To achieve these objectives, we present a system architecture that separates communication functions from vehicle route computation functions. As shown in Figure 1, the wireless sensor network constitutes the communication subsystem (CSS) and is responsible for collecting and forwarding vehicle data to the vehicle routing subsystem (VRS) and for distributing optimum routing information to the vehicles from the VRS. The VRS is responsible for processing the vehicle data and generating optimum routing tables *for every sensor node in the system.*

The CSS is made up of a large number of low-power densely populated roadside sensor nodes (SNs) in a Multi-hop Cellular Network [9], where base stations may be reached by mobiles in multiple hops. This approach is in concert with the advent of low-power multi-hop wireless mesh networks which are expected to extend the reach of many existing communication networks, offer low cost infrastructure, add redundancy, improve reliability and can operate

over rough or difficult terrain that may contain obstructions [1]. The potential of mesh networks has led to their inclusion in key wireless communication standards including WiMAX [6] and HIPERLAN/2 [4]. Viable wireless mesh network solutions have also been explored recently for applications in intelligent transportation systems (ITS) [2].

Base stations (BSs) serve as the interface between the CSS and VRS, responsible for relaying traffic data to the VRS and for broadcasting vehicle routing tables to active vehicles. The SNs are self-organized into a multiple-layer hierarchical routing network [8] which allows data packets carrying vehicle information to be routed efficiently to a BS which is the root of a routing tree. BSs are strategically deployed throughout the network, each capable of broadcasting messages to SNs and vehicles within its broadcast range. Like cellular systems, the ranges of neighboring BSs must overlap in order to ensure complete coverage. The broadcast range of a BS is called a "cell" and the union of the cells of a BS and all its neighboring BSs a "cell cluster" associated with this BS. To illustrate, shown in Figure 2 is a cell cluster associated with BS X. A BS supervises the self-organizing process within its routing tree through which SNs forward data messages upstream to this BS following the tree.

The routing tree owned by a BS can in principle cover the entire network [6] thus allowing every SN in the network to send data to any other BS. However, since we are dealing with a huge network SN simplicity would have to be compromised if every routing tree were allowed to cover the entire network. We thus limit the reach of each routing tree to the cell cluster associated with the root of tree, which is readily achieved by a modified the underlying self-organizing protocol. As a result, a SN can belong to only a limited number of routing trees, dictated by cluster size. Each BS maintains its own routing tree, coordinates medium access, initiates periodic maintenance tasks for its tree, and broadcasts real-time vehicle routing tables associated with every SN within its cell. As a result, the BSs free the SNs from routing related processing burdens and as such, the SNs require little processing power and memory.

A SN has a very small transmission range, and is responsible for only the following functions: 1. continual broadcast of a presence signal that polls vehicles passing by, 2. receipt of data packets from vehicles responding to its poll, 3. insertion of its ID and a time stamp in each vehicle packet it receives and 4. forwarding all packets including those from other SNs to its parent node with respect to the selected BS. While passing by a SN, a vehicle transmits a small packet containing its vehicle ID and its intended destination ID which is encoded geographically (to be discussed later). Upon reception of a vehicle packet, the SN inserts its own ID and a timestamp in the packet and forwards it to its parent node – a neighboring SN, en route to a BS. The CSS supports both up-stream and down-stream communications. Up-stream communication involves a single-hop vehicle-to-SN protocol, a multi-hop SN-to-BS protocol, and a high-bandwidth BS-to-VRS protocol. The first two have been sketched above. With regard to the third, we shall assume that BSs are capable of communicating with the VRS using a separate "root-level" communication infrastructure in which BSs have little constraint in power and bandwidth. Through this infrastructure, the vehicle data are sent to the VRS with little delay. Downstream communication involves distributing optimum vehicle routing tables generated by the VRS to the BSs via the root-level infrastructure and broadcasting the relevant vehicle routing tables directly to the vehicles. A control channel, dubbed Channel R is used by the BSs for broadcasting vehicle routing tables to vehicles within their respective cells.

Through the CSS, the VRS obtains data from vehicles and accordingly calculates link weights within the roadway network. Then based on a distance metric incorporating travel times and optionally other criteria such as system load balancing and special policies assigned to roads or specified by drivers, routing tables are generated for each selected source location in the system. The number of entries in a routing table equals the number of selected destination locations

from a source location and each entry specifies an "optimum next SN" that helps the vehicle decide its action at the upcoming intersection. Computational challenges for this arise from the large size of the sensor network, as reflected by the number of SNs deployed, which is in the order of the number of edges of a topological graph where the vertices are the sensors and the edges are the street/highway paths between adjacent sensors. Without specifying the root-level communication infrastructure and protocol this paper leaves open the VRS architecture. In the following, we discuss only the computational complexity that the VRS as a whole must deal with and a means of reducing it.

Based on the system architecture presented, the required functions for a vehicle are minimal: 1. It constantly listens to Channel S for the presence of roadside sensors; 2. It transmits a short packet containing its ID and its intended destination to an SN upon hearing the broadcast from the latter; 3. It listens to the broadcast on Channel R, selects the routing table associated with the last SN it detected, and from the table the entry that captures its intended destination.

BANDWIDTH REQUIREMENTS

To achieve real-time dynamic vehicle routing, a vehicle sending a packet to the CSS must receive the newest vehicle routing table containing its intended destination before it reaches the next decision point – the upcoming intersection. Let the roundtrip time from when a vehicle emits its packet P to a roadside sensor i until a routing table for sensor i, which incorporates P in its computation, is received by the vehicle may be denoted T_r. This value must be somewhat smaller than T, the time it takes the vehicle to travel from its current location to the next intersection. T_r is comprised of three components: the up-stream communications delay, t_u, the VRS processing delay of updating the routing tables, t_p, and the downstream communication delay, t_d. The real-time objective is met if $T > T_r = t_u + t_{p+} t_d$ where t_d is negligible due to the high bandwidth assumption on the root-level communication. It is thus of interest to assess the bandwidth requirement for the sensor network to contain t_u. In the next section, we discuss complexity issues related to vehicle route computation and present means of reducing the complexities in order to contain t_p. It is readily shown that multi-hop transmission and propagation delays are negligible compared to T and therefore if queuing at the sensors can be constrained, upstream communication delay t_u will not pose a problem for the real-time objective. Since queuing can be contained by having sufficient bandwidth for sensors to forward messages, in the following we analyze a bandwidth requirement for the sensor network.

Macroscopic traffic flow theory relates traffic flow, traffic density and vehicle velocity. The fundamental relationship [5] is $Q=KV$, where Q = traffic flow, K = traffic density and V = average vehicle velocity. From this, traffic density and velocity are inversely proportional and an upper limit on traffic flow and thus vehicle arrival rate from the SN's perspective can be inferred. The Federal Highway Administration has pegged the maximum design capacity per lane of roadway in vehicles per hour as 2,200 [12]. This results in an upper limit of $0.61L$ vehicles/sec on a roadway, where L is the number of lanes. For our model where SNs receive data from vehicles as they pass by, the data rate received by an SN is then proportional to Q. If the length of a data packet sent by a vehicle and forwarded by the SN, including overheads, is P bits, then the maximum bit rate at each sensor is $R = (0.61L) \times P$ in bits per second. This suggests that wireless bandwidth requirements for the sensor network are not burdensome.

The multiple hops needed for vehicle data to get to a BS raise the bandwidth requirement for sensor-to-BS communications. At any particular node, there are two kinds of traffic: traffic that comes directly from vehicles passing the node and traffic that is routed to the node. As stated previously, the maximum rate of data traffic coming from passing vehicles to a sensor is $R_{direct} = (0.61L) \times P$. With the depth of a routing tree of H, the maximum rate at which routed data

traffic flows on the multi-hop network is $R_{routed} = (0.61L) \times H \times P$. This then gives a worst case scenario of approximately $R_{sensor} = (H + 1) \times (0.61L) \times P$ for the multi-hop network. For this scenario, with 4 lanes, P=2,000 bit packets (including sensor ID and timestamps), and a tree depth of 10, we will have R_{sensor} = 53.7 Kbps. This strongly suggests that the communication capacity of such a data network is not a major bottleneck of the system.

MANAGING ROUTING COMPLEXITY

Once the communication delays are contained, the major factor that determines whether the real-time objective can be met is the time it takes the VRS to process vehicle data and to accordingly produce updated routing tables for all the SNs. Basically, the VRS uses vehicle data to continually update link weights between every SN and each of its neighboring SNs, which are then used to determine optimum routes for every source-destination pair of interest. A conventional approach to estimating link weights is based on aggregate flow through an intersection [11]. However, this estimate may not be accurate due to traffic exiting or entering the roadway segment between the two end points. In our approach, based on vehicle ID, sensor ID, and sensor timestamps contained in vehicle data, the VRS may update the travel time between SNs with great accuracy and timeliness by comparing two packets from the same vehicle passing by successive SNs. Note that the VRS must generate a routing table for every SN. Let n denote the total number of SNs in the system, which also equals the number of vertices on the topological graph, and let |E| be the number of edges in the graph. Then if the Dijkstra's algorithm is used, the complexity for producing each routing table (i.e., the per-SN complexity) will be $g(n) = O(|E| \log n)$ [3] which can also be expressed as $g(n) = O(n \log n)$ since $|E| = O(n)$. For a grid street topology, $|E| \sim 3n$ and as a result, the overall computational complexity for the VRS to produce n routing tables is $f(n) = ng(n) = O(n^2 \log n)$. Since n is a huge number, such complexity may pose a problem. Furthermore, distributing large routing tables requires large bandwidth in VRS-to-BS communication and more importantly, huge broadcast bandwidth in BS-to-vehicle communication. In the following, we discuss a means of containing f(n) by shrinking n.

In practice, at a given time the VRS needs to provide optimum vehicle routing information to only active vehicles, i.e., vehicles from which the VRS just received packets. Therefore, we may substantially reduce the processing complexity by exploiting an "on-demand" strategy that: 1. computes new routing tables for only those source locations (SNs) from which the VRS just received vehicle data and 2. includes in a routing table optimum next hops (next SNs) for only those destinations indicated in the vehicle data. In this manner, the total number of source-destination pairs that have to be dealt with by the VRS at any given time is no larger than the number of active vehicles in the system at that time. While this strategy will greatly reduce the number and size of routing tables within a given time interval, it may not yield as much savings on the computational complexity involved in updating them. To circumvent this problem, we first distinguish two types of vehicle routing: intra-cluster and inter-cluster. The former is for vehicles whose destination is within the cell cluster of its current location, and the latter is for vehicles heading for distant destinations beyond their current cluster boundaries. Since a BS owns a (data packet) routing tree that spans a cell cluster, it receives all the vehicle data from SN locations throughout its cell cluster. As such, a BS has all the roadway link weights within the cell cluster and can therefore compute all the possible intra-cluster routes. Thus, a BS is responsible for computing and broadcasting the routing tables for all the SNs within its cell and each of these tables covers demanded intra-cluster destinations. Consequently, a BS involves a computational complexity for intra-cluster routing of $f_0(n) = O(m_0 n_0 \log n_0)$, where n_0 = number of SNs in a cell cluster and m_0 = number of SNs in a cell.

To route a vehicle to a destination beyond the cell cluster associated with its current location, inter-cluster routing is used, with complexity of inter-cluster route computation being contained by a destination address aggregation approach. The basic idea is to aggregate the remote destinations (from a given SN) based on an M-tier hierarchical network partitioning and addressing scheme. Specifically, the network will be geographically partitioned into K equal-size tier-M zones and each Tier-X zone is sub-partitioned into K equal-size Tier-(X-1) zones, where $1<X\leq M$. Figure 3 depicts such a hierarchical partition with K=9, where a Tier-1 zone is smaller than the size of a cell - the broadcast range of a BS - such that a Tier-1 zone will never contain more than one BS. For inter-cluster vehicle routing from a given SN location, each destination (outside the cell cluster of the source SN) is absorbed by a "Tier-Y super node" - a BS representing a Tier-Y zone, where the value of Y reflects degree of aggregation and is determined by either L_2 or L_1-distance between source and destination. This drastically reduces the number of possible destinations that a routing table has include and as a result, the per-SN computational complexity for inter-cluster routing is $g_1(n) = O(n_1 \log n_1)$ where $n_1 = \min\{(K-1)M$, number of BSs (cells) in the system$\}$. Note that the hierarchical zone structure is fixed and is independent of the fuzzy wireless cell and cell cluster boundaries. The routing table broadcasted by the BS on behalf of an SN contains both intra-cluster and inter-cluster destinations where the former consists of complete addresses and the latter aggregated addresses. The inter-cluster entries essentially capture BS-level shortest distances and in that sense they are sub-optimal. We note that selecting the representative BS within an upper-tier zone and link weight between zones remain open issues and are among the future research tasks. Basically, leveraging the high-bandwidth root-level communication infrastructure, a dynamic routing solution can be employed for root-level route computation and maintenance.

REFERENCES

1. D. Beyer, "Fundamental characteristics and benefits of wireless routing ("mesh") networks," in *Proc. Int'l Tech. Symp. Wireless Comm. Asso.*, San Jose, CA, January, 2002.
2. R. Bruno, M. Conti, E. Gregori, "Mesh Networks: Commodity Multihop Ad Hoc Networks," *IEEE Communications Magazine*, March, 2005, pp 123-131.
3. T. Cormen, C. Leiserson, R. Rivest, C. Stein, *Introduction to Algorithms*, MIT Press, 2001.
4. N. Esseling, H. S. Vandra, and B. Walke, "A Forwarding Concept for HiperLAN/2," Proc. Euro. Wireless 2000, Sept, 2000, Dresden, Germany, pp 13-18.
5. D.L. Gerlough, M. Huber, "Traffic Flow Theory: A Monograph TRB Special Report 165" Transportation Research Board, Washington, DC, 1975.
6. IEEE Standard 802.16-2004 "IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems," 2004.
7. A. Iwata, C.-C. Chiang, G. Pei, M. Gerla, and T.-W. Chen, "Scalable Routing Strategies for Ad Hoc Wireless Networks" *IEEE Journal on Selected Areas in Communications, Special Issue on Ad-Hoc Networks*, Aug. 1999, pp.1369-79. And references therein.
8. H. Lee, C. C. Lee, "Layer Scheduling in a Multiple-Layer Self-Organizing Wireless Network," *Proc. of GLOBECOM'02*, Taipei, Taiwan, Nov, 2002.
9. M. Lott, M. Weckerle, W. Zirwas, H. Li, E. Schultz, "Hierarchical Cellular Multihop Networks," EPMCC, March, 2003. See references therein to other MHCNs.
10. Ramachandran, R. Sun, C., Arr, G. Ritchie, S.G. Intelligent Multi-detector fusion for Vehicle Re-identification. *82nd Ann. Mtg. Transportation Research Board*, 2003.
11. Roess, Prassas, McShane, *Traffic Engineering*; 3rd Edition, Pearson Prentice Hall: Upper Saddle River, NJ, 2004.
12. Transportation Research Board, *Highway Capacity Manual 2000*; 4th Edition, National Research Council, Washington, DC, 2000.
13. U.S. Department of Transportation, *Theory of Operations, National Intelligent Transportation Systems Architecture*, Washington, DC, 1998.

FIGURES

Figure 1 A sensor network based vehicle routing system consisting of a large-scale sensor network and communication subsystem (CSS) and a vehicle routing subsystem (VRS). The sensor network provides the communication infrastructure and the VRS does all the computations related to optimum vehicle routing.

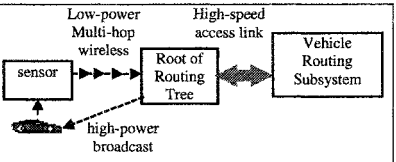


Figure 2 A cell cluster centered at BS X. The range of each BS (triangles in diagram) covers areas in respective circles above. BSs broadcast routing tables to passing vehicles and maintain the routing trees for sensors. In the above figure is a cluster with seven cells, similar to a cellular system.

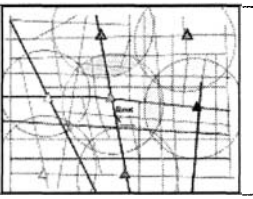


Figure 3 A geographically based hierarchical network partitioning and addressing. A 3-tiered zone hierarchy is shown. The addresses of the tier-1 zones in the E.i tier-2 zone are E.i.a, E.i.b, E.i.c, E.i.d, E.i.e, E.i.f, E.i.g, E.i.h, and E.i.i.

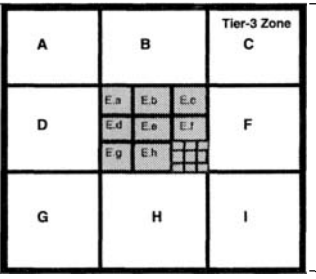


Figure 4. A inter-cluster routing table example

Tier M		Tier M - 1		Tier 2		Tier 1	
Dest	Next BS	Dest	Next BS	Dest	Next BS	Dest	Next BS
B.x.x.x	u	A.a.x.x	x	A.e.a.x	X	A.e.i.a	x
C.x.x.x	u	A.b.x.x	y	A.e.b.x	Y	A.e.i.b	x
D.x.x.x	v	A.c.x.x	u	A.e.c.x	y	A.e.i.c	y
E.x.x.x	w	A.d.x.x	z	A.e.d.x	z	A.e.i.d	x
F.x.x.x	u	A.f.x.x	u	A.e.e.x	x	A.e.i.e	x
G.x.x.x	v	A.g.x.x	z	A.e.f.x	y	A.e.i.f	y
H.x.x.x	w	A.h.x.x	w	A.e.g.x	z	A.e.i.g	z
I.x.x.x	w	A.i.x.x	w	A.e.h.x	z	A.e.i.h	z

Model of a Self-Organizing Traffic Management Hazard Response System

R. Kicinger[1] and M. Bronzini[2]

[1]Department of Civil, Environmental and Infrastructure Engineering, George Mason University, 4400 University Drive, MS 4A6, Fairfax, VA, 22030; PH (703) 993-1658; FAX (703) 993-1521; email: rkicinge@gmu.edu
[2]Department of Civil, Environmental and Infrastructure Engineering, George Mason University, 4400 University Drive, MS 4A6, Fairfax, VA, 22030; PH (703) 993-1504; FAX (703) 993-1521; email: mbronzin@gmu.edu

Abstract

The terrorist attacks of September 11, 2001 and afterwards have caused renewed interest in developing effective policies and strategies for evacuating densely populated areas. The large-scale evacuations caused by Hurricane Katrina and other recent hurricane events have reinforced this need. Unfortunately, the current analytical tools for dealing with such evacuations are sorely lacking, in both theory and practice. The model and its computational implementation presented in this paper attempt to close this gap and make significant progress in traffic management and hazard response systems.

The overall goal of this research is to develop a fundamental understanding of the evolutionary and emergent behavior of transportation systems that are operating under emergency evacuation conditions. Initial ideas on building conceptual models of evacuation scenarios utilizing cellular automata, evolutionary computation, and advanced traffic simulators were presented in the authors' previous paper. This paper describes computational implementations of proposed conceptual models. It also discusses preliminary results of several computational experiments in which the models were used to determine robust configurations of traffic control systems operating under emergency conditions. In these experiments, optimal evacuation strategies were sought for vehicles located within a representative urban area affected by various types of terrorist attacks.

Introduction

Transportation systems that are operating under emergency evacuation conditions exhibit behavior that is substantially different from typical modes of transportation systems' operations. Hence, novel approaches are required to model, analyze, and optimize the performance of such systems. Three hypotheses drive this research:

1. A typical transportation network can be understood as a complex system with many entities and actors, all pursuing their own somewhat limited objectives and acting with variable and limited information inputs. All of the actors in the system often make decisions with little or no knowledge of the impacts of their decision on the performance of the overall system.
2. The emergent behavior of the system and its subsystems is of great interest for finding effective technology and policy approaches to improving performance.
3. Systems operating in crisis mode exhibit self-organizing behavior, so finding optimal operational strategies involves understanding and capitalizing on this attribute.

Agent-based models, with their use of individual agents following localized decision rules for interacting with other agents and the environment, appear to offer a powerful approach to capturing the many interactions in transportation systems. On the other hand, cellular automata (CAs) have been successfully used to model self-organization phenomena in various biological, chemical, and man-made systems. Hence, they offer an enormous potential for developing a new class of models of emergent behavior for traffic management and hazard response systems.

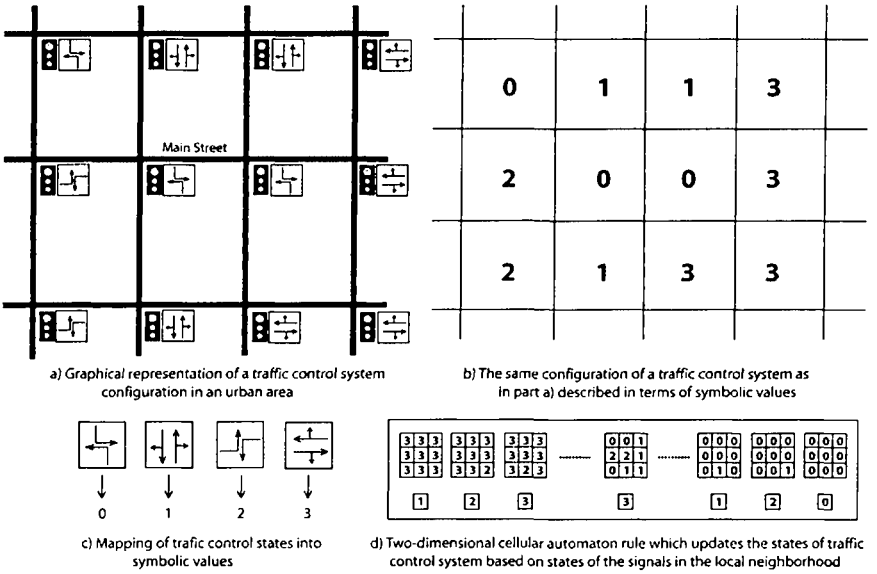
When such models are found, they can be subsequently used by transportation engineers to create effective operational strategies for robust hazard response systems. Evolutionary algorithms (EAs) can be used to identify optimal CA models in the vast solution spaces, i.e., they act as search and optimization mechanisms. They can also be employed to adapt and fine-tune solutions to specific types of transportation systems, geographical locations, etc.

Conceptual Model of a Self-Organizing Traffic Management Hazard Response System

The conceptual model of a self-organizing traffic management system was proposed in (Bronzini and Kicinger 2006). In this section, a brief description of the model and its assumptions is provided in order to facilitate the discussion of its implementation presented in the remainder of the paper.

In the model of a self-organizing traffic management hazard response system, two-dimensional cellular automata (2D CAs) (Wolfram 1994) are used to model emergent behavior of transportation systems operating under emergency conditions. Figure 1 shows a simple example of the application of this model in which the configuration of states of traffic signals within several city blocks is represented as a 2D CA. The spatial (north-south-east-west) relationships among elements of the traffic control

system and their relative distances (1 block away, 2 blocks away, etc.), are represented explicitly as by cells of a 2D CA (see Figure 1b).



a) Graphical representation of a traffic control system configuration in an urban area

b) The same configuration of a traffic control system as in part a) described in terms of symbolic values

c) Mapping of trafic control states into symbolic values

d) Two-dimensional cellular automaton rule which updates the states of traffic control system based on states of the signals in the local neighborhood

Figure 1. Simple model of an urban traffic control system based on two-dimensional cellular automata.

Individual states of traffic signals can be easily represented by discrete values of a cellular automaton as shown in Figure 1c. Here, four possible states of traffic signals are encoded as integer values 0-3. Finally, a 2D CA rule presented in Figure 1d is used to update the state of each traffic signal in this network based on the states (values) of traffic signals located in its local neighborhood. The value shown in the bottom row of Figure 1d defines the state of this traffic signal at the next phase. The advantage of such representation is that range and directions of local interactions among elements of an urban traffic control system can be explicitly modeled.

In the course of traffic simulation, the configurations of traffic signals are updated multiple times using multiple iterations of a 2D CA rule. Each iteration of the rule uniquely defines the states of all traffic signals at a specific simulation time period.

Because one of the goals of this research is to identify optimal evacuation strategies for a given urban area, cellular automata representations of traffic control systems are optimized by evolutionary algorithms. The goal is to find optimal initial configurations of states of a traffic control system as well as optimal 2D CA rules which update these states during an evacuation process, so that the number of vehicles that leave the affected area is maximized and their travel time is minimized. The actual implementation of this simple model is presented in the following section.

Model Implementation

The conceptual model of a self-organizing traffic management system proposed in (Bronzini and Kicinger 2006) and briefly described in the previous section was implemented in a computer system called Emergent Designer (Kicinger et al. 2005). It is an integrated research and design support tool which utilizes models of complex systems, including cellular automata and evolutionary algorithms, to represent engineering systems and the related design processes.

In order to verify the usefulness of the proposed model, a new application domain, i.e., traffic simulation, was added to Emergent Designer as shown in Figure 2. A cellular automata representation of a simple traffic network in an urban area was developed and combined with an evolutionary algorithm (for more detailed description see (Bronzini and Kicinger 2006)). Solutions generated by this evolutionary design process consisted of initial configurations of traffic signal states and 2D CA rules which updated the configurations of traffic signals at multiple phases of the traffic simulation process. These symbolic solutions were first translated into a traffic simulation file (a TRF file in this case) and subsequently executed by CORSIM, an advanced mezoscopic traffic simulator which was integrated with Emergent Designer. CORSIM simulated the movement of vehicles in the affected urban area and provided two measures of effectiveness, i.e., the number of vehicles which left the affected area during the course of simulation as well as the total travel time of vehicles in the affected area, through its output file. These measures provided feedback for an evolutionary algorithm and defined the quality, or fitness, of solutions, produced during the evolutionary design process.

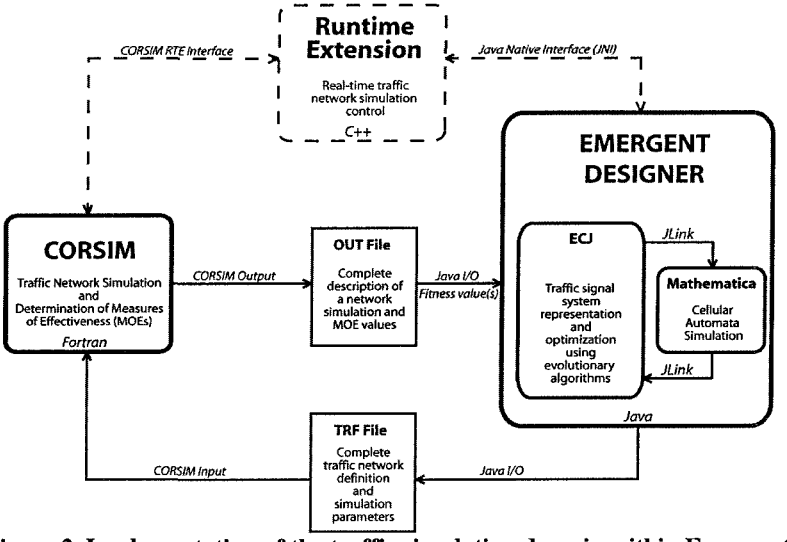


Figure 2. Implementation of the traffic simulation domain within Emergent Designer.

Due to limitations of the current version of CORSIM (TSIS 5.1), the entire process of updating traffic signal states at various phases of simulation was encoded within TRF files in terms of multiple time periods. In the future, this approach will be replaced by direct communication with CORSIM Runtime Extension (RTE) as shown in the upper part of Figure 2.

Initial Results

Initial computational experiments with the proposed self-organizing traffic management hazard response system focused on determining the feasibility of the model and comparisons of its efficiency with other traffic control strategies. In order to achieve these goals, a simple transportation network of an urban area spanning several city blocks of, e.g., Washington, DC, was developed as shown in Figure 3. Several simulation runs were performed to identify effective evacuation strategies for this urban area. The progress of evolutionary design processes was determined by calculating the percentage improvements in the number of vehicles which left the affected area during the course of simulation as well as in minimizing the total travel time of vehicles located in the affected area.

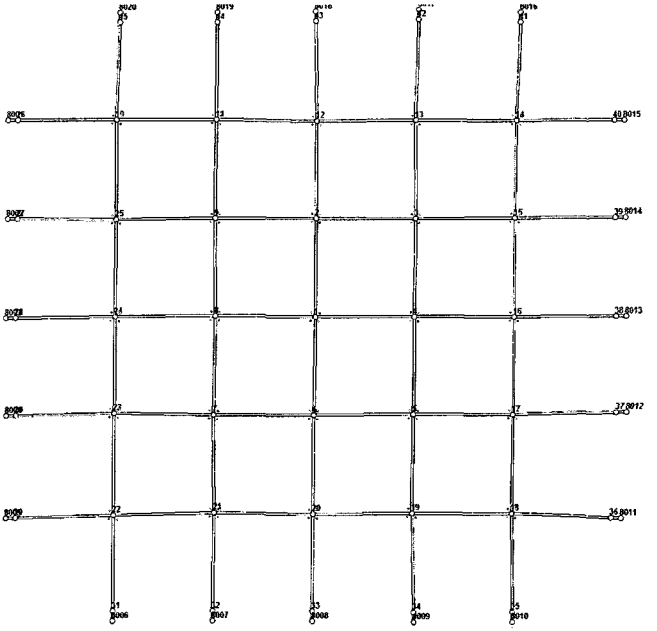


Figure 3. Simple model of a transportation network in an urban area used in initial computational experiments.

Table 1 shows the values of domain and evolutionary computation parameters used in the initial computational experiments reported in this paper.

Table 1. Parameters and their values used in the reported experiments

Domain Parameters	Value(s)	EA Parameters	Value(s)
Number of network nodes	*65*	Representation	*2D cellular automata*
Number of network links	*80*	CA rule types	*Standard*
Number of traffic signals	*25*	Type of EA	*Genetic Algorithm*
Traffic signal states	*2 (corresponding to combined states 0-1 and 2-3 in Figure 1c)*	Population size	*10*
Traffic Assignment	*O-D Table*	Mutation rates	*0.05*
Phase length	*30s*	Crossover rates	*0.8*
Simulation time	*420s*	Fitness	*Number of vehicles leaving the affected area or total travel time*
		Number of runs	*3*
		Termination criterion	*1,000 fitness evaluations*

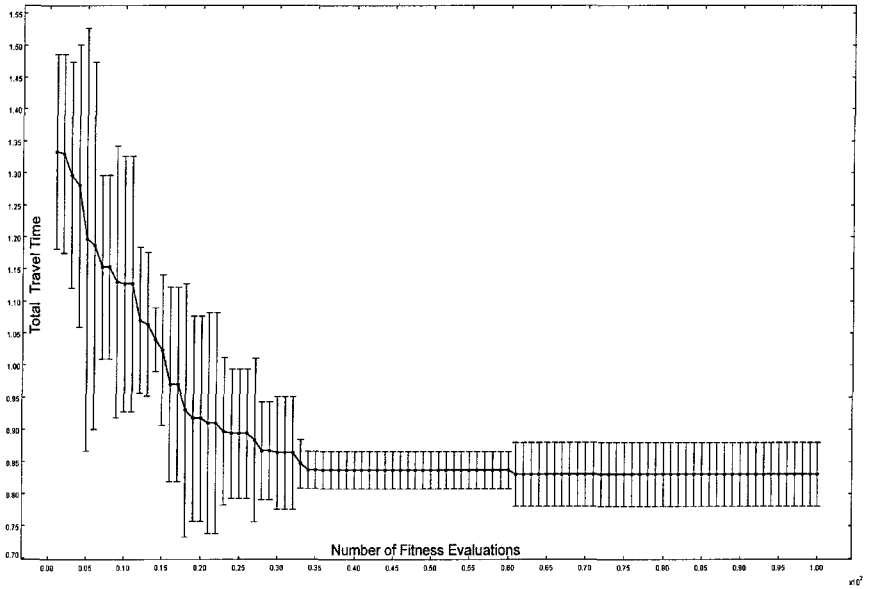


Figure 4. Minimization of the total travel time of vehicles located in the affected urban area obtained by 2D cellular automata models of traffic control systems evolved by evolutionary algorithms.

Figure 4 presents initial results of computational experiments described in Table 1. It shows a gradual improvement of generated 2D cellular automata models of traffic control systems measured by the total vehicle travel time in the affected urban area as a function of the number of fitness evaluations (a measure of an evolutionary progress). The vertical bars represent 95% confidence intervals of mean best-so-far values presented in this graph. Figure 4 shows that the approach presented in this paper is feasible and significant improvement in reducing the total travel time can be achieved. In this specific case, a 35% reduction of total travel time was achieved between initial solutions generated by an evolutionary algorithm and final solutions produced at the end of evolutionary design processes.

Initial Conclusions and Future Work

In this paper, a computational implementation of a self-organizing traffic management hazard response system was discussed and the architecture of a computer system which implements it was presented. It was shown on a simple transportation network that the proposed model and its computer implementation can be effectively used to find evacuation strategies for a representative urban area. On the other hand, the conducted research has also revealed some limitations of existing traffic simulation software.

Even though the initial results are promising, much work remains to be done in order to more comprehensively evaluate the usefulness and efficiency of the proposed approach. First, other examples of transportation networks need to be developed and tested in order to identify more reliably the strong and weak points of this model and the points where greater simulation realism is needed. Next, the complexity of the simulations needs to be expanded in at least three directions: more complex traffic control systems; introduction of one-dimensional CA models of vehicle movements (replacing some elements of the vehicle system simulation); and incorporation of pedestrian traffic into the evacuation scenarios.

References

Bronzini, M., and Kicinger, R. "Conceptual model of a self-organizing traffic management hazard response system." *Compendium of Papers CD-ROM Transportation Research Board 85th Annual Meeting, Washington, DC, January 22-26, 2006*, Washington, DC.

Kicinger, R., Arciszewski, T., and De Jong, K. A. (2005). "Emergent Designer: An integrated research and design support tool based on models of complex systems." *International Journal of Information Technology in Construction*, 10, 329-347.

Wolfram, S. (1994). *Cellular automata and complexity: collected papers*, Addison-Wesley, Reading, MA.

Design, implementation, and test of a wireless peer-to-peer network for roadway incident exchange[1]

Trevor Harmon [2] James Marca[3] Ray Klefstad [2] Peter Martini [2]

ABSTRACT

Vehicular traffic monitoring and control has had a strong infrastructure bias—data is collected centrally, processed, and then redistributed to travelers and other clients. There are several efforts to decentralize traffic monitoring by leveraging advanced local area wireless technology. This paper describes our implementation of such a traveler-centric system, called Autonet. Each Autonet client exchanges network knowledge wirelessly with other, nearby clients. It is demonstrated that knowledge about traffic state can be propagated using this system. The client programs were also used to test the actual throughput possible for messages sent from one vehicle to another using 802.11b wireless hardware. These measurements establish the maximum throughput at about 4,000 incidents for two vehicles moving in opposite directions at highway speeds.

INTRODUCTION

For the next generation of intelligent transportation systems, we envision a decentralized, autonomous communications network between vehicles that leverages the unlicensed spectrum provided by the government and the off-the-shelf hardware that has been developed to exploit this spectrum (802.11a/b/g, DSRC[4] band, etc.). The system we propose expects individuals to share their own travel data freely and allows each traveler to buy as much or as little a device as they need. We call this concept the *Autonet*.

This paper documents an initial implementation of the core vehicle-to-vehicle client application for Autonet. The primary goal is to move from simulation to the real world. The results document the *current* capabilities of hardware for vehicle-to-vehicle communication of traveler information, and they provide working protocols and communications parameters for future simulation studies. The in-vehicle client application will also be useful in future studies of driver interaction and market acceptance.

BACKGROUND

Each Autonet peer does two things: it identifies changes from baseline conditions, and communicates these changes to other vehicles. Vehicles must identify situations that are unexpected, without reference to a central authority, and must transmit this information to other vehicles. The Autonet protocols and systems arbitrate between the information gathered by all vehicles so that a correct interpretation of current conditions can be established and propagated.

[1]This work was supported in part by the California Department of Transportation, Division of Research and Innovation, and by Grant #0339243 from the National Science Foundation.

[2]University of California, Irvine, Department of Electrical Engineering and Computer Science. Irvine, California 92697.

[3]University of California, Irvine, Institute of Transportation Studies. Irvine, California 92697.

[4]Dedicated Short Range Communications, standardized by the U.S. Dept. of Transportation.

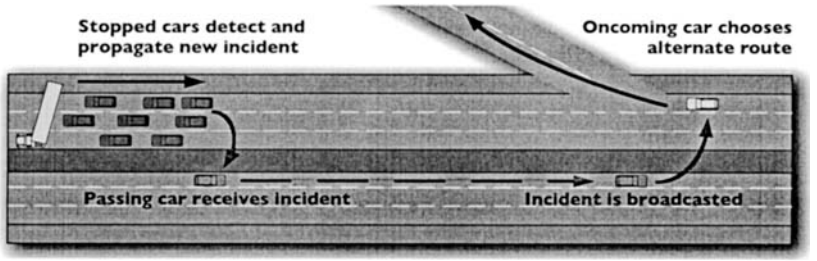


Figure 1. By running variations of this basic scenario on actual roadways, we could judge the prototype's effectiveness.

We explored various ideas for communicating information in the Autonet and decided on the scenario shown in Figure 1. In this scenario, a message about traffic state is passed from one direction of traffic flow to the other for rebroadcast to upstream vehicles. We tested this scenario with a proof-of-concept prototype built entirely from COTS (commercial off-the-shelf) hardware. Although 802.11b seemed like the obvious choice for this prototype's wireless technology, this decision brought with it a host of questions and somewhat contradictory answers.

Ebner (Ebner et al. 2003) recommended against using 802.11b, stating that simulation results indicated that its performance would be unacceptable in urban environments and at high speeds. In contrast, Singh (Singh et al. 2002) conducted several field tests and concluded that 802.11b performance is "suitable for inter-vehicle communications." These papers are not in conflict, but rather interpret "suitability" in different ways. Ebner was striving for a seamless approximation of the Internet on the road over 802.11b, while Singh had a much more modest goal of vehicle-to-vehicle communication.

In other related work, Nadeem (Nadeem et al. 2004) designed a system around 802.11b and reported some results mixing real world tests and simulation for vehicles moving in the traffic stream. Wu (Wu et al. 2004) developed a detailed protocol for sending a message from one geographic area to another, which was tested in simulation using the 2 Mbps 802.11 DCF MAC protocol. Finally, Aziz (Aziz 2003) used 802.11b in infrastructure mode to enable vehicle-to-vehicle communication between vehicles with fixed IP addresses while traveling along a corridor.

The Autonet has modest requirements for vehicle-to-vehicle communication. It does not require a complete, multi-hop network to obtain quantitative improvements in the level of system awareness for participating drivers. Given that a single-hop network is all that is requied for the initial Autonet vision, we concluded that 802.11b was more than adequate for building our field test units.

DESIGN

As a first step in reaching the goals of Autonet, we have developed a prototype Autonet computer. This prototype demonstrates the feasibility of Autonet and provides a platform for testing our algorithms in a realistic environment on public roads using COTS hardware.

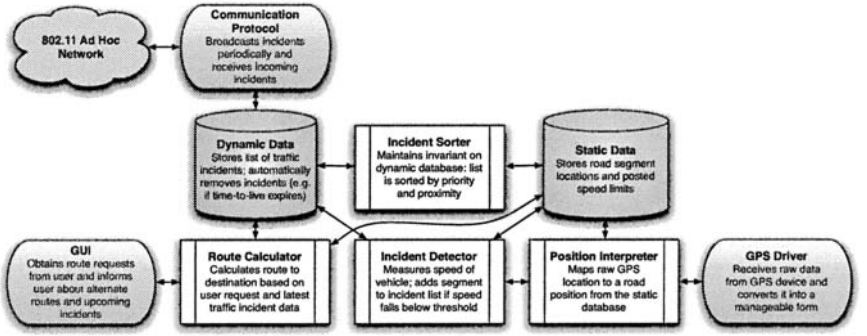


Figure 2. The major software components of the Autonet prototype.

Software components

Figure 2 shows the major software components and the relationships between them. Each of these components is designed to operate concurrently with the others, performing operations on incidents as they come in and responding to user requests asynchronously. For simplicity, the components can be thought of as operating in a sequential fashion as in the following scenario:

1. The Communication Protocol receives a message from a passing Autonet vehicle. The message describes the location, time observed, and severity of the incident.
2. The Dynamic Data component receives the new incident and checks it for validity. Valid incidents are placed into dynamic storage.
3. The Incident Sorter monitors the Dynamic Data for changes and re-sorts it according to some evaluation function. This function takes into account the spatial and temporal proximity of the incident, and its severity, among other criteria.
4. The Route Calculator monitors the Dynamic Data for changes and recalculates the shortest route to the current destination, incorporating new incident information.
5. The Communication Protocol queries the database for the highest priority incidents, as this set may have changed with the addition of new incidents. It broadcasts the set to other Autonet-enabled vehicles.

Communication protocol

The communication protocol impacts the overall quality of the system. It is responsible for exchanging incidents with other Autonet-enabled vehicles, and it ensures that users have accurate knowledge of the true traffic state. It accomplishes these tasks by broadcasting a list of incidents—segments of road observed to be slower than expected—while simultaneously listening for incident reports from the surrounding vehicles.

The message format for incident exchange is simple; the decision of which incidents to send is not. The problem is that over time, a group of Autonet-enabled vehicles may build up a list of tens of thousands of incidents. According to our measurements (see Section 6), two vehicles passing each other in opposite directions can each exchange only a few thousand incident messages with 802.11b

technology, even under the best of conditions. Therefore, the communication protocol must decide which portion of the incident list to broadcast. We address this problem by assigning a priority to each incident, calculated with an evaluation function that takes into account:

- **Deviation from expected speed** The higher the deviation from the expected speed, the more important is the incident. Currently, we define expected speed as the posted speed limit for the road, but for future prototypes we will define it more accurately as the historical average of vehicles passing on the road segment.
- **Size of the road** Highways have higher priority than local roads, etc.
- ***Distance from local position*** *Nearby incidents are more likely to have a direct* impact on travel than distant incidents.
- **Time** The older the information, the less accurately it describes current conditions.

The Incident Sorter applies its evaluation function to sort the incident data by priority. The communication protocol then cycles through the highest priority incidents and broadcasts them continuously for passing vehicles to collect.

IMPLEMENTATION

To show that the Autonet concept is feasible, we constructed prototypes of an Autonet device and installed them in three vehicles. These prototypes helped us perform experiments in a real-world environment using COTS hardware: basic laptop computers, an 802.11b wireless Ethernet card in ad hoc mode, and GPS receivers that communicated with the laptops over Bluetooth. For software, we implemented the components shown in Figure 2 in the Java language.

Incident detection

Our Incident Detector, as discussed in Section 4, provides a basic algorithm for detecting new incidents. The inputs are speed limit information from the static database, and the current speed obtained from the Position Interpreter. An incident is declared if the vehicle drops below 80% of its expected speed. Although this technique works well for freeway driving, it is inadequate for roads with flow interruptions such as intersections and traffic controls.

Although we could have attempted to mitigate these problems, much work has already been done in the transportation research community on incident detection, and we did not wish to re-invent the wheel for our initial prototype. As we discover a known set of capabilities and resources for an Autonet peer, we will then turn to integrating the rich body of literature on incident detection techniques into the unique requirements of the Autonet devices.

User interface

To the casual viewer, the display of an Autonet device would appear no different from commercial navigation systems that are becoming more popular. The Autonet user interface would improve upon these more conventional systems, however, by displaying up-to-the-minute traffic information collected from other Autonet-enabled vehicles. By combining this information with the driver's chosen route, Autonet's interface can then display a more appropriate route (if found) superimposed over a map of the vehicle's surrounding area.

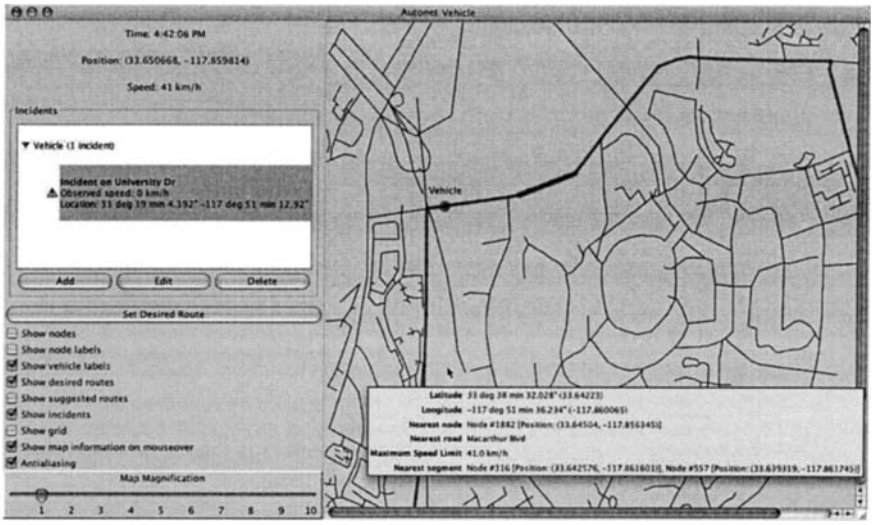


Figure 3. The prototype shown in this screenshot is used to explore new feature ideas and to verify the performance and correctness of Autonet.

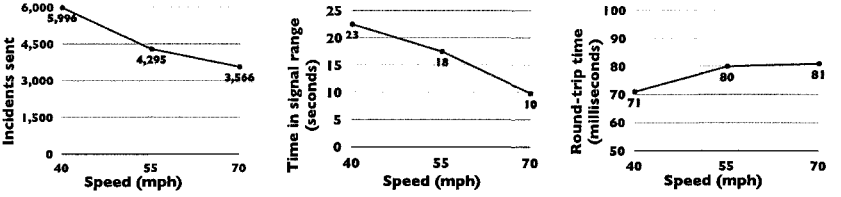
Figure 3 shows our prototype's user interface. The left side of the screen shows the vehicle's speed and position as reported by the GPS device, and a list of known incidents. This list includes directly observed incidents, and second-hand information received from other Autonet vehicles. At the bottom are controls for adding, editing, and removing incidents and for manipulating various attributes of the map display. The right side of the screen shows the road map of the vehicle's location. The map shows the vehicle's position and its chosen route.

EMPIRICAL RESULTS

To benchmark the prototype, we focused on the real-life scenario shown earlier in Figure 1. Three vehicles (at minimum) needed to be equipped with our Autonet prototype to test this scenario. We successfully executed a number of variations of this scenario on streets and freeways in Irvine. In addition, we ran a variety of custom benchmarking programs for testing the performance of 802.11b technology in the Autonet environment. Figure 4 shows the result of this effort.

For Figure 4(a), we tested the ability of 802.11b to transfer incidents between two vehicles that pass each other in opposite directions at identical speeds. The horizontal axis refers to the *absolute* speed of the vehicles. The plot shows the average number of incidents that each vehicle could send over four trial runs.

As expected, the number of incidents exchanged drops sharply as the speed of the vehicles increases. The decline is largely due to the shorter time in which the two vehicles are within signal range, but we suspect that the Doppler effect also plays a role in limiting throughput with increasing speed. We note, however, that even when two cars travel at highway speed—70 miles per hour—they are able



(a) Raw throughput for incident transfer.

(b) Time the vehicles remained within signal range.

(c) Round-trip time for query/response.

Figure 4. These graphs show our benchmark test results of 802.11b communication, aided by external antennas, between vehicles moving at identical speed in opposite directions on public roads.

to send and receive over 3,000 incidents each. These are very promising results, as they show that one vehicle can inform another about every single segment of road for a typical urban area of 64 square kilometers. If we assume that one out of every five segments is an incident, the area grows to 324 square kilometers.

For Figure 4(b), we measured the amount of time during which two vehicles, when traveling in opposite directions, could remain in 802.11b signal range of each other and successfully exchange incidents. The surprisingly long periods during which the two vehicles were able to remain in contact can be attributed to the external antenna mounted on each vehicle's roof. Although we had hoped to collect data without this antenna for comparison purposes, we were unable to exchange incidents reliably after removing it, even at slow speeds. The consequence for Autonet is that any production version of the device would require the consumer to install an external antenna atop their vehicle.

Finally, Figure 4(c) shows the round-trip time for message exchange using 802.11b. Here, we define "round-trip time" as the time required for one vehicle to send an incident request to another and fully receive a response to the query. Based on these measurements, combined with the measurements from Figure 4(b), we calculate that when using a request/response method of incident exchange, a minimum of approximately 100 incidents can be received per vehicle. (That is, at the worst case of 70 mph, the vehicles are in signal range for ten seconds, so 10 seconds $\div$ 81 milliseconds = 123.)

CONCLUSION

This paper presents the results of a real-world implementation of a vehicle-to-vehicle communication system which enables decentralized traffic information propagation. Future work will focus on defining the minimum requirements and capabilities of an Autonet peer. These capabilities will be developed in tandem with simulation efforts aimed at determining more robust communication protocols and incident detection algorithms.

REFERENCES

Aziz, F. M. (2003). "Implementation and analysis of wireless local area networks for high-mobility telematics." Master's thesis, Virginia Polytechnic Institute and State University.

Ebner, A., Rohling, H., Wischhof, L., Halfmann, R., and Lott, M. (2003). "Performance of UTRA TDD Ad-Hoc and IEEE 802.11b in vehicular environments." *57th IEEE Semiannual Vehicular Technology Conference*, Vol. 2, IEEE, Jeju, South Korea. 960–964.

Nadeem, T., Dashtinezhad, S., Liao, C., and Iftode, L. (2004). "TrafficView: Traffic data dissemination using car-to-car communication." *Mobile Computing and Communications Review*, 8(3), 6–19.

Singh, J., Bambos, N., Srinivasan, B., and Clawin, D. (2002). "Wireless LAN performance under varied stress conditions in vehicular traffic scenarios." *Proceedings. IEEE 56th Vehicular Technology Conference, Fall 2002*, Vol. 2. IEEE, 743–747.

Wu, H., Fujimoto, R., Guensler, R., and Hunter, M. (2004). "MDDV: a mobility-centric data dissemination algorithm for vehicular networks." *Proceedings of the first ACM workshop on Vehicular ad hoc networks*, New York, NY. ACM Press, 47–56.

Harnessing Advanced Technologies to Eliminate Recurring Freeway Congestion

Patrick DeCorla-Souza, AICP[1]
Federal Highway Administration

Abstract: *The Operate-Design-Build-Operate (ODBO) highway infrastructure-financing approach presented in this paper attempts to eliminate recurring freeway congestion by introducing congestion-based tolls prior to infrastructure investment with all surplus revenue dedicated for infrastructure expansion. Not only would the existing highway system operate more efficiently; but so would the expanded system. Up-front toll revenues would help pay for expensive urban freeway expansion projects, making them more financially feasible. New public-private partnership approaches are suggested that employ outcome-based contracting systems and financial incentives to maximize public mobility goals.*

Implementation of road pricing is faced with numerous constraints, including economic, financial, operational, political, social, and technological considerations. This paper attempts to deal with multiple considerations on a comprehensive basis, illustrating ways to accommodate high levels of system performance and economic efficiency within a broad framework of public acceptance and political reality.

Why Price Roads?

Some states are finding that, in the future, existing sources of revenue from vehicle taxation will be sufficient only to maintain and operate the existing transportation system, but not for expansion or rehabilitation. States are looking to tolling and public-private partnerships to make up for the shortfall in revenue. It is envisioned, for example, that new highway lanes and new highways in metropolitan areas will be financed from tolls that will be collected on the new infrastructure. This financing model is relatively easy for the public and transportation stakeholder groups to understand and accept, in light of the dwindling value of tax receipts. However, due to high costs for urban freeway expansion, in many cases full financing of such projects may not be supportable from toll revenue alone, delaying needed projects.

On the other hand, from both engineering and economic perspectives, efficiency of existing, new and expanded highways may be maximized by introducing *peak period* tolls on congested segments, with toll rates set at levels that reduce demand sufficiently to eliminate congestion caused by oversaturated conditions. Such congestion-based pricing increases vehicle throughput by eliminating the loss of throughput that occurs under severely congested conditions, while also increasing travel speeds. Once freeway vehicle density (measured in vehicles per mile) exceeds a certain critical value, both vehicle speed and vehicle flow (measured in vehicles per hour) drop precipitously. Empirical evidence suggests that vehicle throughput may

[1] Patrick DeCorla-Souza, Office of Policy and Governmental Affairs, Federal Highway Administration, 400 Seventh St. SW, Room 3324, Washington, DC 20590;
Tel: (202)-366-4076; Fax: (202)-366-7696; e-mail: patrick.decorla-souza@dot.gov
AUTHOR'S NOTE: The views expressed in this article are those of the author and not necessarily those of the Federal Highway Administration or the U.S. Department of Transportation

drop as much as 50 percent, while speeds may drop to 15 mph or lower (U.S. Department of Transportation 2004). Even if demand decreases after the breakdown of traffic flow, the freeway does not recover its full efficiency until much later, because queued vehicles keep vehicle density high. At these high densities, the freeway is kept in breakdown flow condition much after the peak demand period has ended (Chen and Varaiya 2002). With road pricing, a variable toll dissuades some motorists from using freeways approaching critical density and prevents surges in demand that lead to a breakdown of traffic flow, thus maintaining a high level of vehicle throughput. Due to the increased throughput on free-flowing freeways, diversions may occur from parallel arterials to the freeway, reducing congestion in the entire travel corridor.

The idea that prices should be highest at times of heavy demand has been established in other transportation sectors such as air travel, and has been widely accepted in principle. As with any market-pricing mechanism, road pricing helps allocate limited supply (of road space) and enhances efficiency.

The Operate-Design-Build-Operate (ODBO) Approach

Since tolls in peak periods can increase freeway system efficiency and consequent social benefits, one is drawn to an intriguing proposition. Why not introduce tolls on existing congested freeways, only during rush hours, to keep demand from oversaturating the facilities – then allocate the revenues as a "down payment" on costs for infrastructure expansion, in a later phase? This financing approach may be described as an Operate-Design-Build-Operate (ODBO) approach.

Revenue obtained during Phase 1 (i.e., prior to infrastructure expansion) would increase the chances for financial viability of high-cost urban freeway expansion projects. Since motorists who pay tolls in advance of infrastructure expansion might feel that they are not getting "physical" infrastructure improvements for their payments, they could be allocated discount credits towards tolls charged after infrastructure expansion. Improved transit and ridesharing services, funded from surplus toll revenue, could provide new travel options for those not able to afford the tolls. To address equity issues, low-income commuters who have no reasonable access to transit for their commute trips could be provided with toll credits or toll discounts.

Initially, in Phase 1, private partners could be selected under a short-term contract to invest in infrastructure for traffic management and peak-period tolling on the existing system, and to maintain and operate the infrastructure and collect tolls. A private partner's skills would be valuable for deployment of the complex schemes and innovative technologies that would be needed. "Shadow" tolls could be used to compensate the private partner. Shadow tolls are highway usage payments made by a third party. Although the private partner would set the *real* toll rates to manage demand and ensure that traffic is free-flowing, *all toll revenue would go to the public sector*, and the public agency would reimburse the private partner with a flat shadow toll paid for each vehicle *served at free flow speeds* during rush hours when tolling is in effect. This model is termed the Concurrent Real And Shadow Tolling (CRAST) model (DeCorla-Souza 2006a).

The CRAST model would reduce revenue risk to the private partner while managing public cost risk, and could lead to more innovative ways to manage traffic since the private partner would seek to maximize free-flowing traffic conditions *and* vehicle throughput. The public partner could require that public transit and authorized vanpools be provided with toll-free service. The private partner would not lose revenue as a result because it would still receive shadow tolls for serving those vehicles. The public partner would not need to be bound by a "non-compete" clause in the partnership agreement that would prohibit it from making improvements to alternative routes. Such improvements to toll-free roads might cause real toll rates to drop and real toll revenue to be reduced. But this would not cause concern to the private partner, because the private partner would continue to receive its full shadow toll payments. The public partner would be the one to lose revenue. Travel demand in growing metropolitan areas is so heavy that freeways could always be filled to free-flow capacity in rush hours by dropping the real toll rate down to zero if necessary, thus having no effect on total shadow toll payments to the private partner.

High real (user-paid) rush hour toll rates on some freeway segments would indicate the urgent need for capacity enhancements and at the same time provide some of the funding needed for the investment. Based on observed toll rates and corridor traffic forecasts, freeway segments could be prioritized for capacity enhancements in Phase 2. Private proposals could be solicited to address expansion needs. The preferred proposal, which could be from the Phase 1 partner, could then be carried through the environmental review process with assistance from the winning private partner. After a Record of Decision (ROD) on the environmental document, the operation of the facility would be transferred to the winning private partner, and Phase 2 would begin. The private partner would proceed to finance, design and build the project, and would operate and collect tolls on the facility during the design and construction phase and after the expansion is completed. If the private partner were paid using the CRAST model, there would be an incentive to safeguard against disruption of traffic flow during the construction phase.

Surplus revenue from rush hour tolls on existing facilities in Phase 1 would provide some of the needed finances to make Phase 2 shadow toll payments, especially during the early years of Phase 2, when real toll revenue may actually drop below levels attained prior to roadway expansion, because toll rates would fall with the increase in available supply of road space. (The private partner would need to attract motorists with lower values of time in order to maximize vehicle throughput on the increased capacity.) Tolled traffic data from Phase 1 would reduce the difficulty in forecasting revenue from tolls after infrastructure expansion. This would reduce uncertainty of revenue forecasts. Public agencies would have more confidence in revenue forecasts, reducing uncertainty relating to the adequacy of real toll revenue to support shadow toll payments to the private partner.

Multimodal Strategies

To successfully reduce peak period traffic levels without resorting to exorbitant tolls, it is important that reliable, convenient travel alternatives be available for those who do not wish to pay the new rush hour tolls. The first step is therefore to have alternatives to solo driving in place before rush hour tolls are implemented. Travel

time advantages relative to solo driving are critical for success of transit and vanpool services. An extra "*rush hour lane*" could be established in each direction on all freeways by re-striping existing highway pavement to allow shoulder use by transit and authorized vanpools. Restricting use of the shoulder lane to authorized vehicles with trained drivers would ensure that safety would not be compromised.

The Phase 1 private partner could be responsible for making the roadway modifications needed to introduce rush hour lanes. It is important that rush hour lanes and new vanpool/express bus systems be in operation at least three months prior to introduction of rush hour tolls, with free transit and vanpool trial periods to encourage use. This will allow the public to get familiar with the new modal options and allow rush hour tolls to be introduced with fewer concerns from the public about the viability of travel alternatives.

A modification of the CRAST model in Phase 2 may provide incentives to the private partner to innovate with ways to encourage use of alternative modes. Photographic systems employing near infra-red cameras have achieved some success in counting vehicle occupants, and have been shown to have some potential for further improvement and deployment (University of Minnesota Department of Computer Science 1999, Dalton 2005). Phase 2 contracts could be designed under which the private partner would be paid on the basis of the number of *person* trips (rather than vehicle trips) carried on the freeway facility during peak periods, monitored using advanced technologies to count vehicle occupants. For example, if a vehicle has four occupants, the private partner would be paid four times the shadow toll per person.

With shadow toll payments based on *person* throughput rather than vehicle throughput, private partners would have an incentive to design Phase 2 proposals that encourage greater *person* throughput during rush hours by promoting transit and ridesharing modes. They might work with other private and public partners to provide or subsidize vanpool or express bus services and/or appropriate collection and distribution services for express bus trips. They might market travel options to the commuting public. The winning bidder could be selected based on the proposed shadow price per person trip plus consideration of the amount of new *person trip* capacity that would be provided, its adequacy to serve forecasted travel demand, and degree of financial self-sufficiency of the multimodal proposal, i.e., the extent to which real toll revenue would be adequate to support "per person" usage payments. This approach would maximize the flexibility of private bidders to come up with the most innovative, cost-effective and financially viable proposals and would minimize public and social costs.

Region-wide Implementation

In an attempt to evaluate the potential of the ODBO approach, the author estimated costs and revenues from region-wide implementation of a multimodal investment and pricing strategy in a prototypical large metropolitan area with about 4 million people and about 300 miles of freeway (about 2,000 lane miles). Phase 1 envisioned application of rush hour pricing on all lanes of the entire *existing* freeway system with only minor modifications to provide rush hour lanes. Private partners would be sought to invest in and operate sub-networks of the freeway system for a period of

about 5 years, with options to extend the contract period. Phase 2 envisioned long-term public-private partnerships to provide 600 new lane miles on the freeway network in an attempt to keep up with future growth in travel demand. Price signals (i.e., higher toll rates on some segments) in Phase 1 would provide an indicator of the priority that should be given to capacity additions proposed for Phase 2. Private partners would be sought to build, finance and operate the improved facilities for the longer term.

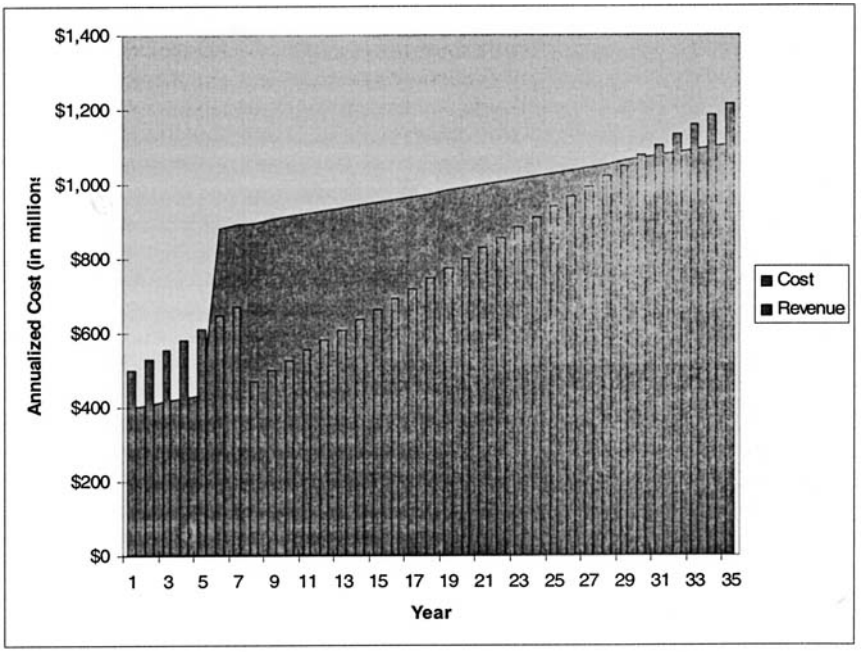


Figure 1. Regionwide Multimodal Costs and Revenues

In order to simplify this illustrative analysis, it was assumed that *all* improvements will be completed at the end of year 7. Also, all estimates were made in real 2005 dollars. Procedures used to estimate costs and revenue are discussed elsewhere (DeCorla-Souza 2006a and 2006b). The revenue analysis assumed that, in order to address double taxation concerns, it will be necessary to rebate to freeway motorists the fuel taxes paid on fuel consumed by their vehicles during the tolled time periods. Results of the multimodal financial analysis are presented in Figure 1. The results suggest that toll and transit fare revenue will be adequate to support the multimodal system in Phase 1. In Phase 2, there will be a multimodal budgetary shortfall each year until year 30 (see Figure 1). However, after capital costs for freeway expansion are fully amortized at the end of Phase 2, there will be continuing and significant annual surpluses, since the annualized costs for lane additions would drop to zero. Thus, a longer-term (e.g., 50-year) financing arrangement involving

private equity may be adequate to fill the multimodal deficit gap in the early years. Alternatively, public dollars that would have otherwise been needed to pay for freeway maintenance and operation under public operation of the facility could be used to pay for some of the shortfall.

Summary and Conclusions

The Operate-Design-Build-Operate (ODBO) highway infrastructure-financing approach presented in this paper attempts to eliminate recurring freeway congestion by introducing congestion-based tolls *prior* to infrastructure investment with all surplus revenue dedicated for infrastructure expansion. Not only would the existing highway system operate more efficiently; but so would the improved system. Up-front toll revenues would help pay for expensive urban freeway expansion projects, making them more financially feasible. New multi-modal public-private partnership approaches are suggested that employ outcome-based contracting systems and financial incentives to maximize public mobility goals.

Acknowledgements

The author would like to acknowledge helpful comments on earlier versions of this paper received from several colleagues at the U.S. Department of Transportation, as well as from various other transportation finance experts, particularly Stephen Lockwood of PB Consult. However, the author alone is responsible for any errors or omissions, and the views expressed are those of the author alone and not necessarily those of the U.S. DOT or the FHWA.

REFERENCES

Dalton, Alastair. (2005). Infrared Cameras Will See Through Forth Bridge Dummy Runs. *The Scotsman.* Tuesday, November 29, 2005.

DeCorla-Souza, Patrick. (2006a). Implementing Congestion Pricing on Metropolitan Highway Networks with Self-Financing Public-Private Partnerships. Forthcoming in *Journal of the Transportation Research Forum.*

DeCorla-Souza, Patrick. (2006b). Improving Metropolitan Transportation Efficiency with Fast Miles. *Journal of Public Transportation.* Volume 9, no.1. Center for Urban Transportation Research, Florida.

University of Minnesota Department of Computer Science. (1999). *Automobile Passenger Counting in the HOV Lane.* Prepared for Minnesota Department of Transportation.

U.S. Department of Transportation (2004). *Report on the Value Pricing Pilot Program through March 2004,* http://knowledge.fhwa.dot.gov/cops/hcx.nsf/All+Documents/AD276ECC2E3A077885257005006B5614/$FILE/March%202004%20Report%20of%20Congress.pdf

Mapping Misuse Cases to Functional Fault Trees in order to Secure Positive Train Control Systems

Mark Hartong[1], Rajni Goel[2] and Duminda Wijesekera[3]

[1]Federal Railroad Administration; Washington, DC email: mark.hartong@fra.dot.gov
[2]Howard University. Washington, DC email: rgoel@howard.edu
[3]George Mason University, Fairfax, VA email: dwijesek@gmu.edu

Abstract

Use cases specify the higher-level functional requirements of a system under design and misuse cases specify its possible misuses. Analyzing them together prevents the latter from occurring while ensuring that the former are implemented. Independently, functional fault trees (FFT) hierarchically breakdown anticipated system failures with respect to its functional architecture. This work presents an algorithm that transforms higher-level use-misuse case to FFTs, and thereby allows the application of the analytical methods available for the latter to the former. The utility of such a mechanism is illustrated by studying the security vulnerabilities that can be introduced to Positive Train Control (PTC) systems - wireless command and control systems for the safe operation of freight and passenger trains.

Introduction

Use cases and misuse cases, respectively, specify functional requirements and potential abuses of systems under design. Together they specify system functionality, while seeking guarantees against known abuses of a system under design. Positive Train Control Systems (PTC) are currently under design and rely on wireless communications to convey control signals in a timely manner, and a potential exists for known protocol vulnerabilities of the latter to have safety critical consequences on the former. Consequently, we employ use cases and misuse cases to model the functional requirements and potential abuses of the PTC system under design. Traditionally, functional faults trees (FFT) have been used in the safety critical design arena that hierarchically specify and analyze safety requirements at a higher level, such as analyzing cut sets etc. This paper presents an algorithm that maps higher-level use-misuse cases to functional fault trees and demonstrates how the analytical tools developed for FFT aid analysis of the requirements specified by the use-misuse scenarios. With the mapping, the intuitive graphical depiction of functional requirements, including the mishaps that are be to be avoided by a system, can still be automatically subjected to the rigorous analysis of FFT based tools. An ideal case study for our algorithm is the specifications of Positive Train Control (PTC) systems, wireless communication based control systems for the safe operation of passenger and freight trains. The train's functionality has been considered safety critical, and the dependency of their controls upon wireless communication, known to contain vulnerabilities, make PTC systems an ideal candidate for analysis through Use-Misuse Case specifications.

Use and Misuse Cases

Use Cases specify functional requirments of a system in terms of its operations, and constitute the first stage of a Unified Modeling Language (OMG 2004) based process to specify, visualize and document the artifacts of software systems and business requirements of non-software systems. Because security is a critical component in software based systems, Use Cases have been enhanced by Misuse Cases (Opdahl & Sindre 2001) to specify the foreseeable interactions between a potential mal-actors and the system under design so that the designed system prevents specified misuses, while providing specified functional requirements.

Functional Fault Trees

FFT are a tree structured hierarchical specification of potential faults in a system's functional architecture, which may result in system failure. With well-understood techniques and methodologies, this approach has been successfully used over the past 40 years for requirements analysis. Traditionally, developers have used structured FFT Analysis (Vesely et al 1981) for developing risk-based safety requirements, and FFT's have directly specified security requirements (Helmer et al 2001). In formal FFT specifications, event relationships resulting in a failure are expressed using Boolean algebra. By simplifying such Boolean algebra expressions, these events can be expressed as a union of minimal cut sets that individually represent combinations of basic faults that result in the system failure so that their removal can result in a system that is devoid of known modes of failure. Associating a probability for each basic failure results in an estimation of the probabilities of various failure modes (each of the minimal cut sets), and the aggregate probability of failure of a system.

Transforming Use-Misuse Cases to Funtional Fault Trees

Though the equivalency of FFTs and the UML notation has been suggested (Bitsch 2002, or Hawkins et-al 2003) a formal translation of Misuse Cases to FFTs is nonexistent. In developing such a transnlation first a syntax for Use-Cases, Misuse Cases and FFT's is specified using Bacus Naur Form (Naur 1960). Then the mapping from Use and Misuse Cases to a FFT is given in the following algorithm.

Algorithm UML_to_FFT (in: Use_Misuse_Tree; ; Out: FFT_Tree)

procedure parse uml-fft (**in/out**: vertex use/misuse tree)
begin
visit (vertex use/misuse tree)
for each child (vertex use/misuse tree)
if child is unvisited
parse uml-fft (child)
else
case
1: child is <<*misuse*>> and parent is <<*use*>> and relation child to parent is <<*threaten*>> then fft_event is <<*conditioning_event*>> where <<*conditioning_event*>> is child /*note use case for conditioning */
2. child is <<*misuse*>> and parent is <<*misuse*>> and relation child to parent is <<*aggravates*>> then fft_event is <<*conditioning_event*>> where <<*conditioning_event*>>is child /*note misuse case for conditioning *

3. child is <<*use*>> and parent is <<*use*>> and relation child to parent is <<*conflicts*>> then fft_event is <<*conditioning_event* >> where <<*conditioning_event*>> is child /*note use case for conditioning *
4. child is <<*use*>> and parent is <<*misuse*>> and relation child to parent is <<*mitigates*>> then fft_event is <<*conditioning_event*>> where <<*conditioning_event*>>is child and /*note misuse case for conditioning */
5. child is <<*use*>> and parent is <<*misuse*>> and relation child to parent is <<*aggravates*>> then fft_event is <<*conditioning_event*>> where <<*conditioning_event*>> is child and /*note misuse case for conditioning *
6. ((child is <<*misuse*>> and parent is <<*use*>>) or (child is <<*use*>> and parent is <<*use*>>)) and relation child to parent is <<*prevents*>> then fft is <<*intermediate or external or undeveloped event* >><<*conditioning event*>> <<*inhibit*>><<*intermediate event*>> where <<*intermediate or external or undeveloped event* >> is child and <<*intermediate event*>> is parent
7. child is <<*misuse*>> and parent is <<*use*>> or <<*misuse*>> and relation child to parent is <<*includes* >> then fft is <<*undeveloped or external or base or intermediate event* >><<*gate*>><<*intermediate gate*>> where <<*basic or external or undeveloped or intermediate event* >> is child, <<*gate* >> is one of {*and, or, exclusive or, priority and, negation*} and <<*intermediate event*>> is parent
8. child is <<*use*>> and parent is <<*use*>> and relation child to parent is <<*includes* >> then fft is <<*undeveloped or basic or external or intermediate event* >><<*gate*>><<*intermediate gate*>> where <<*basic or external or undeveloped or intermediate event* >> is child, <<gate >> is one of {*and, or, exclusive or, priority and, negation*} and <<*intermediate event*>> is parent

end case
end if
end for
end parse
begin
/* main body */
starting at root of UML tree
parse uml-fft (UML Tree) /*generates FFT fragments */
join FFT fragments
end

PTC Wireless Security Requirements

Communications based PTC systems (FRA 1999) are command and control systems designed to prevent train-to-train collisions, enforce speed restrictions, including civil engineering restrictions and temporary slow orders and protect roadway workers with their equipment operating under specific authorities. PTC communications security issues (Carlson et al. 2003, Craven 2004) impose additional risk based requirements on PTC systems that developers must consider. Although the vulnerabilities arise out of exploiting common shortcomings of communicating subsystems, they manifest in a specific system by disrupting the functionality in a predictable manner that can be captured through structured analysis.

Figure 1 illustrates a subset of the Misuse cases that affect a subset of a sample Use Case. Potential mal-actors are abstracted to a single attacker in the Misuse Case, and the office/dispatch, wayside, and mobile unit operators are PTC actors in the Use Cases. All actors (attackers) communicate by exchanging messages, the format of which is implementation and system dependent. The Use Case eventually "chains" back to one or

more of the base PTC functionalities. This structure, when traversed from the negated PTC base functionality to the originating Use Case, is a tree or directed graph structure, where Use Cases are vertices, and relationships are edges.

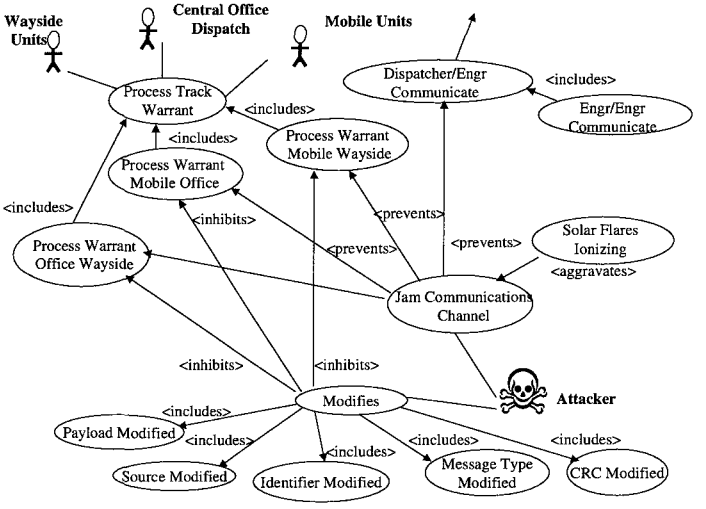


Figure 1: PTC Use-Misuse Case Fragment

Using the algorithm, the Use-Misuse Case diagram fragment of Figure 1 translates and reduces to the FFT fragment of Figure 2. Then, using cut set minimimization as illustrated in the Steps 1 through 4 of Figure 3, the cut set FFT of Figure 3 is derived. Completing this process for known Use and Misuse Cases, the resulting minimal cut sets determine all possible combinations of Misuse Cases that result in security failures, the relative damage that can result, and those parts of the PTC system that are susceptible to shared vulnerabilities. By associating probabilities of occurrence with the individual Misuses Cases, and utilizing standard FFT probability calculations, quantitative measures for the systems susceptibility to attack and the value of protective measures and countermeasures can be determined to augment the qualitative minimal cut sets and common cause failures. This is not achievable using normal Use-Misuse Case Analysis technique.

Fault tree probability calculations are fairly straightforward - the failure probabilities of the basic events are combined in either "and" or "or" gates which are then combined to find the probability of occurrence for each sequence in each of the trees. When a basic event is used in more than one location in the fault trees this simple approach cannot be used. These common events destroy the independence of the gates above them, making the straightforward approach unusable. Instead, for small numbers of common components, it is possible to calculate every combination of the common components in either a failed (p=1.0) or not-failed (p=0.0) state weighted by the actual probability of those components being in those states. This approach can get quite time-consuming for large numbers of common components, and a Monte Carlo trial approach is often substituted. In this approach, the common components are sampled, based on their true failure probabilities, to be either failed

or not failed. These are then combined with the other basic events to calculate the fault trees sequence probabilities.

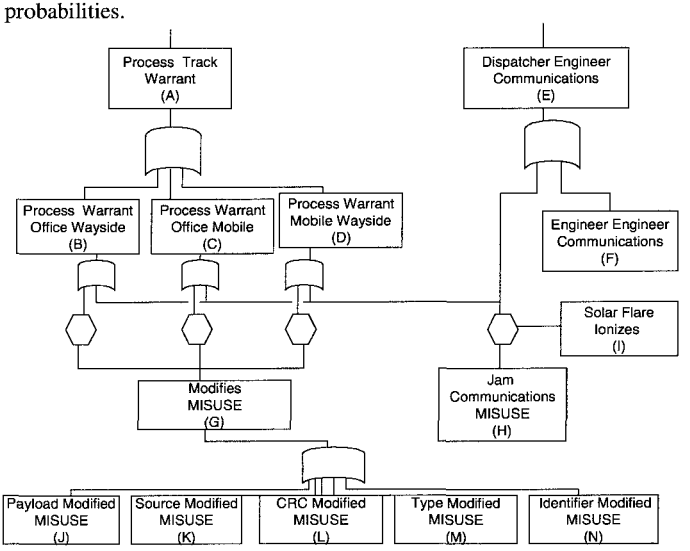


Figure 2: Translated FFT Fragment

Step 1. A =B+C+D
B =NOT G+ (NOT H mod I)
C= NOT G+ (NOT H mod I)
D= NOT G+ (NOT H mod I)
G= J+K+L+M+M+N

E = F + (NOT H mod I)

Step 2. A = NOT G+ (NOT H mod I) + NOT G+ (NOT H mod I) + NOT G+ (NOT H mod I)
G= J+K+L+M+M+N

NOT E =NOT(F + NOT (H mod I))

Step 3 NOT A = G+ (H mod I)
G= J+K+L+M+M+N

NOT E = NOT F * NOT (NOT H mod I)

Step 4 NOT A = J+K+L+M+M+N + (H mod I)
NOT E = NOT F * H mod I

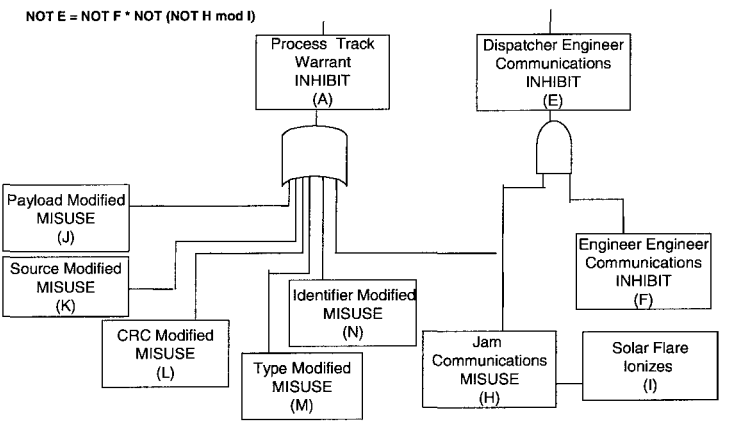


Figure 3: Minimal Cut Set FFT

Summary & Future Work

A shortcoming of Use-Misuse Case analysis is a lack of support for quantitative analysis, a critical issue in ongoing studies of PTC systems. Combining Use-Misuse Case analysis with FFT equivalency to identify ways in which a mal-actor can prevent the functional objectives of a system from being enforced, as well as quantitatively and qualitatively evaluate system failure modes resolves this shortcoming. However, additional study and formalization of Use-Misuse Case and FFT equivalency language as well as automation of the translation process is required to fully exploit the potential advantages of this technique.

References

Bitsch, "Requirements on Methods and Techniques in Perspective to Approval Process for Railway Systems", Proceedings of the 2nd International Workshop on Integration of Specification Techniques for Applications in Engineering - Satellite Event of ETAPS 2002.

Carlson, Frincke, Laude, "Railway Security Issues: A Survey of Developing Railway Technology" Proceedings of the International Conference on Computer, Communications, & Control Technology, International Institute of Informatics and Systemic, 2003

Craven, "A Brief Look at Railroad Communication Vulnerabilities", Proceedings 2004 IEEE Intelligent Transportation Systems Conference Washington, D.C

Federal Railroad Administration "Report of the Railroad Safety Advisory Committee to the Federal Railroad Administrator, Implementation of Positive Train Control Systems" August 1999

Hawkins, Toyn, Bate "A Approach to Designing Safety Critical Systems using the Unified Modeling Language", Critical Systems Development with UML -Proceedings of the UML'03 Workshop, Oct 2003, San Francisco, CA

Helmer, Wong, Slagell, Honavar, Miller, and Lutz, "A Software Fault Tree Approach to Requirements Analysis of an Intrusion Detection System" Proceedings, Symposium on Requirements Engineering for Information Security. Center for Education and Research in Information Assurance and Security, Purdue University, March 2001

Naur, Peter (ed.), "Revised Report on the Algorithmic Language ALGOL 60." *Communications of the ACM*, Vol. 3 No.5, May 1960.

Object Management Group, "UML Superstructure Specification, v2.0 (formal/05-07-04) Aug 2005,

Opdahl & Sindre,"Capturing Security Requirements through Misuse Cases", , Proceedings, NorskInformatikkonferanse2001, Universiteteti Tromsø, Norway, Nov 2001

Vesely & Hassl, "NUREG-0492 Fault Tree Handbook", US Nuclear Regulatory Commission, Jan 1981, US Government Printing Office, Washington DC

Evaluating the Effectiveness of "Turning Traffic Must Yield to Pedestrians (R10-15)" Sign

Ganesh Karkee[1], Srinivas S. Pulugurtha[2], and Shashi S. Nambisan[3]

[1]Transportation Research Center, University of Nevada, Las Vegas, 4505 Maryland Parkway, Las Vegas, NV 89154-4007; Tel. (702) 895-1393; Fax (702) 895-4401; Email: ganesh@trc.unlv.edu

[2]Department of Civil Engineering, University of North Carolina, Charlotte, 9201 University City Boulevard, Charlotte, NC 28223-0001; Tel. (704) 687-6660; Fax (704) 687-6953; Email: sspulugu@uncc.edu

[3]Transportation Research Center, University of Nevada, Las Vegas, 4505 Maryland Parkway, Las Vegas, NV 89154-4007; Tel. (702) 895-1325; Fax (702) 895-4401; Email: shashi@ce.unlv.edu

Abstract

The countermeasure "Turning traffic must yield to pedestrians" (Manual on Uniform Traffic Control Devices (MUTCD) code sign R10-15) is used to address problems such as pedestrians not waiting for signals or an acceptable gap before crossing the streets resulting in conflicts between right turning vehicles and pedestrians. The aim of this paper is to evaluate the effectiveness of this installation. Various measures of effectiveness (MOEs) were identified. These MOEs are pedestrian / vehicle conflict, presence of pedestrians in the crosswalk during the flashing DON'T WALK and during the all red, percent of vehicles blocking the crosswalk, percent of right turn on red drivers coming to a complete stop, percent of turning drivers yielding to pedestrians, percent of pedestrians who look at the start of WALK signal for turning vehicles, pedestrian delay, and vehicle delay. The study site is Harmon Avenue / Paradise Road intersection located in the Las Vegas metropolitan area.

A "before-and-after" data collection strategy was applied to test any significant difference in the identified MOEs between the two study periods. Data were collected during AM and PM peak hours. Statistical tests, test for two proportions and two-sample t-test, were used to test the significance of differences in MOEs during the two study periods. The result shows that motorists yielding behavior while turning either on red or green increased during the after study period. A significant reduction was observed in vehicles blocking the crosswalk while a significant increase in vehicles stopped completely before turning on red ($P < 0.001$). Average pedestrian delay increased during the after study period from 44 sec/pedestrian to 61 sec/pedestrian whereas the average vehicle delay increased from 67 sec/vehicle to 76 sec/vehicle. The installation of R10-15 effective increases the yielding behavior of turning traffic at green in presence of pedestrians which also leads to increase in both pedestrian and vehicle delay.

Introduction

The "Turning traffic must yield to pedestrians" sign (MUTCD, 2004) reminds motorists to yield for pedestrians before turning either on red or green. The Manual on Uniform Control Devices (MUTCD) code for "Turning traffic must yield to pedestrians" sign is R10-15. In this study, this sign was installed at a signalized intersections where right turn on red (RTOR) was permitted. This sign is mounted on the signal post on top of the traffic signal. Figure 1

shows the placement of this sign and crosswalks in south and east approaches of the intersection. The identified problems at Harmon Avenue / Paradise Road are pedestrians not waiting for signals or acceptable gaps before crossing the streets, and conflicts between right turning vehicles and pedestrians. The installation of "Turning traffic must yield to pedestrians" sign was expected to address these identified problems. The sign is expected to alert motorists about the presence of pedestrians downstream of the intersection.

Site Description

The R10-15 sign was installed at the intersection of Harmon Avenue / Paradise Road. A mixed land use pattern is observed around the intersection. Land use is classified as residential, commercial, and recreational (hotels and casinos) proximate to the study location. This intersection is a signalized four-legged intersection. Three approaches, north, south, and east approach, consist of two through lanes in each direction and two left turn pockets. The west approach has two through lanes in each direction and a left turning pocket. Both roads are classified as minor arterials with posted speed limits of 35 mph. Traffic counts, average daily traffic (ADT), along this segment of Harmon Avenue and Paradise Road are 18,000 and 41,000 vehicles, respectively (NDOT, 2004).

Figure 1. Turning traffic must yield to pedestrians (R10-15) sign.

Methodology

A "before-and-after" study was used to evaluate the effectiveness of the R10-15 sign. The "before" condition data were collected without intervention. The "after" condition data were collected nine weeks post installation to avoid novelty effects of the sign. Data were collected for five hours during AM and PM peak hours during each study period. Several measures of effectiveness (MOEs) were identified to quantify pedestrians' and motorists' behavior. These MOEs are as follows:

i. Turning drivers yield to pedestrians: Yielding behavior of motorists while turning either on red or green was recorded. Motorists were scored as "yielding" if a motorist allowed a pedestrian to cross before executing a turn. These observations were converted to the percentage of yielding or not yielding of the total observations.
ii. Vehicles that blocked the crosswalk: A vehicle was scored as blocked the crosswalk if the vehicle occupied approximately 50 percent or more of the crosswalk.
iii. Drivers who turn right on red coming to complete stop: Drivers were recorded as coming to complete stop on right turn on red if they stopped completely before executing RTOR.
iv. Pedestrian delay: Waiting time for pedestrians before they start crossing was recorded.
v. Vehicle delay: Vehicle delay of all approaches of the intersection was recorded.

vi. Pedestrians who looked at start of the walk phase for turning vehicles: Pedestrians were scored in this category if they turned their head over the shoulder and looked for turning traffic when pedestrian signal changed to WALK signal.
vii. Pedestrians in the crosswalk during the flashing DON'T WALK phase: Pedestrians were scored in this category if they were in the crosswalk when the flashing DON'T WALK sign was displayed on the pedestrian signal head.
viii. Pedestrians in the crosswalk at the end of all-red phase: If pedestrians were in the crosswalk during the all-red phase these observed pedestrians were scored in this category.
ix. Pedestrians trapped in the middle of crossing: Pedestrians could not complete crossing in one attempt and has to wait for an acceptable gap in the middle of the road for another attempt.
x. Pedestrian and vehicle conflicts: An event is scored as a conflict if a vehicle or a pedestrian has to change a course or speed due to others action. This observation was recorded in an interaction between a vehicle and a pedestrian.

Statistical tests

The test for two proportions and two sample t-test were used to compare the two sets of data for the two study periods. Most of MOEs were obtained in terms of proportion of the total observations. The test for two proportions was used to evaluate any differences in proportions in the before and after study periods. Two sample t-test was used to evaluate differences in pedestrian delays. These statistical tests were performed at the 95 percent confidence level (α= 0.05).

Null hypothesis (H_0) and alternative hypothesis (H_a) are set whether the values of MOEs in two study periods are significantly different. These hypotheses are as follows:

H_0 = the mean of the MOEs values are the same for both the before and after study period

H_a = the mean of the MOEs values during before study is less than that of the mean during the after study period

The alternative hypothesis is set in such a way that the mean values of the MOEs during the before study is more or less than that of the mean during the after study. The values of MOEs in setting an alternative hypothesis are based on the considered MOEs and their potential impact after installation. H_0 is rejected if P-value is ≤ 0.05 whereas it is not rejected if P-value is > 0.05.

Results

The "before-and-after" study results show that the installation of the "Turning traffic must yield to pedestrians" sign has increased the percentage of motorists yielding at RTOR from 61.3 percent to 73.3 percent (P = 0.156). Similarly, the percentage of motorists yielding at right turn on green increased from 73.5 percent to 76.7 percent (P = 0.615) during the "after-study" period. The installation of the sign, "Turning traffic must yield to pedestrians," shows an increase in motorists yielding while turning either on red or green even though these differences were not statistically significant at the 95 percent confidence level.

Before the installation of the "Turning traffic must yield to pedestrians" sign, about 11 percent of vehicles blocked the crosswalk before executing a turn. However, after the sign was installed, the percentage of motorists blocking the crosswalk reduced to zero ($P < 0.001$). The observed stopping behavior of motorists before RTOR indicates that about 74 percent of motorists stopped completely before the sign was installed. The percentage increased to about 97 percent ($P < 0.001$) after the sign was installed. The values of MOEs during before and after study periods, their difference and statistical significance are shown in Table 1. The value of MOEs in percentage before and after study periods are depicted in Figure 2.

The average vehicle delay increased from 66.8 seconds/vehicle before the installation of the sign to 75.6 seconds/vehicle after the installation of the sign. This could be due to increase in the percentage of motorists yielding to pedestrians. It is expected that pedestrian delay would decrease as motorists' yielding behavior to pedestrians has improved after the sign was installed However, data show that pedestrian delay increased from 44 seconds/pedestrian before the installation of the sign to 61 seconds/pedestrian after the installation of the sign. The reason is not known as to why pedestrian delay has increased after the installation of the "Turning traffic must yield to pedestrians" sign.

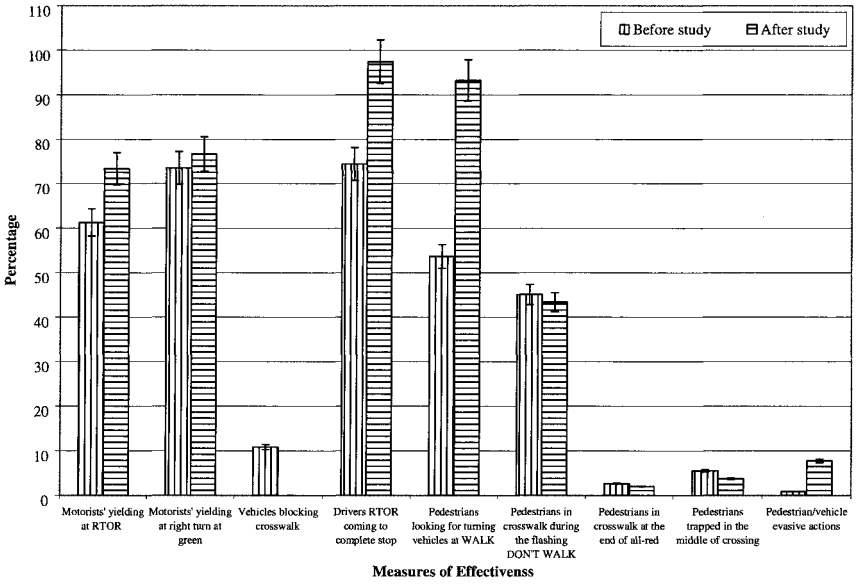


Figure 2. Difference in MOEs during before and after study periods.

The sign, "Turning traffic must yield to pedestrians," is intended for motorists. However, the before-and-after study result indicates some positive influence on pedestrians' crossing behavior. The percentage of pedestrians looking for turning vehicles at the beginning of the WALK signal increased from 53.7 percent before the installation of the sign to 93.2 percent after the installation of the sign ($P < 0.001$). As the motorists' yielding increases, motorists might stop upstream of the crosswalk. Therefore, more pedestrians watch for turning vehicles before crossing. Marginal differences were observed in the percentage of pedestrians who were in the crosswalk during the flashing DON'T WALK phase and during the all red phase before and after the installation of the sign. The percentages of pedestrians who were in the crosswalk during the flashing DON'T WALK phase and the all-red time decreased by 1.7 percent ($P = 0.586$) and 0.6 percent ($P = 0.525$), respectively after the installation of the sign.

Table 1. MOEs during Before and After Study Periods and their Statistical Significance.

S. No.	Measures of Effectiveness	Before		After		(Before - After)	P-value	Null hypothesis
		Sample size	Value	Sample size	Value			
1	Motorists' yielding at right turn on red (in the presence of pedestrian at turn or approach), %	31	61.29	30	73.33	-12.04	0.156	Do not reject
2	Motorists' yielding at right turn on green (in the presence of pedestrians), %	102	73.53	90	76.67	-3.14	0.615	Do not reject
3	Percentage vehicles blocking crosswalk	129	10.85	235	0.00	10.85	0.000	Reject
4	Percentage of drivers executing right turn on red coming to complete stop	129	74.42	235	97.45	-23.03	0.000	Reject
5	Pedestrian delay (sec/ped)	556	44.37	355	61.09	-16.73	0.000	Reject
6	Vehicle delay at intersection (sec/veh)	1,356	66.83	1,275	75.64	-8.81	N/A	N/A
7	Percentage of pedestrians who looked at start of the WALK phase for turning vehicles	542	53.69	370	93.24	-39.55	0.000	Reject
8	Percentage of pedestrians who were in the crosswalk during the flashing DON'T WALK phase	639	45.07	390	43.33	1.74	0.586	Do not reject
9	Percentage of pedestrians who were in the crosswalk at the end of all-red	639	2.66	390	2.05	0.61	0.525	Do not reject
10	Percentage of pedestrians who were trapped in the middle of crossing	618	5.50	373	3.75	1.75	0.194	Do not reject
11	Percentage of pedestrian/vehicle evasive actions, change course/slow to avoid motorists	609	0.82	349	7.74	-6.92	1.000	Do not reject

Note: $\alpha = 0.05$

Observations also showed an increase in the percent of motorists turning on permitted left-turn yielding to pedestrians. This has led to a decrease in the percentage of pedestrians trapped in the middle of the road after installation of the sign. Data indicate that the percentage of pedestrians who were trapped in the middle of the road while crossing decreased from 5.5 percent during the before study to 3.7 percent ($P = 0.194$) during the after study.

The percentages of evasive actions were 0.8 percent and 7.7 percent before and after condition data collection period, respectively. The difference in the percentage of evasive action between before and after period was significantly different ($P < 0.001$) at the 95 percent confidence level but in an unexpected direction.

Conclusions

The installation of the sign "Turning traffic must yield to pedestrians" led to an increase in motorists' yielding behavior while turning either on red or on green. Even though the sign, "Turning traffic must yield to pedestrians," was intended for motorists, some improvements in pedestrians' crossing behaviors were also observed. These behaviors include: decrease in the percentages of pedestrians in the crosswalk during the flashing DON'T WALK and at the all-red time, a decrease in the percentage of pedestrians trapped in the middle of the road while crossing, and an increase in the percentage of the pedestrians looking for turning vehicles. Overall, observations indicate that installation "Turning traffic must yield to pedestrians" sign at intersection improves safety and yielding behavior of both motorists and pedestrians.

Acknowledgments
The authors thank the Federal Highway Administration (FHWA), the Regional Transportation Commission (RTC) of Southern Nevada, the City of Las Vegas, the Clark County Department of Public Works (CCDPW), the Nevada Office of Traffic Safety, and the Nevada Department of Transportation (NDOT) for their financial support. The authors also acknowledge staff of these agencies for their technical support and guidance. The students of the University of Nevada, Las Vegas – Transportation Research Center are thanked for their help and support in collecting data for the project.

References

Nevada Department of Transportation. Traffic Information Division. (2004). 2003Annual Traffic Report. Nevada Department of Transportation, Website: http://www.nevadadot.com/reports_pubs/traffic_report/2003/pdfs/Clark.pdf, Accessed on September 27, 2004.

United States Department of Transportation. Federal Highway Administration. (2004). Manual on uniform traffic control devices for streets and highways 2003 edition. United States Department of Transportation, Federal Highway Administration Web site: http://mutcd.fhwa.dot.gov/pdfs/2003r1/pdf-index.htm, Accessed on October 14, 2004.

Performance Evaluation of Parallel Genetic Algorithms for Adaptive Transit Signal Priority

Guangwei Zhou[1], Albert Gan[2], Chengjun Zhan[3], and Xiaoxia Zhu[4]

[1] Lehman Center for Transportation Research, Florida International University; PH (305) 348-4103; FAX (305) 348-2802; email: gzhou002@fiu.edu
[2] Lehman Center for Transportation Research, Florida International University; PH (305) 348-3116; FAX (305) 348-2802; email: gana@fiu.edu
[3] Lehman Center for Transportation Research, Florida International University; PH (305) 348-4103; FAX (305) 348-2802; email: czhan003@fiu.edu
[4] Lehman Center for Transportation Research, Florida International University, PH (305) 348-3171; FAX (305) 348-2802; email: xzhu004@fiu.edu

Abstract

Adaptive Transit Signal Priority (TSP) is increasingly becoming a dominant form of preferential treatment for transit vehicles along urban arterials. The optimization of real-time adaptive traffic signals is both complex and demanding. Conventional optimization methods, including calculus-based, enumerative, and random search methods, lack both the speed and robustness needed for such applications. Parallel Genetic Algorithms (PGAs) have the potential to overcoming these obstacles. This paper applies PGAs to optimize an adaptive strategy for TSP. The results show that the PGAs can offer more efficient and faster optimization for the adaptive TSP strategy in terms of convergence speed and required computation resources, especially for the complex case of congested traffic conditions.

Introduction

Transit Signal Priority (TSP) is a preferential strategy designed to help transit vehicles to pass through signalized intersections with less delay by modifying the normal signal operation. Existing TSP strategies can be categorized as passive, active, and adaptive. Passive priority operates continuously regardless of whether there is a transit vehicle arrival, while active (or actuated) priority is responsive to transit service requests based on pre-determined signal priority designs. Current research on TSP has focused predominantly on adaptive real-time strategies, which address the "on-time" response of signal timing to not only transit vehicles, but also the real-time traffic conditions, by applying real-time optimization based on some performance criteria, such as delay and vehicle stops.

The optimization of real-time adaptive traffic signals is both complex and demanding. Conventional optimization methods, including calculus-based (such as hill-climbing), enumerative, and random search methods, lack both the speed and robustness needed for such applications. This has led to the use of Genetic Algorithms (GAs)—a machine-learning search method based on the mechanics of nature selection and natural genetics (Goldberg, 1989). Traditional GAs (i.e., simple GAs) suffer from slow convergence and are not suitable for real-time applications. Advances in the field of evolutionary computation now hold potential to overcoming some obstacles of simple GAs. Parallel GA (PGA), which is able to structure the populations into a number of subpopulations running in parallel on either a

single processor or across multiple processors, can provide more efficient and faster solutions to complex transportation problems such as this.

The objective of this research was to evaluate the performance of PGA for adaptive TSP. To achieve the objective, the adaptive TSP strategy was first designed and the corresponding models established. The optimization problem was then coded into PGA to search for optimal solutions. The PGA performance was evaluated at different traffic congestion levels in terms of convergence speed and required computation resources, based on which the conclusions were drawn.

Adaptive TSP Optimization Strategies and Model Formulation

Conventional TSP strategies, including passive and active TSP, aims to provide a green signal for transit vehicles upon their arrivals at a signalized intersection. The basic objective behind these strategies can be stated as a binary logical decision process, as follows:

1. If possible, allow transit vehicles to go through an intersection without stopping;
2. Otherwise, make their waiting time as short as possible.

This basic objective is realized by arbitrarily interrupting the signal timing plans intended to optimize the intersection or system-wide operation, such as extending the green time and early terminating the red time for approaches with transit arrivals. To reduce the negative impact of such interruptions on the existing signal design, additional strategies such as green time reimbursement and coordination recovery are sometimes implemented. For the adaptive TSP strategy proposed in this study, the objectives described above for transit operation are integrated into the objectives for the system-wide level. As such, the adaptive TSP would attempt to achieve equilibrium between the system level and the user level (transit vehicles).

The common approach to providing priority to transit vehicles has been to assign a weighting factor to the priority approach or movement as part of the optimization objective function. The delay estimation in such an objective function is generally based on either macroscopic analysis models such as the HCM delay model, or mesoscopic models based on traffic streams. However, assigning a global weighting factor to the entire approach or movement tend to overstate the importance of the transit approach or movement. To avoid this problem, this study used a microscopic delay estimation model that is based on individual vehicle units. The weighting strategy was applied to only individual transit vehicles that requested TSP service. The delay estimation models for the subject network can be formulated as follows:

$$AD = \frac{1}{m}\sum_{i}\sum_{j}\sum_{k}\sum_{m}\left[(1+w(t))(D_{k,m}^{i,j}(t) - A_{k,m}^{i,j}(t))\right] \quad (1)$$

where

AD = average vehicle delay for the subject network,
i = number of intersections included in the network,
j = number of signal phases in timing plan for an intersection,
k = number of lanes included in a signal phase for an intersection,
m = number of vehicles arriving at the intersection during the optimization horizon,
w(t) = dynamic weighting factor assigned for arriving transit vehicles,

D = departure time of individual vehicles from intersection stopline, and
A = ideal arrival time of individual vehicles at the intersection stopline in absence of signal control.

Genetic Algorithm Solution

In this study, genetic algorithm was used to search for the optimal or near-optimal solution for the problem stated above. The objective was to minimize the total delay of the intersection, subject to various constraints, such as minimum green times, minimum and maximum cycle lengths, allowable queue accommodation space, and so on.

Minimize $$AD = \frac{1}{m}\sum_i\sum_j\sum_k\sum_m\left[(1+w(t))(D_{k,m}^{i,j}(t) - A_{k,m}^{i,j}(t))\right]$$

Subject to:

$$\sum_j GR_j = 1$$

$$G_j \geq \min G_j$$

$$\min C \leq C \leq \max C$$

$$\sum_j \min G_j = \min C$$

$$MaxQ_j^k(t) < AllowQ_j^k(t)$$

$$G_j, C > 0$$

$$x \in X \text{ (set of condidate phasing plans)}$$

A crucial step toward applying GA for adaptive TSP optimization problem is to formulate the genetic representation (i.e., coding scheme) for potential solutions. The following chromosome structure was constructed to represent the decision variables to be optimized:

{ Phase sequence;	cycle length;	green split 1;	green split 2;	.....	green split k; }
{(01100...);	(01100...);	(10100...);	(01011...);	.....	(01110...) }
{ PhI;	CI;	GI_1;	GI_2;	.....	GI_k; }

The string length for each decision variable depends on the optimization resolution for each variable. Before the decision variables can be substituted into the objective function for the calculation of the fitness value, they have to be decoded with a chromosome structure. The following decoding scheme was developed to convert the decision variables from integers to the units required by the parameters of signal control:

$$x = PhI \bullet \frac{\max PhNum}{2^{m_i} - 1} \tag{2}$$

$$C = \min C + (\max C - \min C) \bullet \frac{CI}{2^m - 1} \tag{3}$$

$$G_j = \min G_j + (C - \min C) \bullet \frac{GI_j}{\sum_j GI_j} \tag{4}$$

where maxPhNum, minC, maxC, and $\min G_j$ represent the maximum number of phases, the allowable minimum signal cycle, maximum signal cycle, and minimum phase duration including yellow-plus-all-red intervals for phase j, respectively; and PhI, CI and GI_j represent

the integers converted from the binary chromosome strings for phasing plan number, cycle length, and green split j, respectively.

Through the process of decoding, some constraints were incorporated into the system model. However, the inequality constraint for queue accommodation could not be incorporated into the system model directly. A penalty function was introduced into the objective function to transform the constrained optimization problem into an unconstrained problem shown in Equation (5). In this equation, r is the penalty coefficient, and $\Phi[\ ]$ is the penalty function.

$$\text{Minimize } f(i, j, k, m, t) = AD(i, j, k, m, t) + r \cdot \sum_{j} \Phi[\text{AllowQ}_{j}^{k}(t) - \text{MaxQ}_{j}^{k}(t)] \quad (5)$$

Since genetic algorithms only maximize the fitness value, the model in Equation (5) has to be transformed into a maximization problem by subtracting the minimization function from a coefficient, CO_{max}, that carries large value:

$$\text{Maximize Fitness Value} = CO_{max} - f[i, j, k, m, t] \quad (6)$$

Island parallel GA is a type of parallel GA developed by Goodman (1996) and was used as the search engine for the optimal solution to the adaptive TSP problem. Island parallel GA is capable of simulating in parallel many sub-populations by dividing the population for single or simple GA into small clusters or islands, each of which operates as an individual simple GA. Periodically the individuals from an island or a sub-population are allowed to migrate to other islands or sub-populations. A notable advantage of island parallel GA is its ability to reduce the chance of premature convergence to a local optimum—a classical problem incurred by the simple GA. More information on island parallel GA can be found in Goodman (1996) and Chen (2004).

Experimental Design

The performance of PGAs was evaluated in this study by compared with simple GAs under both under- and over-saturated traffic conditions for an isolated signalized intersection. In the application of GA for adaptive traffic optimization problems, the major concerns are its convergence speed and the required computation resources. These three types of GAs were compared in terms of convergence speed and the required computation resources: simple GA, parallel GA with 4 population islands, and parallel GA with 8 population islands The input for the major GA parameters was listed in Table 1. The topologies for parallel GA with 4 and 8 populations are a ring structure shown in Figure 1. The intersection was tested for all the three GAs for both under- and over-saturated levels.

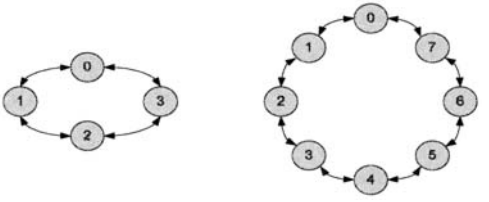


Figure 1. Topologies for 4-Population and 8-Population PGAs

Table 6.1: Input Values for Major Parameters of GAs

	Simple GA	PGA with 4 Populations	PGA with 8 Populations
Number of populations	1	4	8
Sub-populations	120	30	16
Maximum generation	200	200	200
Crossover probability	0.8	0.8	0.8
Mutation probability	0.005	0.005	0.005

Results

Figures 2 and 3 show the variations of the best individual's fitness value for different GAs over the generations of evolution, respectively, for under- and over-saturated intersection traffic volumes. As shown in Figure 2, the parallel GAs with four populations or with eight populations can provide a much faster convergence when compared with simple GA (GA with one population). However, the parallel GAs cannot improve the fitness value significantly.

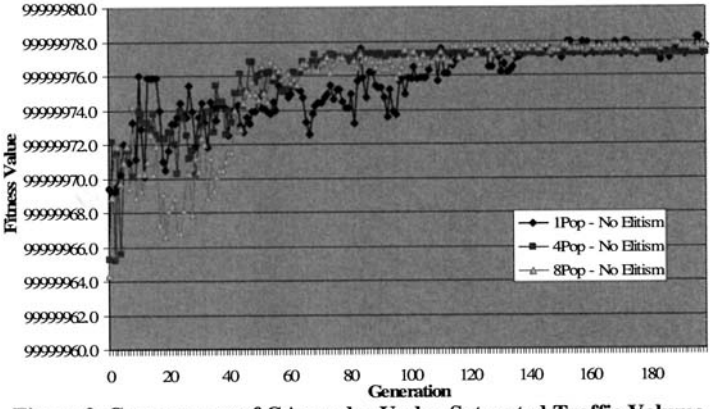


Figure 2. Convergence of GAs under Under-Saturated Traffic Volume

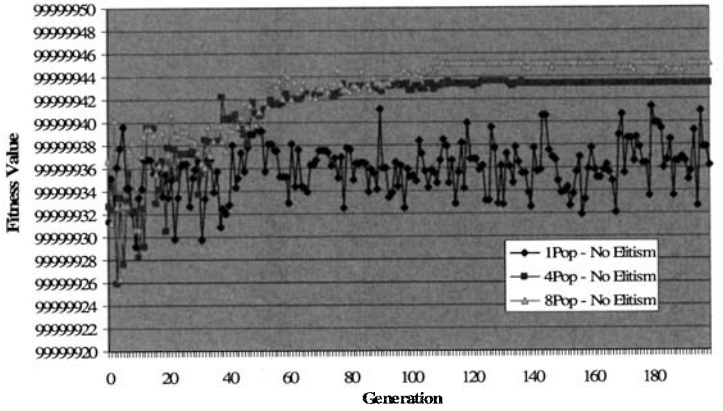


Figure 3. Convergence of GAs under Over-Saturated Traffic Volume

After the experiment tests for under-saturated traffic condition, it was stipulated that the insignificant improvement in fitness value of PGAs might be contributed to the not-so-hard nature of the optimization problem for under-saturated traffic condition. Therefore, the same experiments were designed for over-saturated traffic conditions, and tested for the three types of GAs. The experiment results for over-saturated traffic condition were illustrated in Figure 3. This figure shows that parallel GAs not only converge faster but also achieve a much higher fitness value than the simple GA. The reason for this is understandable. For under-saturated traffic conditions, the solution to the adaptive signal control problem is not strongly constrained by the limitation on queue accommodation in the objective function. It is relatively easy to find a solution that does not violate that constraint. However, for over-saturated traffic conditions, the constraint of queue accommodation limitation is very easy to be violated. It is difficult to even find a feasible solution to the over-saturated problem. The complex nature of over-saturated traffic optimization problem leads simple GA to a random walk during the search of optimal solutions.

To compare the computation resources required by each GA, the same calculation loads were tested for the three types of GAs in the same laptop computer with AMD Turion 64 mobile processor, 1.60 GHz CPU, and 1G RAM. Figure 4 shows that the parallel GA requires much less computation resources under the same calculation load, and thus can provide more efficient searching. When compared with simple GA, PGA with four populations saves 16% of CPU time, and PGA with eight populations save 46% of CPU time.

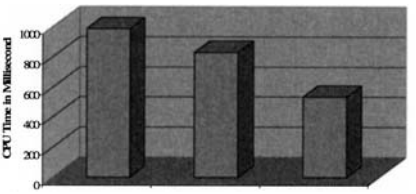


Figure 4. CPU Time Required by Each Run of Different GA Types

Conclusions

This paper applies parallel genetic algorithms (PGAs) to optimize an adaptive strategy for transit signal priority (TSP). The results show that, when compared with simple GA, PGAs can offer more efficient and faster optimization of adaptive TSP strategy in both convergence speed and required computation resources, especially for the more complex cases that involved congested traffic conditions.

References

Golderg, D. E. (1989), "Genetic Algorithms in Search, Optimization, and Machine Learning." Addison Wesley.

Chen, H., G. Abi-Lebdeh, and E., Goodman (2004), "Improving Performance of Genetic Algorithms for Transportation Systems: The Case of Parallel Genetic Algorithms", Compendium of Papers for the 83rd Transportation Research Board Annual Meeting, Washington, D.C.

Goodman, E. (1996), An Introduction to GALOPPS, the "Genetic Algorithm Optimized for Portability and Parallelism," Release 3.2, (software and User's Guide), http://garage.cps. msu.edu/software/galopps.

Augmented Reality Applications to Traffic Operations

Ghada S. Moussa[1], Essam Radwan[2], and Khaled F. Hussain[3]

ABSTRACT

A useful and very significant leap forward in simulation technology is to be able to evaluate synthetic simulated conditions in realistic settings. This technology is based on Augmented Reality "AR". Augmented reality is a paradigm which creates a combination of real and virtual objects in REAL TIME (for example in movies they combine real and virtual objects but not in real time) in which the user cannot tell the difference between the real and augmented world. The AR technology can offer a very realistic environment for driving enhancement as well as driving performance testing under different scenarios. This can be achieved by adding virtual objects (people, vehicles, hazards, and other objects) to the normal view while driving an actual vehicle in a safe environment.

This paper documents a summary of AR technologies on the market, illustrates the different components of the **A**ugmented **R**eality **V**ehicle **S**ystem **"ARV System"** built at the University of Central Florida, and discusses the validation and application of such system to traffic engineering research. The ARV System is a promising tool for transportation research for improving roadway safety and decreasing the number of traffic collisions.

[1]Ph.D. Student, Department of Civil & Environmental Engineering, University of Central Florida, Orlando, FL 32816-2450; PH (407) 823-1056; FAX (407) 823-3315; email: ghada@mail.ucf.edu

[2]Ph.D., P.E., F. ASCE, Professor , CATSS Executive Director, Department of Civil & Environmental Engineering, University of Central Florida
Orlando, FL 32816-2450; PH (407) 823-4738; FAX (407)823-4676; email: aeradwan@mail.ucf.edu

[3]Ph.D., Assistant Professor, School of Computer Science, University of Central Florida, Orlando, FL 32816-2362, PH (407) 823-4452; FAX (407) 823-5419; email: khaled@cs.ucf.edu

Introduction

The fidelity of any traffic experiment's results is greatly based on driving feeling. A great part of the driving feeling depends on the driver's interaction continuously with the vehicle's steering wheel, braking, and gas pedals as well as the surrounding environment (Johansson & Nordin 2002). Performing a real experiment to evaluate human performance under certain traffic scenarios might be very expensive with high degree of risk to drivers. With driving a real vehicle in a safe controlled environment, we will guarantee the fidelity of the experiment's result as well as the safety to drivers. This is the main idea of using the Augmented Reality Vehicle System "ARV System" built at the University of Central Florida. Our ARV System can provide a very realistic environment for testing human performance under different scenarios while driving a real vehicle. Our ARV System is based on the Augmented Reality "AR" technology. The basic idea of the AR technology is to add virtual (computer-generated) objects, audio and other sense enhancements to a real-world scene in real-time (Hussain & Kaptan 2004). These enhancements can help users to perform real-world tasks (Azuma 1997). Applying the AR technology while driving an actual vehicle, can reduce hazards and costs, compared to real experiments as well as providing high degree of motivation to the driver. That makes applying the AR technology into traffic engineering applications a promising approach. Figure 1 shows an example of a view that the user might see from an AR system showing a live scene with a virtual vehicle.

(a) (b)

Figure 1 a) live scene, b) the result of inserting a virtual vehicle into the live scenes.

The Augmented Reality Technology

In AR technology, real world and virtual objects are combined in a real time. There are two main techniques for combining real and virtual objects; optic technique and video technique (Johansson & Nordin 2002). While optic technique uses an optical combiner for combining real and virtual objects, video technique uses a computer or a video mixer for combining video of the real world, from video cameras, with virtual images (computer-generated) (Azuma 1997). Both AR techniques (optic and video)

can display the final view to the user using a Head Mounted Display "HMD", monitor-based display, and/or hand-held display.

There is a wide range of applications in many areas has been developed using the AR technology such as; medical visualization and training, mechanical maintenance and repair, manufacturing and assembling, military training, and commercial applications (Hussain & Kaptan 2004). The common thing between all AR applications is the requirement to align virtual images with objects in the real scene.

Research Objectives

Applying the Augmented Reality "AR" technology into traffic engineering area is a new and challenging task. Our main goal is to investigate the feasibility of applying AR technology into traffic engineering area. In order to achieve that goal, the following objectives are defined:

- Explore the ability to adapt the AR technology to road driving
- Study drivers' behavior under AR, and on-the-road situations

To achieve these objectives, the following tasks are needed:

1. **Build** a new tool "ARV System", based on the video see-through AR technology.
2. **Evaluate** the ARV System.
    a. Evaluate the outcomes from the system, and
    b. Evaluate the system's effects on the driver's performance.
3. **Use** the ARV System to perform traffic studies.
    a. Left-turn gap acceptance, and
    b. Horizontal visibility blockage

The Augmented Reality Vehicle System

Our AR Vehicle System "ARV System" is based on the video see-through HMD technology. The ARV System has three main components connected to a powerful computer. These components are video camera, HMD, and GPS. The ARV System can be installed in any vehicle as shown in Figure 2, where the driver wears a flip-able HMD while driving a real vehicle in a real safe environment. In our experiment, the ARV System will be installed in a dual controlled vehicle. The dual controlled vehicle is a vehicle that can be controlled by any of two persons; the driver or the one sitting beside him/her. The safe controlled environment is a rented racetrack. During the experiment, the experiment's participant will be allowed to drive the dual controlled vehicle while wearing the flip-able HMD. Through the HMD, participant will see an augmented video. The augmented video is a combination of an in-time video of the real surrounding road and virtual images of vehicles, traffic signs, traffic light, buildings, trees, and other objects depending on the scenario. The combination

is done through the computer so that the driver will not be able to tell the difference between the real and virtual object in the scene. At any time of the experiment participant is able to withdraw from the experiment. The participant can flip the HMD and see through the front windshield of the vehicle the real surrounding road and he can stop the vehicle either by himself or by the second driver sitting beside him.

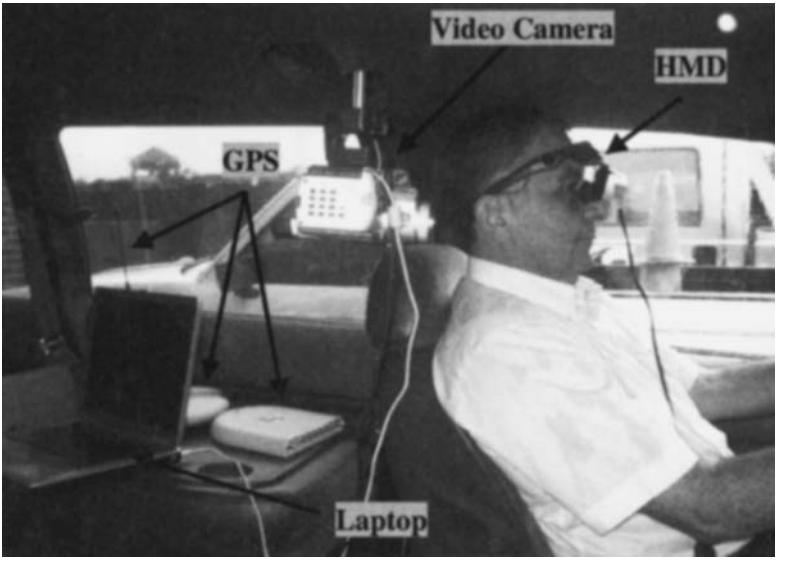


Figure 2 the ARV System installed in a vehicle

Evaluate the Augmented Reality Vehicle System

As any new tool, before using the ARV System in real traffic studies it is very important to evaluate it. In evaluating the ARV System, two stages needed to be considered;

1. evaluate the outcomes from the system
2. evaluate the system's effects on driving performance

In evaluating the outcomes, we need to make sure that the outcome from this new tool is trustable and can be used for real world applications. The ARV System has to be able to

a. duplicate real scenes
b. generate new scenes

As synthetic images are augmented to the real view, in real time, augmenting those images has to be done in a way that they are well aligned with real objects in the real view. By comparing real-world's images with their corresponding augmented images "final view", we will be able to verify that.

In evaluating system's effects on the driving performance, a human factor study will be conducted to study if there is a statistically significant difference between; driving the vehicle without the ARV System, and driving the vehicle with the ARV System. Drivers will be asked to perform some tasks while driving which can be used to investigate certain issues as follow;

a. stopping the vehicle at a certain point **(for distance judgment)**
b. driving at a constant speed **(for speed judgment)**
c. lane Keeping **(for driving comfort)**

In addition, drivers will be asked general questions regarding the ARV System realism. The questionnaire will include questions regarding quality of images, whether they felt any motion sickness while driving or not, whether they felt safe while driving or not. As well as asking them about their opinion of how the ARV System affected the way they felt and drove, and is it safe tool for future research.

Proposed Traffic Studies using the ARV System

Two traffic studies will be performed using the ARV System installed in the dual controlled vehicle; left-turn gape acceptance and horizontal visibility blockage.

1. Left-Turn Gap Acceptance

When left turner vehicles don't have the right-of-way as in unprotected left-turn lane at two way stop controlled intersections, left turning become a complex and driver decision procedure. Drivers in left-turn lanes need to find acceptable gaps to cross the opposing through traffic. Misjudged gaps might cause serious accidents and/or high intersection delay.

We will study left turning behavior from major road to minor road through oncoming (opposing) traffic at a two way-stop controlled intersection. Effects of the left-turn driver' age and gender as well as the opposing vehicles' speed on the left turning maneuver and gap acceptance will be studied. Drivers will drive the dual controlled vehicle in a real safe road (a rented racetrack) while seeing, through the HMD, an augmented world with upcoming virtual vehicles in the opposing direction (with different gaps between them).

2. Horizontal Visibility Blockage

One of the main real life problems is the horizontal visibility blockage. The horizontal visibility blockage happens when the driver's view to his left and/or right is blocked. This can occur when someone is driving a passenger car, such as Saturn, Honda Accord, Nissan Sentra, or Ford Taurus, closely behind a Light Truck Vehicle "LTV", such as van and sport utility vehicles "SUVs". A sudden stop of the LTV, in the shadow of the blindness of the following car driver, may cause high probability of a rear-end collision.

We will study the contribution of Light Truck Vehicles "LTVs" to potential rear-end collisions at un-signalized intersection. Also, drivers' behavior, including speed and gap maintenance when driving behind LTVs and behind passenger cars will be studied. Then the effect of adding an additional mirror on the side of the road on solving horizontal visibility problem will be

evaluated. Drivers will drive the dual controlled vehicle in a real safe road (a rented racetrack) while seeing, through the HMD, an augmented world with virtual surroundings, vehicles, and mirror.

For the chosen studies, appropriate measures will be collected to capture drivers' behavior and responses to these settings. Those measures include speed variance, lane keeping, response time, acceleration/deceleration rates, gap acceptances and gap rejections, and crash occurrence if any. The collected data will be analyzed using statistical methodologies and conclusions will be drawn from these experiments. A minimum of twenty-five subjects of both sexes varying in age from 18 to the mid-50's will participate in the study. A preliminary study or a pilot study will be done with this design and necessary changes would be incorporated for the final run of experiments.

Discussions and Conclusions

At this stage of the on-going research, the video camera is fixed as close as possible to the HMD (the driver's eyes) as shown in Figure 2 to get the same driver's prospective view of the real world. In the mean time, we assume that the driver is not moving his head. In case of the drive's head is moving a driver's head tracker need to be used.

The realism of the final view that the driver sees is influenced by the quality of the virtual object. In order to get more illusion 3-D virtual object, high quality 2-D images of virtual object, from different viewpoints can be recorded. Then the morphing technique can be applied to get 2-D images from any viewpoint. Afterward, these images can be augmented to the real scene to create the illusion of the 3-D. Also, fast movements of virtual objects can be more realistic by adding motion blur to the animation. Another factor that affects the final view's realism is the sun lighting effect. This effect can be considered by using the sun lighting model for casting shadows and indirect light on the objects in the image.

References

Azuma, R. (1997). "A survey of augmented reality." *Teleoperators and Virtual Environments*, 6 (4), 355-385.

Hussain, K. F. and Kaptan, V., (2004). "Modeling and simulation with augmented reality." *International Journal on Operations Research*, 38 (2), 89-103.

Johansson, M. and, Nordin, J. (2002) *A survey of driving simulators and their suitability for testing Volvo cars*, Department of Machine and Vehicle Systems, Chalmers University of Technology, Göteborg, Sweden.

Bayesian Estimation of Statewide OD Flows using Link Volumes Estimated from Combined Information in Remotely Sensed Data and Ground Counts

Prem K. Goel[1], Mark R. McCord[2], Shiling Ruan[1], and Morton O'Kelly[3]

[1]Dept. of Statistics; The Ohio State University, 404 Cockins Hall, 1958 Neil Ave., Columbus, OH 43210; PH (614) 292-8110; FAX (614) 292-2096; email: goel.1@osu.edu; ruan@stat.ohio-state.edu

[2]Dept. of Civil and Env. Engineering and Geodetic Science; The Ohio State University, 470 Hitchcock Hall, 2070 Neil Ave. Columbus, OH 43210; PH (614) 292-2388; FAX (614)292-3780; email: mccord.2@osu.edu

[3]Dept. of Geography, The Ohio State University, 1036 Derby Hall, 154 North Oval Mall, Columbus, OH 43210; PH (614) 292-8744; FAX (614) 292-6231; email: Okelly.1@osu.edu

Abstract

Vehicles are detectable in high-resolution satellite and airborne imagery of highway segments. Given the increasing availability of such remotely sensed data and recent results on estimating annual average daily traffic (AADT) and quantifying their uncertainty, the image-based data could conceivably be useful for other transportation planning purposes. In this work we develop a Bayesian methodology for estimating statewide annually averaged origin-destination (OD) flows using link-level AADT estimates obtained from remote sensed imagery as well as from traditional ground counts. We demonstrate the methodology, which accounts for different levels of uncertainty involved, using a simulated set of statewide intercity OD flows and a highway network, and quantify the benefits of adding increased levels of remotely sensed data.

Introduction

With the increase in statewide travel demand models, statewide origin-destination (OD) flows are of increasing importance. In this problem, link-level traffic volumes are realized as vehicles travel from origins to destinations on intercity highway links. Thus the link volumes contain information on underlying OD flows and can conceivably be used in their estimation. In the OD estimation problem, a study area can be divided into n zones, connected by a network of links. Let μ_{ij} denote the OD flow between zone i and zone j, where $i, j = 1, \ldots n$ (the intrazonal trips can be safely disregarded). Let h_{ij}^a denote the known proportion of OD flow between zones i and j using link a. If the traffic volume on the link a is V_a, then the OD flows and link volumes satisfy (Maher, 1983):

$$\sum_i \sum_j h_{ij}^a \mu_{ij} = V_a, \quad \forall a. \tag{1}$$

The n(n-1) values of $\{\mu_{ij}\}$ stacked in a column-vector is referred to as the OD vector μ. For the link a, when the values of $\{h_{ij}^a\}$ are stacked in row-vector h_a, in the same order as the $\{\mu_{ij}\}$ in the vector μ, then (1) can be expressed as $h_a \mu = V_a$. Given the observed volumes V on a collection of links, stacked in a column vector, and the corresponding row vectors h

stacked in a matrix H, the observation equations (1) can be expressed as $H\mu = V$. The matrix H is called the incidence matrix for the observed link volumes. The OD estimation problem is a difficult "inverse" problem of finding a vector μ satisfying the observation equations. In most cases, the number of OD pairs is much bigger than the number of observed link volumes, thus there is no unique solution. In addition, the volume estimates themselves may have estimation errors, which may lead to an inconsistent system of observation equations.

It is obvious from (1) that accurate OD estimation depends on the quality of link volume information $\{V_a\}$. A commonly used summary measure of link volume is annual average daily traffic (AADT) (FHWA, 1992; McShane, *et al.*, 1998). In general, two types of AADT estimates are available from ground-based link count data (see, *e.g.,* McCord, *et al.*, 2002). The first is from (permanent) automatic traffic recorders (PATRs). Daily (24-hour) volumes available from PATRs for every day of the year, with possibly a few missing values, lead directly to AADT for that segment. These estimates are considered to be the most accurate, containing very little estimation uncertainty. The second type of estimates are based on a sample from movable automatic traffic recorders (MATRs), which are used to collect traffic counts for typically two consecutive 24-hours volumes.

Besides ground-based traffic recorders, satellite and air-borne imagery might also provide traffic flow information. Many state Departments of Transportation (DOTs) regularly collect air-based images for mapping and inventory purposes that could be made available for volume estimation. However, an image represents a snapshot of traffic conditions at an instant in time, and image-based volume estimates will be noisier than those obtained from ground counts. Here we are interested in using AADT estimates from the ground- and image-based data to estimate an average OD matrix and to determine if the noisier image-based data can improve the OD estimates. We employ a methodology that appropriately accounts for the different levels of uncertainty in the AADT estimates based on the diverse sources of information. With a simulated set of intercity data, we quantify the benefits of adding increasing levels of remotely sensed image data.

Statistical Methodology

We adopt a Bayesian approach (Maher, 1983) to estimate the ODs. Let μ denote the unknown annually averaged ODs vector. Let μ_0, C_0 be the mean and co-variance matrix of the prior distribution of μ, which would be based on surveys and past information. Assume $\mu \sim MVN(\mu_0, C_0)$, where *MVN* denotes a multivariate normal distribution. Let V denote the vector of AADT estimates on a set of links. Using (1), let $V = H\mu + \varepsilon$, where H is the incidence matrix corresponding to these set of links, and $\varepsilon \sim MVN(0, \Sigma)$ represents the uncertainty assessment associated with the AADT estimates. Then the posterior distribution of μ given observation $V = v$ is also *MVN* with mean μ_1 and variance C_1 given by

$$\begin{aligned} \mu_1 &= \mu_0 + C_0 H'(\Sigma + HC_0H')^{-1}(v - H\mu_0) \\ C_1 &= C_0 - C_0 H'(\Sigma + HC_0H')^{-1} HC_0 \end{aligned} \tag{2}$$

This approach has several advantages. It utilizes prior information on the OD flows in the estimation process– *e.g.,* previous OD estimates, and recognizing that the AADT estimates from different sources of data may have possibly different level of uncertainty, provides a way to combine AADT estimates produced from several sources of data in a unified

approach. In addition, it provides an assessment of uncertainty associated with the posterior distribution that can be used for decisions based on the ODs.

As mentioned earlier, information in imagery containing snapshots of traffic on highway links can be used to estimate the link AADTs (McCord et al., 2003). Compared to the traditional method of estimating AADT from MATR samples, the image-based estimates contain additional uncertainty because the instantaneous information in the snapshot needs to be expanded to daily volumes (Jiang, 2005). This uncertainty is captured by the coefficient of variations cv_g and cv_s of AADT estimates obtained from 24-hour ground counts and images, respectively. Assuming that an AADT estimate on a link sampled with an MATR is based on two 24-hour counts on the link, and that an image-based AADT is obtained from m independent images of the link, the variances of these AADT estimates are given by

$$\begin{aligned} \text{var}(AADT^{MATR}) &= 1/2 * AADT^2 * cv_g^2 \\ \text{var}(AADT^{img}) &= 1/m * AADT^2 * cv_s^2. \end{aligned} \tag{3}$$

In this paper, we use $cv_g = 0.15$ and $cv_s = 0.25$, values based on empirical studies in McCord *et al.* (2003) and Jiang (2005). Using these values, it follows from (3) that the variance of an AADT estimate, based on $m = 5.55$ images, is the same as the variance of an estimate based on a two-day MATR sample. Thus, approximately 6 images on a link seem to provide AADT information equivalent to one traditional MATR sample. Of course, with other assessed estimates of uncertainty in AADT estimates, this number would be different. The AADT estimates based on PATR observations are considered accurate, and we assume zero variance associated with the PATR data.

Network-based Numerical Study

State DoTs have a large inventory of imagery of highway segments, usually obtained for mapping purposes, which could be used when estimating OD flows from link volumes. Above, it was indicated that approximately six independent images produce the same uncertainty in AADT estimates as two 24-hour traffic counts. However, since estimating ODs from link volumes depends on the interactions between the network and OD structure, it is not straightforward to determine the value of a given quantity of imagery for OD estimation. Therefore, a network-based simulation study was devised to determine this value.

A simplified version of the Ohio intercity highway network was developed to demonstrate the methodology and quantify the benefits of adding imagery information. The network contains 29 centroids, with 812 ($=29\times28$) OD pairs, and 386 bi-directional links. Crude, but realistic, OD flows were approximated based primarily on population and network distances. These OD flows are considered the true ODs, μ_T, which will be unknown. The true underlying link volumes are given by $H\mu_T$, where the incidence matrix H is determined from shortest distance paths. AADT estimates $\{V = v\}$ on the links assumed to be sampled with MATRs or images are simulated from $MVN(H\mu_T, \Sigma)$, where Σ is a diagonal matrix based on (3) with $cv_g = 0.15$, $cv_s = 0.25$.

In this study, a prior OD mean μ_o was obtained from μ_T, with $\mu_o = (1 \pm 0.5)\,\mu_T$, where the sign was selected at random, and the prior variance C_0 is a diagonal matrix, whose i^{th} element is equal to $(\mu_{0_i} \times 0.5)^2$, $i = 1, \ldots, 812$. The posterior OD μ_1 in equation (2) is the Bayes estimate

of the true OD, which is compared to the true OD vector μ_T to evaluate estimation performance. In each of the 100 simulation runs, a different set of "observed" AADTs is simulated from $MVN(H\mu_T, \Sigma)$. Let μ_{1ik}, i=1,2,...,812, denote the posterior estimates (2) of the i^{th} OD in simulation run k, k=1,2,...,100. The performance measure associated with these estimates is the SNR (signal-to-noise ratio)

$$SNR = \frac{1}{100}\sum_{k=1}^{100} \frac{\sqrt{\frac{1}{812}\sum_{i=1}^{812}(\mu_{1ik} - \mu_{Ti})^2}}{\frac{1}{812}\sum_{i=1}^{812}\mu_{1ik}} \tag{4}$$

In our first simulation, approximately 3% of the links were randomly designated as being equipped with PATRs, and approximately 25% of the links were randomly designated being sampled with MATRs. (These % are roughly equivalent to the current practice.) Thus AADT estimates from PATR or MATR data were assumed to be available on 28% of the links as a baseline. Figure 1 plots the SNR as a function of the percentage of randomly chosen links in addition to the baseline (indicated as 0%), whose AADTs were assumed to be estimated from imagery. (The curves extend to values of 72% = 100%-28% additional links sampled, indicating that AADTs on all links were estimated in some way.) These additional links were assumed to sampled with m = *1, ..., 6* independent images, and the variances of the corresponding AADT values were determined from (3).

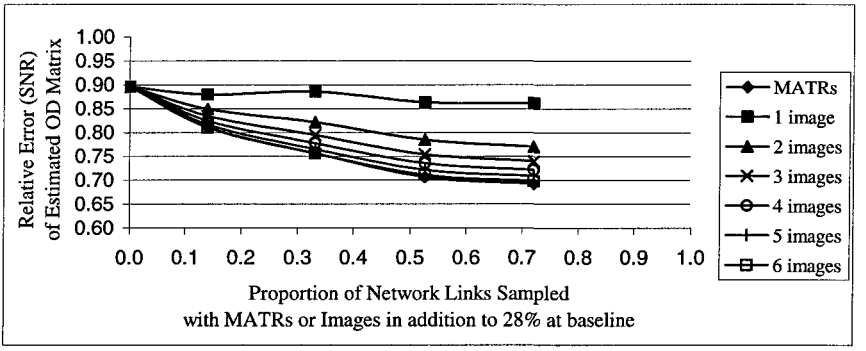


Figure 1. Relative Error in OD estimates vs. Proportion of additional links sampled: Baseline equal to 3% PATR and 25% MATR

Figure 1 indicates that adding just one image on some links or even on every link beyond the baseline, the performance of OD estimates is almost same compared to the baseline. This can be partly explained from the fact that the high level of uncertainty in AADT estimates based on one image can lead to a few very noisy counts, which can affect many OD estimates. However, having two images available on even a small proportion of additional links can markedly reduce the OD estimation error. Note that the curve corresponding to MATRs on additional links, at the bottom, is almost same as curve corresponding to six images on the same links. Thus for OD estimation, on the average, 6 images per link provide information equivalent to two daily counts on the same link.

To further explore the value of adding imagery in OD estimation, the second simulation used the same 3% of the segments assumed to be equipped with PATRs but only 12.5% assumed

to be sampled with two-day MATR counts (i.e., with the baseline of 3% + 12.5% = 15.5%). Additional links were then assumed to be sampled with m images ($m=1,...,6$), as in the first simulation. The results are shown in Figure 2. (In this case, the curves are determined up to an additional 84.5% when AADTs are estimated on all segments.) The results in Figure 2 once again show that having only one image does not have much of an impact on OD estimation, but that having two images appears to have a marked effect.

As expected the SNR error at baseline is greater in Figure 2 than in Figure 1 (approximately 0.95 vs. 0.90), since there are 12.5% fewer links assumed to be sampled with MATRs. Considering 3% PATR coverage and 25% MATR coverage to be representative of present practice, the 0.90 SNR in Figure 1 approximates "without imagery" OD estimation performance. To achieve this same SNR value in Figure 2, between approximately 8% (for 6 images) and 15% (for 2 images) additional links must be sampled with images.

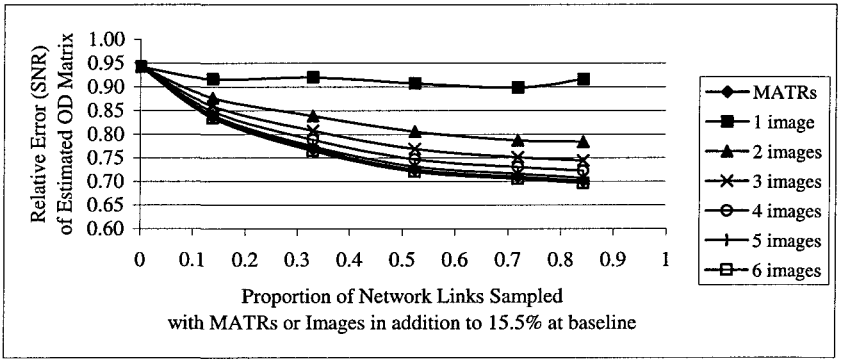


Figure 2. Relative Error in OD estimates vs. Proportion of additional links sampled: Baseline equal to 3%PATR and 12.5% MATR

Thus, if approximately 15% of the links could be imaged twice per year, the number of costly and dangerous ground-based MATR samples could be cut in half, and the same quality in the OD estimates would be obtained. Based on conversations with the Ohio DOT and analysis beyond the scope of this paper, we estimated that roughly 10-15% link-images are produced every year for various purposes. Although this value is not sufficient to produce 2 images on 15% of segments, it indicates that *existing* imagery could conceivably be used to improve the quality of OD estimates or to maintain the same quality of estimates while reducing the number of MATR samples taken per year by some meaningful percentage (to be determined from further studies), if the imagery contains multiple samples of the same segment.

As mentioned above, the interaction between the network structure and the OD matrix (as modeled through the incidence matrix H) would lead to samples v on different links having different impact on the estimated OD matrix μ_l (see equation (2)). To acknowledge this impact, the simulations were repeated for the two baseline scenarios but with a different set of links assumed to be equipped with PATRs, or sampled with MATRs, and the additional links sampled with images generated in different order. The numerical results (details not presented here) again indicate that just one image available on even a large percentage of additional links has little impact on the quality of the estimated OD flows, but having two images available on a relatively small number of links can markedly improve the estimation quality. The SNR values at baselines were, interestingly, very close to the 0.90 and 0.95 values of Figures 1 and 2. However, when reducing the percentage of MATR links from 25% to 12.5%, the percentage of additional links that must be sampled with images to obtain

equivalent SNR increased to between approximately 18% (for 6 images each) and 24% (for 2 images each). Although these percentages are roughly twice those for the conditions of Figures 1 and 2, it is interesting to note that the range in the percentage of additional segments that must be imaged is fairly small. Thus the marginal value of having a second image of the same link is large, but the marginal value of additional images on the same link decreases rapidly.

Conclusion

A Bayesian approach is adopted to estimate average statewide OD flows from AADT estimates while incorporating the different uncertainties in the AADT estimates that result when the AADTs are estimated from remote sensed images, or ground counts. Simulation studies beyond the scope of this paper show that ignoring different uncertainties in the AADT estimates greatly decreases the accuracy of the OD estimates.

Network-based simulation studies show having only a single image of even a large percentage of segments may not be beneficial in OD estimation, whereas having two images of additional segments could be valuable in decreasing the error in estimates of OD flows obtained from link volumes and in reducing the number of costly and dangerous ground counts that would presently produce these volumes.

Sensitivity of these results to the assignment matrix H, the prior OD mean and variance, as well as relative levels of uncertainty attached with AADT estimates is left for future study. In addition, since different links affect OF flows differently, the location of the sampled segments would appear to be an important factor in the results. Larger numerical studies are required to investigate this effect, and such studies are left for future work. Selecting the links according to some optimality criterion is also an interesting question to pursue.

Acknowledgment: This research was partially supported through the U.S. DOT National Consortium on Remote Sensing in Transportation-Flows at The Ohio State University. We thank Dr. Zhuojun Jiang for analytical and numerical help in producing inputs for simulation.

References

FHWA (1992) *Traffic Monitoring Guide,* U.S. DOT, Publication No. FHWA-PL-92-017.

Jiang, Z. (2005) *Incorporating image-based data in AADT estimation: methodology and numerical investigation of increased accuracy.* Ph.D. thesis, The Ohio State University,

Maher M.J. (1983) Inferences on Trip Matrices from Observations on Link Volumes: a Bayesian Statistics Approach. *Transportation Research*-B, Vol 17B. No 6, pp. 435-447, 1983.

McCord, M.R., Goel, P.K., Jiang, Z., Bobbit, P. (2002) Improving AADT and VMT Estimates with High Resolution Satellite imagery: simulated Analytical results. In *Applications of Advanced Technology in Transportation,* Proceedings of the 7th international conference, Eds. C.P.Wang, S.Madanat, S.Nambisan, and G. Spring, Reston, VA: ASCE, pp. 632-639.

McCord, M.R., Yang, Y., Jiang, Z., Coifman, B. and Goel, P.K. (2003) Estimating AADT from Satellite Imagery and Air photos: Empirical results. *Transportation Research Record No. 1855,* Transportation Research Board, Washington, D.C., pp. 136-142.

McShane, W.R., Roess, R.P., and Prassas, E.S. (1998) *Traffic Engineering,* 2nd ed, Prentice-Hall, Englewood Cliffs, N.J.

An Improved Compound Clustering Algorithm in Vehicular Ad-Hoc Networks

Peng Fan, James G. Haran, Peter C. Nelson, John Dillenburg
University of Illinois at Chicago, Department of Computer Science,
851 S. Morgan (M/C 152), Room 1120 SEO, Chicago, IL 60607-7053 USA
{pfan, jharan, nelson, dillenbu}@cs.uic.edu

Abstract

The application of Mobile Ad Hoc Network (MANET) technologies in the service of Intelligent Transportation Systems (ITS) has brought new challenges in maintaining communication clusters of network members for long time durations. Stable clustering methods reduce the overhead of communication relay in MANETs and provide for a more efficient hierarchical network topology. During the creation of Vehicle Ad Hoc Network (VANET) clusters, one vehicle node per cluster is selected as the cluster head to act as the routing node. In this paper, we propose a Compound Utility Function (CUF) clustering algorithm which takes into consideration the degree, position, velocity and acceleration of a vehicle altogether, and the invocation of this algorithm is not periodic as in earlier research, but reactive on the dynamism of the nodes. In the paper we experimentally compare CUF with Highest-Degree and Lowest-ID algorithm. The results show that CUF gives the best cluster stability.

1. Introduction

With the burgeoning need for the service of Intelligent Transportation Systems (ITS), communication between vehicles is considered a prime area where Vehicular ad-hoc-networks (VANETs) are likely to be deployed in the near future. The U.S. Federal Communications Commission (FCC) has recently allocated the 5.85-5.925 GHz portion of the spectrum to inter-vehicle communication (IVC) and vehicle-to-roadside communication (VRC) under the umbrella of dedicated short-range communications (DSRC [DSRC]). This has fuelled significant interest in applications of DSRC to driver-vehicle safety applications, infotainment, and mobile Internet services for passengers.

Vehicles in a VANET environment move within the constraints of traffic flow while communicating with each other via wireless links. Ad hoc networks use less specialized hardware for infrastructure support and leave the burden of network stability on the individual nodes within the network. Without routers, or other dedicated communication hardware, a possible method to optimize communication within the network is to develop a hierarchical clustering system within the network. To support the dynamic nature of the VANET environment, the clustering must be periodically updated to reflect topological changes and vehicle movements. Clustering within the network must be very fast to minimize time lost to clustering [Johansson, 2004].

A significant amount of research [Sivavakeesar, 2002][Basagni, 1999][Basagni, 1997] focuses on optimal methods for clustering nodes in MANETs. VANETs, however, pose new challenges in cluster head selection and network stability. VANETs must follow a tighter set of constraints than MANETs, and therefore require specialized clustering algorithms. First, nodes or vehicles must follow constraints set in place by the real road network topology. Second, vehicle movements follow well-understood traffic movement patterns. Third, vehicles generally travel in a single direction and are constrained to travel within a two-dimensional movement. Given these movement restrictions it is possible to approach clustering more intelligently and possibly discover a better clustering methodology for VANET environments.

Many simulation tools such as ns-2 [NS-2] have been designed for MANET implementations. These tools, however, fail to adequately model the needs of a VANET network. Therefore, for the purpose of this study, a micro-simulation tool [TS-3.0] is specially modified to perform randomized vehicle-based clustering under traffic constraints.

In this paper, according to previous utility concept [Fan, 2005] we propose a compound utility function algorithm. Most of the existing clustering algorithms are invoked periodically, but the invocation of our algorithm is delayed as long as possible to reduce the computation cost. Frequent updates result in high

communication overhead among vehicles. We show by simulation experiments that our method yields better results as compared to two classic algorithms.

The rest of the paper is organized as follows: in Section 2, a summary of background knowledge and related work are presented. We propose our new algorithm in Section 3. In Section 4 we discuss the simulation results. Section 5 concludes the paper and discusses some directions of future research.

2. Backgrounds and Related Work

2.1. Clustering in MANETs

Clustering algorithms are applied in communication networks to organize all nodes into groups and to obtain a hierarchical network organization. Figure 1 provides and example of the organization of twelve nodes into three clusters. The basic communication capability between the twelve nodes is outlines as connections between the lower tier of the hierarchy. In the upper tier, the three cluster head nodes are displayed with connections between them representing the possible message paths under the cluster-constrained network [Bettstetter, 2002].

Under this architecture each cluster head aggregates local member topology and acts as a relay point for communication between its members and members of other clusters. This reduces the messages exchanged between individual network nodes and the overhead of information stored within those nodes [Garg, 2004].

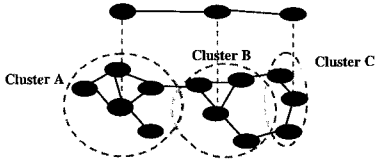


Fig.1. Clustering within a 12-node MANET environment

Cluster formation in MANETs has attracted considerable attention in recent Times [Sivavakeesar, 2002][Basagni, 1999][Basagni, 1997]. Several algorithms have been proposed, such as Lowest-ID algorithm [Gerla, 1995][Ephremides, 1987][Jiang, 1999] and Highest-Degree [Gerla, 1995] algorithm. In Lowest-ID, if a node hears from a cluster head with a lower ID than itself, it resigns and uses that node as a cluster head instead. In Highest-Degree, it uses the degree of the nodes instead of ID to elect cluster heads. The performance of these two algorithms was shown in [Basagni, 1999] [Gerla, 1995]. Although many other clustering algorithms existed, they are just designed for different proposes. This research focuses on these two algorithms due to their constant time complexity and good scalability.

2.2 Transportation-specific Utility Clustering Methodology

In our previous research [Fan, 2005] [Haran, 2005], it has focused on the discovery of traffic-optimized clustering schemes. Each vehicle implements some form of utility analysis of each proximally located possible cluster head. Periodically, each vehicle will broadcast general network information such as ID and current degree as well as vehicle-specific traffic statistics such as position, velocity, and acceleration. Upon receipt of this information, each vehicle chooses a cluster head by evaluating the utility of each potential head. The node with the highest utility is selected as the cluster head. The belief is that these traffic-specific utility functions will be better predictors of the common traffic situations that lead to cluster dissociation. Here we give an example of utility function:

Closest Position to Average (CPA): A vehicle attempts to choose as its cluster head in order of the absolute difference of candidate's position to the average position of all proximal vehicles.

These detailed steps outline the procedure for implementation of the utility function:

1. Each vehicle determines the vehicles within range by polling the local broadcast region and tracking the candidate cluster head set C. All vehicles with broadcast range are considered candidate cluster heads.

2. Using candidate set C and the state information received by broadcast; each candidate is evaluated using CPA function.

3. The cluster head is chosen in decreasing order of CPA value. The petition for cluster membership is broadcast to the chosen vehicle. The chosen vehicle will deny the request the vehicle with the next highest utility is selected and this step repeated.

The simulation result highlighted the performance of the Lowest-ID clustering algorithms as optimal for the constrained MANET environment provides by VANETs. On the other hand, four different utility functions presented in this research, *(i) the Closest Velocity to Average* and *(ii) the Closest Acceleration to Average* algorithms provided fairly stable clusters, however stability degraded as transmission range increased. *(iii) The Closest Velocity* and *(iv) the Closest Position to Average* algorithms showed somewhat stable performance but were prone to cluster head changes.

3. Compound Utility Algorithm

3.1. Design Philosophy

As discussed previously, each of the existing clustering algorithms is invoked periodically and none of them leads to an optimal selection of cluster head since each deals with only a subset of parameters which impose constraints on the system. Therefore, as the search for better solutions for this problem continues, we propose the use of a non-periodical compound utility function that takes into account several parameters like the node degree, mobility and stability. The goal of this research is to improve on the basic MANET clustering by applying domain-specific logic to aid in cluster head selection.

3.2. Basic Features of the algorithm

Motivated by the work [Chattterjee, 2003], we take into account vehicle degree, velocity, position to decide how well suited a vehicle is for being a cluster head. The following features are considered in our utility algorithm:

1. Compare to previous design, the cluster head election procedure is neither periodic nor entire and is invoked as rarely as possible. This reduces system updates and hence computation and communication costs.
2. Each cluster head can ideally support up to MAX (a predefined number) nodes to ensure efficient MAC functioning. Limiting or optimizing the number of vehicles in each cluster can achieve a high throughput of the system.
3. Stability and mobility are important factors to decide the cluster heads. When one of the ordinary vehicles leaves a cluster and joins another existing cluster, the amount of information exchange between the vehicle and the corresponding cluster head, is local and relatively small. The information update in the event of an inter-cluster change is much more than an intra-cluster.

3.3 Compound Utility Function

Based on the preceding discussions, we propose a new Compound Utility Function (CUF) that effectively combines several important parameters with certain weighting factors chosen according to the system model needs. The flexibility of changing the weight factors helps us apply our utility function to various problems. The CUF is actually cluster head election procedure; every invocation of CUF does not necessarily mean that all the cluster heads are replaced with new ones. Since every vehicle is aware of its distance to adjacent clusters and to its current cluster head. When a vehicle is no longer able to attach itself to the cluster head, the head tries to hand over the vehicle to one of the adjacent cluster. On the other hand, even new vehicles are trying to join to an existing cluster head instead of triggering an invocation. The desire is to improve upon the initial clustering logic to obtain better stability over the simulation time. The detail description of CUF is below:

$$Utility = W_1 \cdot DD + W_2 \cdot SD + W_3 \cdot AVT + W_4 \cdot AAT$$

$$\sum W_i = 1$$

1. Find the neighbors of each vehicle. This gives the degree D of this vehicle. Compute the degree-difference $DD = |Max - D|$.
2. For every vehicle, compute the sum of distance SD with all its neighbors.
3. Compute the average velocity AVT for every vehicle within some time period T. This gives a measure of mobility.
4. Compute the average acceleration AAT for every vehicle within some time period T. This gives a measure of stability.
5. Calculate the utility function for each vehicle, and choose that vehicle with the smallest utility as the cluster head.

All these parameters are normalized, and the corresponding weights W_1, W_2, W_3, W_4 are kept fixed for a given scenario. It is to be noted that the elected cluster head and its neighbors are no longer eligible to participate further in the remaining part of the election process. The election process is continued until every node is found to be either a cluster head or a neighbor of some cluster head.

The first component DD, contributing to the combined utility, helps in efficient MAC functioning because it is always desirable for a cluster head to handle up to a certain number of vehicles in its cluster. The motivation of SD is mainly because a cluster centroid is the best position to manage a cluster, and it will reduce the transmission delay time as well as keep a cluster relatively stable. Therefore the closer a vehicle to its centriod, the better a cluster head it will

be. The last two components are due to mobility and stability of the vehicles. As discussed previously, a vehicle with relatively stable mobility to its neighbors is always a better choice for a cluster head.

It can be argued that the Lowest-ID and Highest-Degree algorithms are all special cases of the CUF algorithm. As for Lowest-ID algorithm, the assignments of the ids are random. We can assume that the ids being assigned are based on mobility measurement AVT. The lowest id is assigned to the least mobile vehicle and highest id for the most mobile, in that case, $W_1 = W_2 = W_4 = 0$. The Highest-Degree considers only the degree of a vehicle and disregards all other parameters ($W_2 = W_3 = W_4 = 0$).

3.4 Connectivity with Traffic Flow

Clustering ensures that the nodes within a cluster are able to communicate among themselves through the cluster heads, each of which acts as the central node of a star. However, the volume of traffic flow may put a huge influence to inter-cluster communication due to density of nodes.

We define connectivity as the probability that a vehicle is reachable from any other vehicle. For a single component graph, any node is reachable from any other node and the connectivity is 1. If the network does not result in a single component graph, then we can say that all the nodes in the largest component N can communicate with each other and the connectivity can be the ratio of N to the cardinality of the graph. Thus,

$$Connectivity = {}^{N}/_{TotalNodes}$$

We will investigate the relationship between connectivity and traffic flow in later simulation section. The transmission range of a cluster head can be made large enough by adjusting the power in such a way so as to yield a connected network.

4. Simulation Study

This study modified Traffic Simulation 3.0 [TS-3.0], an Intelligent-Driver Model (IDM) micro-simulation tool built to monitor traffic flow under various basic highway configurations. This environment simulates accelerations and braking decelerations of drivers (i.e. longitudinal dynamics), and uses the Minimized Overall Braking Induced by Lane changes (MOBIL)[MOBIL] lane change model. All model parameters and the initial simulation source code are available at [TS-3.0].

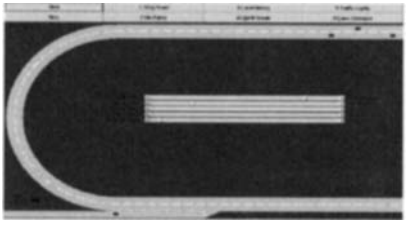


Fig. 2. On Ramp Screenshot of Traffic Simulation

4.1 Implementation

The source code for the aforementioned simulation tool was modified to perform fixed interval cluster formation using either of three experimental algorithms (Lowest-ID, Highest-Degree, CFU). To aid in algorithm visualization, the graphical display of the micro-simulation environment was modified to display vehicle clusters using contrasting colors.

4.2 Metrics

In addition to utility function and display changes, periodic state logging was implemented. This data provided the basis for the simulation result analysis and algorithm comparison. To measure the system performance, three metrics were identified: (i) the average cluster head change per step and (ii) the average cluster size. Metric (*ii*) alone does not accurately depict system performance, so the relative measurement (*ii*)/(*i*) was introduced to provide a reasonable comparison metric between the analyzed algorithms. A method is considered relatively better if it has either better stability using metric (*i*) or larger average cluster size. (*iii*) The cluster connectivity as mentioned before.

4.3 Limitations

The simulation tool only provides couple simulation scenarios, such as Ring Road, On-Ramp, On-Ramp with Traffic Lights, etc. Therefore, these simplified scenarios might not satisfy all topological requirements of traffic-related problems. On the other hand, it does not include a complete set of configurations for a simulated traffic, which could affect the accuracy of the simulated results by using the default value under certain circumstances.

5. Results Discussion

The simulation results represent the performance of each algorithm across various wireless transmission range values (0-300 meters) and maximum vehicle speed (40-140 kilometers/hour) with a fixed maximum cluster degree M = 50 vehicles. The values use for simulation were W1 = 0.6, W2 = 0.2, W3 = 0.1 and W4 = 0.1. Note that these values are arbitrary at this time and should be adjusted according to the model requirement. In addition, all simulations are tested under On-Ramp scenario and the simulation time duration was held constant across all tests. To minimize traffic flow variability between simulations and enable repeatable test results, the randomized features of the model were seeded with the same value at each simulation run.

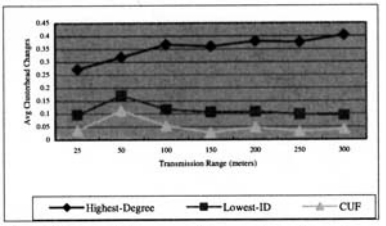


Fig. 3. Cluster Changes vs. Transmission Range

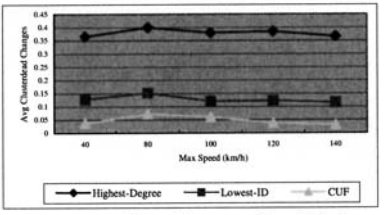


Fig. 4. Average Cluster Change vs. Speed Limit

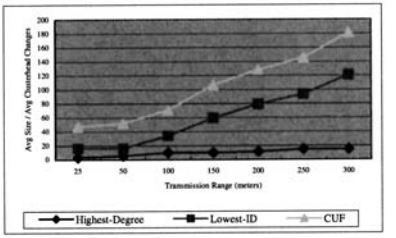


Fig. 5. Clustering Ratio vs. Transmission Range

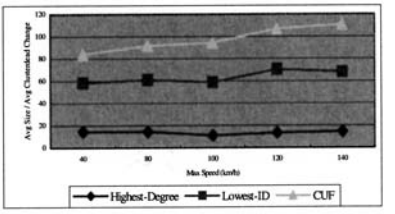


Fig. 6. Clustering Ratio vs. Speed Limit

Figure 3 summarizes the variation of the average number of clusters in unit time with respect to the transmission range. It illustrates the performance of the two algorithms for a reasonably standard traffic flow environment with a fixed maximum speed of 100km/h. In Figure 3, the maximum clusters change value is 1.0 indicating that a vehicle will change its head every unit time. Notably, the CUF clustering algorithm enabled most stable clusters.

Figure 4 shows the effect of varying the maximum speed on the average number of cluster head changes with a fixed transmission range of 150m. Algorithm performance is consistent with those of Figure 3. Speed limits are useful only in heavy-traffic situations [Treiber, 2000]. Figure 5 displays the performance of all algorithms over various transmission ranges. Higher curves indicate better performance. Figure 6 shows the overall performance across various speed limits.

To study how the volume of traffic flow affects the reachability of one vehicle from another, it is essential that the clusters be connected. Simulation was conducted for CUF algorithm with 5 different volume of traffic flow, e.g, 500 vehicle/hour. Figure 7 demonstrates that a well-connected graph can be obtained at the larger volume of traffic flow and longer transmission range.

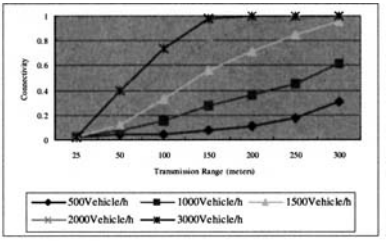


Fig. 7. Connectivity vs. Traffic Flow

6. Conclusion and Future Works

From the previous discussion, the analysis highlights the performance improvement of the CUF algorithms to the well-known Lowest-ID and Highest-Degree algorithms for the constrained MANET environment provides by VANETs. The non-periodical invocation of CUF reduces lots of communication overheads by utilizing a previous clustering topology. More importantly, this paper presents a simulation model of testing clustering algorithms for Vehicular Ad-hoc Networks, providing basic clustering functionalities and some traffic-specific statistics. It also takes into account the simulation scenarios influencing the clustering algorithm results with allowing to adjusting various traffic parameters.

The results show an initial approach and testbed model to analyzing parameterized VANET dynamics from a traffic micro-simulation perspective. Further work will be focused on exploring model functionality and network performance testing.

References

[Basagni, 1997] Basagni, S., Chlamtac, I., Farago, A., A Generalized Clustering Algorithm for Peer-to-Peer Networks, Workshop of ICALP'97, Bologna, Italy, Jul 11-12, 1997.

[Basagni, 1999] Basagni, S., Distributed Clustering for Ad Hoc Networks, Proceedings of I-SPAN'99, IEEE Computer Society, pp.310-315, Australia, Jun 23-25,1999.

[Bettstetter, 2002] Bettstetter, C., and Konig, S. On the Message and Time Complexity of a Distributed Mobility Adaptive Clustering Algorithm in Wireless Ad Hoc Networks, proceeding of European Wireless 2002, Florence, Italy.

[Chatterjee, 2000] Chatterjee, M., Das, S., Turgut, D., "An on-demand Weighted Clustering algorithm for ad-hoc networks", Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM), 2000, pp.1697-1701

[DSRC] Dedicated Short Range Communications (DSRC) home. http://www.leearmstrong.com/dsrc/dsrchomeset.htm.

[Ephremides, 1987] Ephremides, A Design Concept for Reliable Mobile Radio Networks with Frequency Hopping Signaling, Proc of the IEEE, Vol. 75, No. 1, January 1987, pp. 56_73.

[Fan, 2005] Fan, P., Haran, J., Nelson, P., and Dillenburg J., Cluster-Based Framework in Vehicular Ad-Hoc Networks. ADHOC-NOW 2005: 32-42.

[Garg, 2004] Garg M. and Shyamasundar, R.K. A Distributed, Clustering Framework in Mobile Ad Hoc Networks, Proceedings of ICWN'04, Las Vegas, USA, June21 24,2004.

[Gerla, 1995] Gerla, M. and Tsai, J., Multicluster, Mobile, Multimedia Radio Network, Wireless Networks, 1(3) 1995, pp. 255-265.

[Haran, 2005] Haran, J., Fan, P., Nelson, P., and Dillenburg J., An intelligent vehicle approach to mobile vehicular ad hoc networks - clustering optimisation in dynamic traffic networks. ICINCO 2005: 224-231.

[Jiang, 1999] Jiang, M., Li, J., and Tay, Y.C. Cluster Based Routing Protocol, IETF Draft, August 1999.

[Johansson, 2004] Johansson, T. and Carr-Motyckova, L. Bandwidth constrained Clustering in Ad Hoc Networks, In The Third Annual Mediterranean Ad Hoc Networking Workshop. Bodrum, Turkey, June 27 - 30 2004.

[MOBIL] MOBIL http://vwisb7.vkw.tudresden.de /~treiber/MicroApplet/MOBIL.html

[NS-2] Project, V. The network simulator –ns-2

[Sivavakeesar, 2002] Sivavakeesar, S., A Prediction-Based Clustering Algorithm to Achieve Quality of Service in Multihop Ad Hoc Networks, London Communication Symposium, 2002.

[Treiber, 2000] Treiber, M., Hennecke, A., and Helbing, D., Congested Traffic States in Empirical Observations and Microscopic Simulations, Physical Review E 62, 1805 2000.

[TS-3.0] Traffic Simulation 3.0. http://vwisb7.vkw.tu-dresden.de/~treiber

Developing Microscopic Integrated Freeway-Toll Plaza Model

Haitham M. Al-Deek[1] and Nezamuddin[2]

[1]Ph.D., P.E., Professor of Engineering, Department of Civil and Environmental Engineering, University of Central Florida, P.O. Box 162450, Orlando, FL 32816-2450; PH (407) 823-2988; FAX (407) 823-3315; email: haldeek@mail.ucf.edu

[2]Graduate Research Assistant, Department of Civil and Environmental Engineering, University of Central Florida, P.O. Box 162450, Orlando, FL 32816-2450; PH (407) 823-4552; FAX (407) 823-4775; email: ne532690@ucf.edu

ABSTRACT

Since a built-in toll plaza model is not available for use with any of the widely used microsimulation models, almost every researcher independently develops a model using some programming language (VB, Java etc.) and customizes it for his or her study. Most of these models are queuing theory based input-output models, which lack the element of traffic flow theory. This paper presents a simple approach to develop a microscopic toll plaza model by customizing standard tools available in PARAMICS. A model for Holland East mainline plaza in Orlando, Florida has been developed and successfully calibrated for volume and delay values for morning peak-hour. The procedure is generic in nature, and can be used to develop a model for any other toll plaza with appropriate modifications.

I. INTRODUCTION

Simulation modeling is the cheapest and fastest way to study different types of facilities on ground without interfering with their operation. It can be used to study future conditions and hypothetical scenarios. A microscopic toll plaza model can be used to estimate delay at toll plazas and quantify the benefits of open road tolling.

Many researchers have developed toll plaza models for their studies (*Al-Deek 2000, Correa 2004, Lin 2002, Zarrillo 1997*). Usually, these models are not available to other researchers due to practically no distribution in the market, and are often not supported few years after being developed. Some of these models work independent of the actual geometry on ground, and vehicular movements in the transition area (*see Figure 1*) are not simulated microscopically. A vehicle is directly assigned to one of the toll lanes from the uniform toll road segment based upon payment type (manual, coin machine or electronic toll) and queue lengths at toll booths. If a vehicle is traveling in leftmost lane and the rightmost toll lane has the shortest queue length, then the queuing model will assign this vehicle to the rightmost lane, and the vehicle will do unrealistic maneuvering to reach to the assigned toll lane instantly. Vehicles do not take into account headway, gap acceptance, or inter-vehicle interaction to follow a lead car or to perform lane changing maneuvers. For queuing models, ensuring equal queue lengths at all the toll booths and processing the queued vehicles at defined rate are the primary mechanisms which account for the toll plaza simulation. They do not simulate vehicular movements just upstream of the toll plaza microscopically.

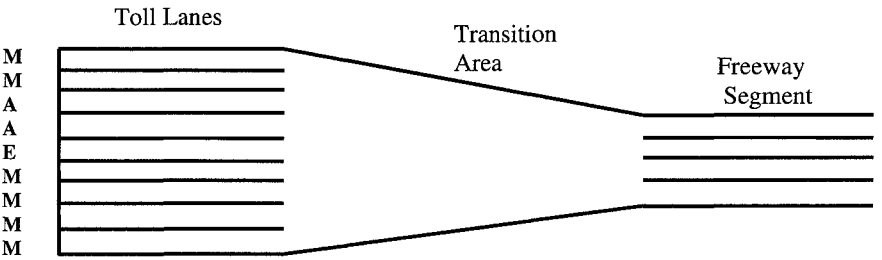


Figure 1: Holland East Toll Plaza – Queuing Model (*Al-Deek et. al 2000, and Al-Deek 2001*)
(M – Manual payment, A – Automatic coin machine, E – Electronic toll collection)
Electronic toll collection (ETC) vehicles can use any lane.

PARAMICS employs car-following and lane-changing models to simulate the movement of individual driver vehicle units (DVU). A DVU is a combined representation of the behavior of driver and the physical characteristics of a vehicle. Each DVU contains a set of parameters (mean target headway, mean reaction time, aggressiveness and awareness, physical dimensions of vehicle, acceleration & deceleration profile ... etc.) which define its behavior completely. Using the car-following model, a DVU changes its speed according to its perception of the speed and acceleration of the DVU in front. It also adjusts its speed after seeing brake lights of the DVU immediately ahead. Lane changing in PARAMICS is done when a suitable gap in target lane is available. Acceptable gap is based on the target headway, which is again linked with the car-following model. Target headway for a DVU varies around the specified mean target headway and depends upon other parameters. For example, a high aggression value will cause a DVU to accept a smaller headway and hence smaller gap. The very nature of the car-following and lane-changing models in PARAMICS and their application in simulating vehicular movements around toll plaza ensures that this model works in accordance with traffic flow theory (*PARAMICS User Guides V 5.0*).

Tools of PARAMICS used to develop the toll plaza model are discussed in Section II. Section III illustrates the calibration procedure and Section IV draws the conclusions.

II. TOLL PLAZA MODEL

An aerial picture of Holland East plaza is used as an overlay to ensure that the toll plaza model operates under similar geometric conditions as the real plaza (*Figure 2*).

Figure 2: Holland East Toll Plaza - PARAMICS Model

After the geometry is complete, the primary tasks are to:

a) Instill a tendency in the vehicles to smoothly move towards one of the appropriate payment lanes as they approach the toll plaza.
b) Process only valid vehicle types through toll lanes. For example, automatic coin machine (ACM) lane will process only ACM vehicles and ETC vehicles, but it will bar manual payment vehicles.
c) Assign appropriate stop time (time duration for which vehicles remains stationary at booth to pay toll) for manual payment and ACM vehicles to pay toll.

Important tools of PARAMICS required to complete these tasks are discussed below.

i. Vehicle Type Manager

Vehicle Type Manager (*Figure 3, left*) can be used to define a vehicle type as per the requirement of the modeling. From the operational point of view of the toll plaza, three types of vehicles have been created depending upon their mode of toll payment.

MB - This vehicle type pays toll by using manned booth lanes.
ACM - This vehicle type pays toll by using automatic coin machine lanes.
E-PASS - This vehicle type pays toll by using any lane equipped with ETC.

These vehicle types are assigned different colors so that they can be easily identified during simulation runs: MB- Green, ACM- Blue and E-PASS- White.

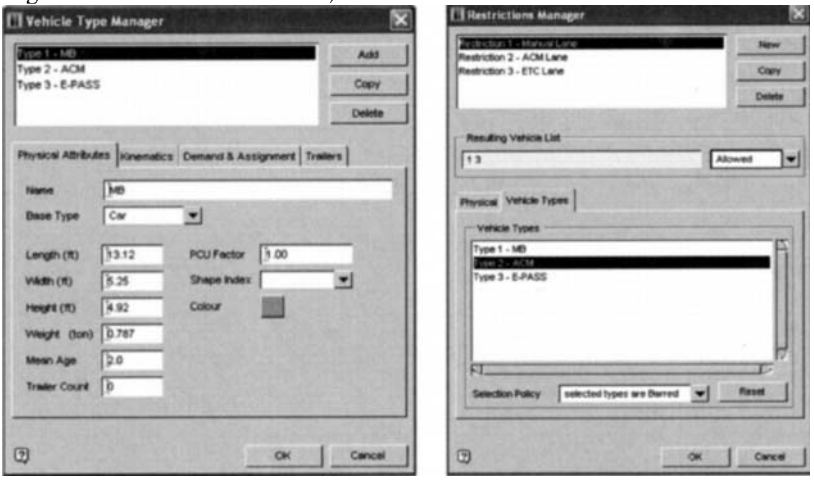


Figure 3: Vehicle Types (left) and Lane Restrictions (right)

ii. Restrictions Manager

Restrictions Manager (*Figure 3, right*) can be used to define restrictions on different types of vehicles. Once a restriction is defined, it can be applied to a particular lane of any road segment in the network. Vehicles falling into that restriction category will not be allowed to enter the restricted lane. Three types of restrictions have been defined to represent different toll payment lanes at the toll plaza.

Manual lane restriction allows MB and E-PASS vehicles to pass through the toll lane and bars ACM vehicles. Similarly, ACM and ETC lane restriction are defined.

iii. Next-lane Allocation

As the free flow speed section of the road approaches the toll plaza, smooth transition is provided to accommodate more and more lanes. During the simulation, vehicles have a tendency to occupy same old lanes, and the newly added lane remains empty. Suppose that a link in the model has five lanes and the downstream link has six lanes. Vehicles will occupy only five rightmost lanes in the downstream link and the leftmost lane remains empty. Next-lane Allocation can be used to map upstream lanes onto downstream lanes. It tells the vehicles in a particular upstream lane to choose from one or more of the downstream lanes as per the settings. Next-lane has been used in such a manner that all the downstream lanes are properly utilized (*Figure 4*).

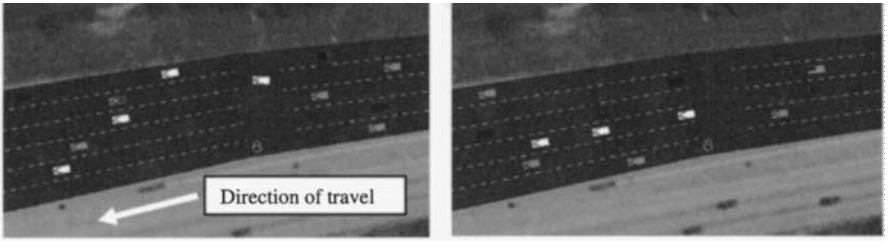


Figure 4: Lane Usage – Before (left) and After (right) Next-lane Allocation

iv. Lane choices Rule

This is the primary tool used to put intelligence in the vehicles, so that they move towards the correct toll lane(s) while approaching the plaza. Lane choices rule does not tie a vehicle to a particular toll lane beforehand; it creates a tendency in the vehicle to move towards defined lane groups and then depending upon the vehicle type (MB, ACM or E-PASS), lane restrictions (Manual, ACM or ETC lane) and queue lengths, vehicle chooses one of the toll lanes based on the logic embedded in the microscopic nature of PARAMICS. *Figure 5 (left)* illustrates an example of lane choices rule for MB vehicles (Type: 1) while traveling on link 7:8 (i.e., road segment

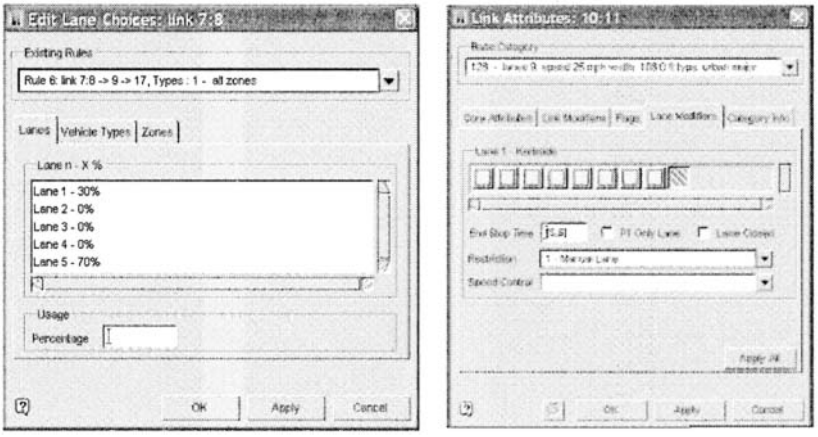


Figure 5: Lane choices rule for MB (left) and Manual Lane at Toll Booth (right)

connecting nodes 7 and 8, *see Figure 2*) and going towards nodes 9 and 17 near the toll plaza.

Link 7:8 has five lanes and in accordance with lane choices rule, MB vehicles split themselves into these lanes. Since there are 2 manual payment lanes in the right and 4 in the left at the toll plaza, MB vehicles are split into two lanes (Lane 1-30% & Lane 5-70%). It is to be noted that PARAMICS does not enforce this rule strictly; instead it creates a general tendency in MB vehicles to occupy rightmost (Lane 1) and leftmost lane (Lane 5) while traveling on an upstream link 7:8. Therefore some MB vehicles also travel on lanes other than these two lanes, which is a more realistic case. *Figure 5 (right)* shows an example of manual lane restriction and stop time for a toll lane. After the application of lane choices rules (*Figure 6*), ETC vehicles occupy center

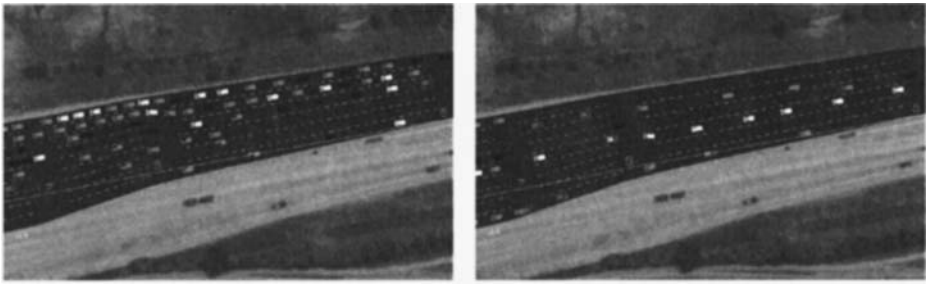


Figure 6: Lane Usage - Before (left) and After (right) Lane Choices Rules

lanes, automatic coin machine vehicles occupy lanes right of center, and manual payment vehicles use left and right lanes. This is appropriate lane usage for the given toll lane configuration (toll lanes from right to left: M-M-A-A-E-M-M-M-M).

III. CALIBRATION

PARAMICS uses large number of parameters to simulate traffic movements. It is practically impossible to adjust all the parameters. Over the years, researchers have identified key parameters that need to be adjusted, such as: mean target headway, mean reaction time, and aggressiveness and awareness of drivers (*Lee 2001*).

During the course of calibration, it was found that the above list of simulation parameters is not adequate for simulating movement near toll plazas. There are many other parameters which directly affect the simulation of toll plaza (*Table 1*). Apart from these parametric calibrations, a significant portion of calibration effort is qualitative in nature. Tools discussed in Section II, particularly next-lane allocation and lane choices rules, play very important role in calibration process. They are used to streamline the movement of different vehicle types near toll plaza, and minimize the interference between them while weaving is taking place. Although numerical adjustment of parameters increases the throughput, it stops at a maximum value which is far below the observed value. After that, the inter-play of qualitative parameters is used to improve the operation at the toll plaza.

There are two kinds of data available for Holland East plaza: Morning peak-hour volume data for individual lanes and for the entire toll plaza (*Klodzinski 2001*) and

peak-hour delay (*Zarrillo 1998*). These data were extracted from video records of Holland East plaza for morning peak hours (7AM-8AM).

Table 1: Calibrated Parameters

Parameters	Default Values	Adjusted Values
Queue Gap Distance (ft)	32.81	26.00
Queuing Speed (mph)	4.47	15.00
Mean Target Headway (s)	1.0	1.5
Mean Reaction Time (s)	1.0	1.0
Minimum Gap (ft)	6.56	3.00

Different combinations of parameter values and lane usage percentages in lane choice rules have been tried to achieve the optimum results. In *Table 2*, volume data in default values column corresponds to the result obtained by only using the qualitative calibration.

Table 2: Calibration Results - Hourly Volume

Lane number	Hourly Volume (Observed)	Using default values		Using adjusted values	
		Hourly Volume	Error (%)	Hourly Volume	Error (%)
Lane 1 (M)	NA	383	NA	365	NA
Lane 2 (M)	380	403	6.05	401	5.53
Lane 3 (A)	NA	588	NA	576	NA
Lane 4 (A)	615	600	-2.44	609	-0.98
Lane 5 (E)	1320	1466	11.06	1387	5.08
Lane 6 (M)	427	442	3.51	425	-0.47
Lane 7 (M)	NA	361	NA	362	NA
Lane 8 (M)	NA	464	NA	449	NA
Lane 9 (M)	NA	466	NA	437	NA
Total	5018	5173	3.09	5011	-0.14

Observed field delay for the same day is recorded as 3654 veh-min during morning peak hour. It is important to understand how delay is calculated in field, so that the same procedure is applied when the model is being calibrated. While calculating delay, length of queue is noted down from the video data for each lane at an interval of one minute. It is assumed that if a vehicle is in queue at the end of the minute, then it is delayed by one minute. L_{ji} is defined as the queue length in lane i at the end of minute j. The total number of vehicles stationed at the plaza for each minute j, is determined by summing over all N lanes in the plaza. The summation over all minutes of rush hour, n, is the actual rush-hour-delay in units of vehicle-minutes, as shown below:

$$L_j = \sum_{i=1}^{N} L_{ji} \quad \text{and} \quad \text{Rush-hour-delay} = \sum_{j=1}^{n} L_j = \sum_{j=1}^{n} \sum_{i=1}^{N} L_{ji}$$

To follow the same procedure of delay calculation using simulation, number of vehicles queued upstream of the toll plaza is recorded at interval of one minute. Queue lengths calculated over an hour is summed up to calculate total delay, which is 3618 veh-min and has an error of 1 % compared to observed delay value. The calibrated values of stop times are as follows: 3 sec for automatic coin machine lanes, and a uniform distribution between 5 sec and 6 sec for manual payment lanes.

IV. CONCLUSIONS

In absence of a standard built-in microscopic toll plaza model, this paper outlines a methodology to develop a microscopic model using PARAMICS. The application of car-following and lane-changing models of PARAMICS to simulate vehicular movements near toll plaza ensures that this model functions in accordance with traffic flow theory. The model has been successfully calibrated for volume and delay values. Since transportation researchers use one of the network simulators to work on simulation projects, this model can be tested and verified by many researchers and can potentially act as the standardized toll plaza model. Similar attempt can be made for other microscopic simulation models. This model is flexible and can be customized for different configuration of payments lanes. It can also be used to evaluate different scenarios, like open road tolling, lane closure and incident management strategies.

REFERENCES

Al-Deek H.M., Mohamed A.A., Radwan E.A, "New Model for Evaluation of Traffic Operations at Electronic Toll Collection Plazas", Transportation Research Record 1710, pp. 1-10. TRB, National Research Council, Washington, D.C., 2000

Al-Deek H.M. "Analyzing Performance of ETC Plazas using New Computer Software." Journal of Computing in Civil Engineering, Vol. 15, No. 4, October 2001, pp. 309-319.

Correa E, C. Metzner and N. Nino, "Tollsim: Simulation and Evaluation of Toll Stations", International Transactions in Operational Research (2004)

Klodzinski, Jack , Methodology for Evaluating the Level of Service (LOS) of Toll Plaza on a Toll Road Facility, Ph.D. Dissertation, University of Central Florida, Orlando, FL, 2001.

Lee D., X. Yang and P. Chandrasekar, "Parametric Calibration of PARAMICS using Genetic Algorithm", Presented at the 80th Annual Meeting of the Transportation Research Board, Washington, D.C., January 2001.

Lin, Feng-Bor, "Delay Model for Planning Analysis of Main-Line Toll Plazas", Transportation Research Record 1776, TRB, National Research Council, Washington, D.C., 2002.

Paramics User Guides V 5.0, Quadstone Ltd., Edinburgh, UK, 2004.

Zarrillo M., A. Radwan and H. Al-Deek, "Modeling Traffic Operations at Electronic Toll Collection and Traffic Management Systems", Computers and Industrial Engineering: An International Journal. 33 3–4 (1997), pp. 857–860.

Zarrillo, M.L, Development and Applications of TPModel: A Queuing Model Describing Traffic Operations During Electronic Toll Collection (ETC), Ph.D. Dissertation, University of Central Florida, Orlando, FL, 1998.

Block Container Trains Formation Plan between Railway Network Container Freight Stations Based on Genetic Algorithm

Yan Haifeng[1], Peng Qiyuan[2], Tan Yunjiang[3]

[1]Doctor, Institute of Transport and Economy, China Academy of Railway Sciences, Num.2, Daliushu Street, Haidian Block, Beijing 100081, China; PH & FAX (0086) 010-51874065; email: yanhaifeng@rails.com.cn
[2]Prof., College of Traffic & Transportation, Southwest Jiaotong University, Num.111, First north Segment, Second Loop-road, Chengdu 610031, China; PH & FAX (0086) 028-87600750; email:qiyuan-peng@263.net
[3]Master, College of Traffic & Transportation, Southwest Jiaotong University, Num.111, First north Segment, Second Loop-road, Chengdu 610031, China; PH & FAX (0086) 028-87600757; email: start1001@tom.com

Abstract

Freight railroad operations involve complex classification and train formation decisions. Optimization with respect to these decisions can be quite difficult due to discrete and non-linear characteristics of the problem. The train formation plan is one of the important elements of railroad system operations. While mathematical programming formulations and algorithms are available for solving train formulation problem but CPU time required for their convergence is excessive. In this paper, 0-1 bi-level Linear Programming (BLP) is examined for obtaining good solutions with reduced complexity for block container trains formation plan (BCTFP). The minimum spending of container hour is taken as linear step function for optimization of the train formation plan. Cooperative multi-colony genetic algorithm (GA) is proposed for the solution of the block container train formation plan. Finally, the main contribution of the paper is to show the reduced complexity of the algorithm i.e. $O(\alpha n^4 \ln \beta n^2)$. The efficiency of the algorithm is also shown for convergence to global optimum.

Introduction

In the freight railroad industry, car movements are carried out by formation of cars into trains through successive classification. In order to achieve maximum efficiency the cars in the same destination are grouped together. This grouping is done according to block plan. Then block with the same direction travel is organized into trains according to train formation plan. After sufficient trains are accumulated then it is dispatched according to the train schedule.

In China the train formation plan (TFP) started to practice in 1950s, and most of advances in this field surged after 1978 (Cao, 1992). Many models are structured, including mixed integer programming (MIP), 0-1 linear programming (LP), 0-1 quadratic programming, synthetic optimization, network optimization

and beneficial directions model. Besides, some heuristics algorithms are put forward, such as branch-and-bound method, cutting plane algorithm, ordered combinatorial tree algorithm, implicit enumerative algorithm, advanced linear programming approximating algorithms, dynamic programming and simulated annealing algorithm.

Due to the discrete and non-linear programming nature of the problem a common element exists in rail operating decision models that they all require substantial investment of computational effort and subsequently implementation time. The nonlinear discrete programming problem is a kind of NP-complete problem (Harmor, Liu 1990). The optimization of BCTFP between railway network container freight stations (RNCFS) is a sub-class of train formation plan (TFP). Existing models and algorithms are not so applicable for the optimization of large-scale railway network due to their computational complexity.

To overcome these difficulties we introduced an Adaptive genetic algorithm (GA) to reduce the overall complexity of the algorithm for the solution of the block container train formation plan. The results are encouraging and the main contribution of the paper is to show the reduced complexity of the algorithm i.e. $O(\alpha n^4 \ln \beta n^2)$. The effectiveness of the algorithm is also shown for convergence to global optimum.

Model Formulation

Related Assumptions and Definitions

Assumption 1: Considering railway network container freight station (RNCFS) as a black box, there are mainly three parameters used: a_{ij} is container volume departing from station i and ending at station j, $(i \neq j)$ in 24 hours. T_{ij} is accumulation cost from station i to station j for block container train (BCT) in 24 hours and t_i is average saving time per container for any unclassified container flows.

Assumption 2: Each a_{ij} is corresponding to a definite traffic pathing ρ_{ij}, and all RNCFS is a subclass of $R=\{\rho_{ij}\}$, in which ρ_{ij} is big enough to accommodate container flow intensity.

Assumption 3: Any two adjacent RNCFS is bound to have BCT.

According to the assumptions, railway network can be described as directed graph $G_J=\{V_J, E_J\}$, where V_J is set of RNCFS, E_J is directed arc set and can be obtained by predigesting traffic pathing between any adjacent RNCFS as a pair of reversed parallel arc.

Theorem 1: ρ_{ij} can be denoted as a ordered set V_{ij}, and elements in V_{ij} denote origin destination and possible transfer stations of a_{ij} in ρ_{ij}, whose order obey by-passing such stations order of a_{ij} (Yan, 2004). Set V_{ij}' is a subset of V_{ij} which eliminates the first and the last element.

Definition 1: For ordered set V_{ij}, if another ordered set A satisfies $A \subset V_{ij}$ and $i, j \notin A$, then A is called as big-subset of V_{ij}.

Definition 2: x_{ij}^j, x_{ij}^k and x_{ij}^{Aij} are 0-1 variables. $x_{ij}^j=1$ when a_{ij} is not classified, otherwise $x_{ij}^j=0$, $x_{ij}^k=1$ when a_{ij} is only classified at station k, otherwise $x_{ij}^k=0$,

$x_{ij}^{Aij}=1$ when a_{ij} is classified at all stations included in set A_{ij}, otherwise $x_{ij}^{Aij}=0$, where A_{ij} is big-subset of V_{ij} for $i, j \in V_J$, $k \in V_{ij}'$ and $i \neq j \neq k$.

Step Function for the BCTFP model

The minimum container hour is taken as objective function for optimization of the block container train formation plan. Time consuming can be calculated as the sum of accumulation cost at original station and classifying cost along the network, because the time consuming in route are equal for containers classified or unclassified.

If block container train operates from station i to j, a_{ij} may be incorporated with any container flow passing over the two stations. The maximum container flow denotes as n_{aij} and corresponding accumulation cost denotes as $T_{ij}\cdot\lceil n_{aij}/M \rceil$, where $M>n_{aij}$ and $M \in Z$. $\lceil \cdot \rceil$ is unary operation, and its' algorithm is $\lceil x \rceil = y$, where $x \in R$ and $y \in Z$ and $x+1>y \geq x$. Corresponding objective function is as follows.

$$Z_{ij} = T_{ij}\left\lceil \frac{n_{aij}}{M} \right\rceil + a_{ij}\left(\sum_{k_1} x_{ij}^{k_1} t_{k_1} + \sum_{A_{ij}} x_{ij}^{A_{ij}} \sum_{k_2} t_{k_2} \right)$$

$$\text{subject to } x_{ij}^{j} + \sum_{k_1} x_{ij}^{k_1} + \sum_{A_{ij}} x_{ij}^{A_{ij}} = 1 \quad (k_1 \in V_{ij}' \ \& \ k_2 \in A_{ij})$$

Then combining all container sub-flows to get step function model M1

$$Z = \min\left\{ \sum_i \sum_j Z_{ij} \middle| \ x \in \Omega \right\} \qquad \text{(M1)}$$

In above function, vector x denotes all the decision-making variables, Ω is solution space for M1. Constraints of M1 are very simple, each container sub-flow corresponds to one equality constraint and there is no intercross among these constraints. However, step function cannot be decomposed, and according to our knowledge there are no mature algorithms at present in the literature.

0-1 Bi-level Linear Programming Model of BCTFP

Ankenbrandt presents that genetic algorithm (GA) takes $O(O(f)\, n\ln n)$ time at worst, where $O(f)$ shows fitness function complexity depends on feasible region (Ankenbrandt, 1991 and Xu, 2000). It can be concluded that for some combinatorial optimization problem, a polynomial-time algorithm can be taken with the help of GA if $O(f)$ is polynomial. Furthermore, with proper selection, GA converges to optimal solution. This shows obvious advantage of GA to solve such kind of problem.

However, M1 characterizes with a number of variables and complex objective function, due to which the length of bit string and the complexity of the fitness function may increase. To avoid this, it is necessary to transform and improve M1.

Ben-Ayed and Blair shown the methodology of knapsack problem into bi-level linear program (Liu, 1999), which proves BLP is a kind of NP-Complete problem and which also depicts that the problem of block container train formation plan can be solved by BLP. Because the number of through direction is

far less than the number of classifying scheme. M1 is changed into the following 0-1BLP model (M2).

$$BLP:\begin{cases} BLP_U: \begin{cases} \max Z=\sum_i\sum_j a_{ij}\sum_k t_k-\sum_i\sum_j T_{ij}x_{ij}-F \\ x_{ij}\in\{0,1\} \quad (i\neq j\&i,j\in V_J\ \&k\in V_{ij}') \end{cases} \\ BLP_L: \begin{cases} \min F=\sum_i\sum_j a_{ij}\sum_k t_k \\ x_{ij}\in\{0,1\} \quad (i\neq j\&i,j\in V_J\ \&k\in l(x_{ij})) \end{cases} \end{cases} \tag{M2}$$

In model M2, x_{ij} denotes whether running through BCT from RNCFS i to j. If through BCT exits, directed arc E_{ij} is constructed and added into network G_J, thus multiple given routes can be get. $l(x_{ij})$ denotes set of RNCFSs corresponded to given routing from i to j.

BLP_U of M2 is 0-1 Linear Programming with the objective of maximum accumulation saving. BLP_L is also 0-1 Linear Programming with the objective of minimum container classifying time along the routing in terms of variable x_{ij}. This model has no other complex constraints; therefore genetic algorithm is considered suitable for this model.

Algorithm Designing

Genetic Algorithm Parameters

(1) Construct matrix $(x_{ij})_{n\times n}$ to express individual v_i, with bit string length of L, and individual space of 2^L. Individual v_i corresponds one-to-one to feasible solution of M2. Besides, v_i is a directed graph incidence matrix.

(2) Take objective function of BLP_U as fitness function F which is the shortest route problem of two random spots and strictly obeys triangle inequality. For this problem, Fold algorithm is the best algorithm at present and its complexity under the most unfavorable situation is $O(n^2)$ (Wang, 1990).

(3) For chromosome with a long bit string, population size S is usually set at kL, where $k=1.5\sim2.0$ (Li, 2002). In order to contain more schema and avoid premature convergence of GA, population drift method (Yang, 2003) is adopted to choose better initial population.

(4) Elitist selection method and a smaller generation gap G=0.3 is adopted.

(5)Construct adaptive uniform crossover operator P_C and mutation operator P_m (Li, 2002 and Yan, 2004).

Cooperative Multi-colony GA

Gene-related may lead to GA deceptive problem, which would lead searching process to deceptive attractor and converge at exploitation in the end (Li, 2002). Because x_{ij} is determined by many train flows and controlled by other chromosome. In block container train formation plan problem, GA characterizes with strong gene-related and multi-modality values. Therefore some measures should be adopted for GA operation.

Cooperative multi-colony GA is able to solve schema deceptive problem. It follows the basic principle in which it independently evolves several sub-colonies at the same time. The only optimal solution for this problem exits when all these sub-colonies converge to it. Otherwise it gives the optimal solution of each sub-colony into set of niches core. Use calculated niches core radius to derate individuals of each sub-colony (Li, 2002 and Yan, 2004). Then the evolution process is repeated until the last optimal solution is worse than original optimal solution. Generally, scales of sub-colonies of cooperative multi-colony GA are same. The architecture of the process of GA parameters selection for our model M2 is illustrated in Fig.1.

Analysis of Time Complexity

The time complexity of the above procedure, can be summarized in following steps:

- Population drifting is used to generate initial population, the time it required is $O(kn^4)$.
- Sub-colonies allocation is in time $O(q)$.
- Each of sub-colonies' heredity evolution search with a scale of n/q, the time it required is $O(kn^4 \ln kn^2/q)$.
- It requires $O(q)$ arranging order of q optimal solutions.
- To compute the q optimal solutions with n^2 components, it takes $O(4qn^4)$ time at worst.
- It needs $O(2qn^4)$ to compute or compare each optimal solution $2n^2$ times.
- q times comparing and assigning to each individual needs $O(2kqn^2)$ time at worst.
- Output needs $O(kn^2)$ time.

Time complexity of each process of GA is tabulated below in Table. 1.

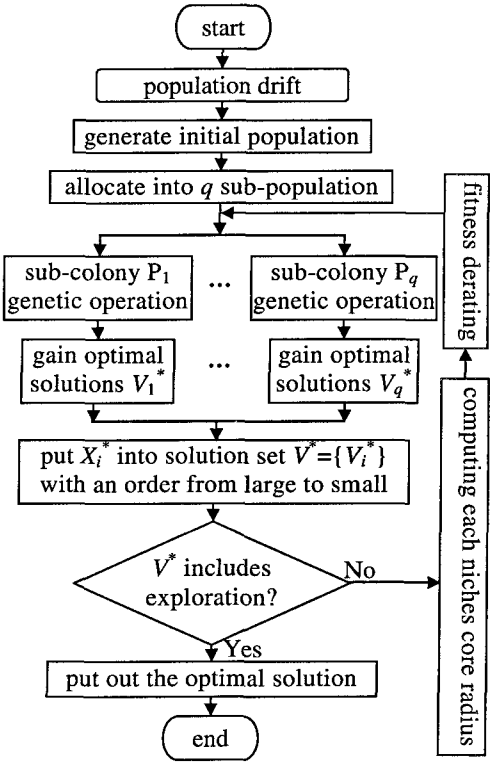


Fig.1 Genetic Algorithm Flow Chart

Conclusion

We presented the block container train formation problem using cooperative multi-colony genetic algorithm. It is concluded that four processes of the whole

algorithm, generating initial population, colony heredity evolution search, calculation of niches core radius and fitness derating, have higher time complexity. The time complexity of the algorithm is $O(\alpha n^4 \ln \beta n^2)$. Besides, it also can be deduced that this algorithm is efficient in global-optimal-converging algorithm.

Table. 1 Time Complexity of Each Process of GA

Calculation Process	Time Complexity
Initial population generation	$O(kn^4)$
Sub-colony allocation	$O(q)$
GA evolution of sub-colony for solution	$O(kn^4 \ln kn^2/q)$
Order arrangement of optimal solution set	$O(q)$
Computing niches core radius	$O(4qn^4)$
Computing fitness adjustment volume	$O(2qn^4)$
Individual fitness derating	$O(2kqn^2)$
Output optimal solution	$O(kn^2)$

References

Ankenbrandt, C.A. (1991). "An extension to the theory of convergence and a proof of the time comptexity of genetic algorithms." *Foundations of Genetic Algorithms*, Morgan Kcufmann, pp. 53-68.

Cao, Jiaming (1992). *Multiple Optimization on Wagon Flow Organizing in the Rail Network.* Southwest Jiaotong University, pp. 6-15, pp. 27-29, pp. 42-57.

Harmor,P., and Liu,Y.P. (1990). "Boolean Algebra Method on Combinatorial Optimization." *Journal of Mathematical Research & Exposition*, (10), pp. 300-313.

Holland, J.H. (1992). *Adaptive in natural and artifical systems: An introductory analysis with applications to biology, control, and artifical intelligence.* 2nd edition, Cambridge, MA: MIT Press.

Li,Minqiang, etc. (2002). *Fundamental Theory and Application of Genetic Algorithms.* Beijing: Science Press, pp. 233-244, pp. 387-399.

Liu,Shuan, etc. (1999). "Genetic Algorithms Based Approach for Bilevel Linear Programming Problem." *Journal of System Engineering*, 14(3), pp. 280-285.

Wang, Shuhe (1990). *Graph Theory and Algorithms.* press of university of science and technology of China, pp. 246-249.

Xu,Zongben, etc. (2000). "Theoretical Development on Genetic Algorithms:A Review." *Journal of Mathematic Progress*, 29(2), pp. 97-114.

Yan,Haifeng (2004). *Study on Containers transport Organization between Railway Network Container Freight Stations.* Southwest Jiaotong University, PhD.

Yang,Linmei, etc. (2003). "An Application of Genetic Algorithms with Initial Group Floating for Mixed Integer Nonlinear Programming." *Computing Techniques for Geophysical and Geochemical Exploration*, 25(3), pp. 253-258.

Criteria for a Steady-state Initialization in a Large-scale Railroad Network Simulation

Feng Xu[1] and Zong Tian, P.E.[2]

[1] Graduate Student, Dept. of Civil & Environmental Engineering, University of Nevada, Reno, Nevada, M/S 258, Reno, NV 89557, Email: xuf2@unr.nevada.edu
[2] Assistant Professor, Dept. of Civil & Environmental Engineering, University of Nevada, Reno, Nevada, Reno, NV 89557, Email: zongt@unr.edu.

Abstract: Computer simulation is well accepted as a technique to support the planning of transportation systems and analyses of infrastructure capacity. With the advance of computer simulation technologies, it requires more and more precise results from the simulators. As a critical step of a simulation experiment, the initialization process and the initial state of the simulator has a profound impact on the simulation process and the accuracy of the simulation experiment. To satisfy the need of providing a more realistic initial state for the simulation, a steady-state initialization process is preferred to the traditional transient process. Based on the railroad transportation system this paper proposed a model structure for the initialization process and the criteria set for checking whether a steady-state condition is reached during the initialization.
Keywords: Computer simulation; Initialization; Steady-state; Railroad transportation system.

Introduction

In general, there are two types of initialization processes, namely the transient and the steady-state. The transient initialization fills the network by generating stochastic variables based on a certain time point. The steady-state initialization, which requires a warm-up running period, attempts to establish a steady state by letting the simulation to run for a certain time and to fill the network with traffic.

The railroad transportation system is a complex system, consisting of many subsystems, which are interrelated. It is very difficult to find a set of transient stochastic parameters to reflect the real condition of a transportation network. On the other hand, the traffic in a railroad transportation system is not completely random, but rather controlled-random. Therefore, the steady-state method is a preferred method for simulating a railroad network.

Unlike the models proposed by other researchers, this proposed initialization process is accomplished in a separate subsystem, which interconnects with the other modules of the railroad system simulator. First, based on the characteristics of the railroad transportation system, a set of criteria is established to judge whether the initialization meets the requirements of steady-state for a simulation study. Then the paper provides the structure of initialization system for railroad transportation system simulator based on the network.

Criteria for Steady-State Initialization

The criteria for steady-state are established based on two conditions: the arriving and departing traffic and deviation from the timetable.

Criterion Based on Arrival and Departure

In railroad transportation systems, the variables of the traffic include not only the arrival and departure times of the trains, but also other attributes of the vehicles forming the trains, such as the destination of each vehicle, the number and the sequence of the vehicles in a train, and the type of a train. These variables are restricted by some other factors and not completely random. The neighboring stations interrelate on each other. It proves very difficult to use stochastic data to replace these variables. The steady-state initialization allows the simulation system to start with an empty network. After a certain simulation time, more trains and vehicles enter the network. This criterion is to decide whether the number and structure of these trains or vehicles satisfy the requirements of a simulation experiment in the scope of a study.

The empty vehicle flows are caused by the difference between the number of loaded and empty vehicles at each station. Previous studies did not specifically consider the empty vehicle traffic in the simulation *(2, 3, 7)*. Since the empty vehicle traffic flow is closely related to the loaded vehicle traffic flow, the initialization would not exactly reflect the real condition of steady-state without enough consideration of the empty vehicle traffic flow. This criterion not only considers the quantity of the loaded and empty vehicles but also considers the composition of the vehicles. So it consists of two parts: the criterion of quantity and criterion of composition.

The criterion of quantity: $$X = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(k_i + p_{ij} + q_{ij} + r_{ij})}{(2m + 2m \times n) \times \lambda} \quad \text{..... (1)}$$

This criterion is to check whether the quantity of the traffic flows after the warm-up running is within the thresholds of the real-world traffic flows. If the number of the simulated vehicle arrivals at station i, M^{i}_{Arri}, is within the range of the historical record, k_i is set to 1, otherwise, k_i is 0. If the number of empty vehicle arrivals of type j at Station i, N^{ij}_{Arri}, is in the range of the historical record, q_{ij} is set to 1, otherwise, q_{ij} is 0. The decision rules of p_i and r_{ij} are similar.

$$k_i = \begin{cases} 1 & M^{i}_{Arri} \in [M^{i}_{Min}, M^{i}_{Max}] \\ 0 & M^{i}_{Arri} \notin [M^{i}_{Min}, M^{i}_{Max}] \end{cases} \quad i = 1,2,\ldots,m \quad \text{..... (2)}$$

$$p_i = \begin{cases} 1 & M^{i}_{Dept} \in [M^{ri}_{Min}, M^{ri}_{Max}] \\ 0 & M^{i}_{Dept} \notin [M^{ri}_{Min}, M^{ri}_{Max}] \end{cases} \quad i = 1,2,\ldots,m \quad \text{..... (3)}$$

$$q_{ij} = \begin{cases} 1 & N^{ij}_{Arri} \in [N^{ij}_{Min}, N^{ij}_{Max}] \\ 0 & N^{ij}_{Arri} \notin [N^{ij}_{Min}, N^{ij}_{Max}] \end{cases} \quad i = 1,2,\ldots,m,\ j = 1,2,\ldots,n \quad \text{..... (4)}$$

$$r_{ij} = \begin{cases} 1 & N^{ij}_{Dept} \in [N'^{ij}_{Min}, N'^{ij}_{Max}] \\ 0 & N^{ij}_{Dept} \notin [N'^{ij}_{Min}, N'^{ij}_{Max}] \end{cases} \quad i = 1,2,\ldots,m,\ j = 1,2,\ldots,n \quad \text{..... (5)}$$

Where: i is the sequence number of stations;

j is the type of empty vehicle;

M^{i}_{Max} and M^{i}_{Min} are the maximum and minimum number of loaded vehicle arrivals per day at Station i;

M'^{i}_{Max} and M'^{i}_{Min} is the maximum and minimum number of loaded vehicle departures

per day from Station i;

N^{ij}_{Max} and N^{ij}_{Min} are the maximum and minimum number of empty vehicle arrivals of type j per day at Station i;

N'^{ij}_{Max} and N'^{ij}_{Min} are the maximum and minimum number of empty vehicle departures of type j per day from Station i;

M^{i}_{Arri} is the number of loaded vehicle arrivals per day at Station i;

M^{i}_{Dept} is the number of loaded vehicle departures per day from Station i;

N^{ij}_{Arri} is the number of empty vehicle arrivals of type j at Station i;

N^{ij}_{Dept} is the number of empty vehicle departures of type j from Station i;

λ is a parameter to be selected by the operator, which reflects the degree of satisfying the initialization criterion. For example, if λ is 0.9, 90 percent of vehicles during initialization need to satisfy the conditions set in Equations 2 through 5.

The criterion of composition:

$$Y = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}(k'_i + p'_i + q'_{ij} + r'_{ij})}{(2m + 2m \times n) \times \lambda} \qquad \text{..... (6)}$$

This criterion is to check whether the composition of the traffic flows after warm-up running meets the composition requirement of the real-world traffic flows. The decision rules are the same as the criterion of quantity. The form of the objective function is designed for easy calculation by computer.

$$k'_i = \begin{cases} 1 & P^{i}_{\text{Arri}} \in \left[P^{i}_{Min}, P^{i}_{Max}\right] \\ 0 & P^{i}_{\text{Arri}} \notin \left[P^{i}_{Min}, P^{i}_{Max}\right] \end{cases} \quad i = 1,2,\ldots,m \qquad \text{..... (7)}$$

$$p'_i = \begin{cases} 1 & P^{i}_{\text{Dept}} \in \left[P'^{i}_{Min}, P'^{i}_{Max}\right] \\ 0 & P^{i}_{\text{Dept}} \notin \left[P'^{i}_{Min}, P'^{i}_{Max}\right] \end{cases} \quad i = 1,2,\ldots,m \qquad \text{..... (8)}$$

$$q'_{ij} = \begin{cases} 1 & P^{ij}_{\text{Arri}} \in \left[P^{ij}_{Min}, P^{ij}_{Max}\right] \\ 0 & P^{ij}_{\text{Arri}} \notin \left[P^{ij}_{Min}, P^{ij}_{Max}\right] \end{cases} \quad i = 1,2,\ldots,m,\ j = 1,2,\ldots,n \qquad \text{..... (9)}$$

$$r'_{ij} = \begin{cases} 1 & P^{ij}_{\text{Dept}} \in \left[P'^{ij}_{Min}, P'^{ij}_{Max}\right] \\ 0 & P^{ij}_{\text{Dept}} \notin \left[P'^{ij}_{Min}, P'^{ij}_{Max}\right] \end{cases} \quad i = 1,2,\ldots,m,\ j = 1,2,\ldots,n \qquad \text{..... (10)}$$

The variables included in Equations 6 through 10 are given below:

P^{i}_{Max} and P^{i}_{Min} are the maximum and minimum percentage of loaded vehicle arrivals to the total vehicle arrivals per day;

P'^{i}_{Max} and P'^{i}_{Min} are the maximum and minimum percentage of loaded vehicle departures to the total vehicle departures per day;

P^{ij}_{Max} and P^{ij}_{Min} are the maximum and minimum percentage of empty vehicle arrivals of type j to the total vehicle arrivals per day;

P'^{ij}_{Max} and P'^{ij}_{Min} are the maximum and percentage of empty vehicle departures of type j to the total vehicle departures per day;

$P^{i}_{\text{Arri}} / P^{i}_{\text{Dept}}$ is the percentage of loaded vehicle arrivals/ departures to the total vehicle

arrivals at Station i;

$P_{\text{Arri}}^{ij} / P_{\text{Dept}}^{ij}$ are the percentage of empty vehicle arrivals/ departures of type j to the total vehicle arrivals at Station i;

The general criterion of arrival and departure traffic is:

$$Z = \mu_1 X + \mu_2 Y, \ \mu_1 + \mu_2 = 1 \quad \text{..... (11)}$$

μ_1 is the coefficient of criterion of quantity, which reflects the weighting of the quantity criterion in the general criterion;

μ_2 is the coefficient of criterion of composition, which reflects the weighting of the composition criterion in the general criterion.

When Z equals to 1, this criterion is met. As the initialization process usually does not produce a 100% match to the historical record, Z is usually smaller than 1. But different values can be selected for Z depending on the purpose of the study. For example, one study requires meeting this criterion at a 70 percent level, so Z needs to be at least 0.7.

Criterion Based on Timetable

Timetable defines not only the arrival and departure times at every station for each train, but also the running time and sequence of each train in the sections between stations. During the process of initialization, more and more traffic enter the network. As there is more than one station in the simulation, the traffic flow of every station will affect each other. There will be some deviation between the time of initialized traffic and the timetable that the simulation experiment needs. This criterion is based on the specific timetable and the initialization data to determine whether the deviation is acceptable.

Assume the initialization time is T_{Ini}, the following data can be obtained. To simplify the discussions, only the traffic flow at one station i is listed.

1. The simulated arrival time set of train m at station i in T_{Ini}: $\{ T_i^m \mid m$ is the number of the train arrived at station $i \}$.
2. The arrival time set of train m at station i in timetable: $\{ T_i'^m \}$.
3. The simulated departure time set of train n at station i in T_{Ini}: $\{ t_i^n \mid n$ is the number of the train departed from station $i \}$.
4. The departure time set of train n at station i in timetable: $\{ t_i'^n \}$.

Based on the above data the following data can be obtained:

In T_{Ini}, the deviation set of the simulated arrival and departure times of the trains and that of the timetable: $\{\Delta T \mid \Delta T = (\Delta T_1, \Delta T_2, \ldots, \Delta T_M, \Delta t_1, \Delta t_2, \ldots, \Delta t_N)\}$ .The distribution of ΔT is assumed to follow a normal distribution, which is considered reasonable.

1. $\{ \Delta T_m \mid \Delta T_m = T_i^m - T_i'^m, m = 1, 2, \ldots, M\}$ is the deviation of the simulated arrival time of the trains and that of the timetable
2. $\{\Delta t_n \mid \Delta t_n = t_i^n - t_i'^n, n = 1, 2, \ldots, N\}$ is the deviation of the simulated departure time of the trains and that of the timetable.

If the deviation is within an acceptable region, the simulated time is considered identical to the time in the timetable. In this case, we assume the deviation μ is 0 with the following hypothesis:

$H_0 : \mu = 0 \quad H_a : \mu \neq 0;$

The values of the sample are $X = \{\Delta T \mid \Delta T = (\Delta T_1, \Delta T_2, \ldots, \Delta T_M, \Delta t_1, \Delta t_2, \ldots, \Delta t_N)\}$

Test statistic is $\frac{\overline{X} - 0}{S/\sqrt{n}} \sim t(n-2)$, $n=M+N$

When $\left|\frac{\overline{X} - 0}{S/\sqrt{n}}\right| \leq Kt(1-\alpha/2; n-2)$, the null H_o is concluded, i.e., the deviation is acceptable *(8)*.

Although efforts have been made to construct the simulation model to better reflect the actual system operations, exactly matching the real-world conditions may be impossible. Errors also exist in the sample values. Hence, a positive integer K is selected to make the interval larger as shown in Equation 12:

$$-Kt(1-\alpha/2; n-2)S/\sqrt{n} \leq \overline{X} \leq Kt(1-\alpha/2; n-2)S/\sqrt{n} \quad \ldots\ldots\ldots\ldots (12)$$

During the time of initialization, T_{ini}, if the mean deviation of the simulated arrival and departure times of the trains and that of the timetable lies in this interval, the initialization meet the requirement of this criterion.

Structure of Initialization System

The distributed modeling framework is the trend of computer simulation techniques. Several system modules or servers work together to provide the system function. The initialization process discussed in this paper is based on a distributed system. Unlike the approaches used in previous research *(1,6)*, the initialization system is a separate subsystem and cooperates with other modules of the simulator. The structure of the initialization system is shown in Figure 3.

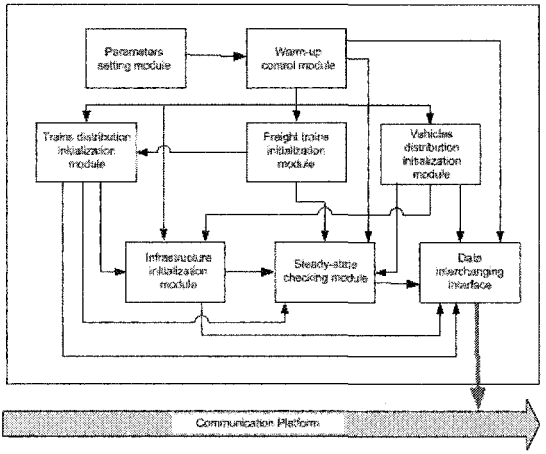


FIGURE 3: Structure of Initialization System

Parameters setting module is the interface for the operators to set the system parameters, including simulation scope, requirement to the infrastructure, special requirements for the abnormal simulation condition and so forth. These are the basis for the warm-up simulation process.

Warm-up control module is the core module of the initialization system. This module

controls the pace of the entire initialization and the data communication. Other modules, such as the *Trains distribution initialization module*, the *Freight trains* ***initialization module*** and the *Vehicles distribution initialization module* get the instructions from this module.

Infrastructure initialization module sets the parameters of the tracks, signals and other infrastructure according to the trains and the vehicles in the simulation scope.

Steady-state checking module determines whether the initialization has reached the steady-state. When the performance measures obtained from simulation meet the criteria of steady-state initialization, this module sends a message to the *Warm-up control module* to terminate the warm-up running process and sends a message to the data interchanging interface to release the initialization data to the others modules.

Concluding Remark

In this paper, we proposed a model structure for the initialization process and the criteria set for checking whether a steady-state condition is reached during the initialization. Unlike the models proposed by other researchers, our proposed initialization process is accomplished in a separate subsystem, which interconnects with the other modules of the railroad system simulator. In the initialization process, the *steady-state checking* module determines whether the initialization process can end by examining the system performance measures against the proposed criteria for steady-state.

There are several issues that will need further research, particularly, the method of choosing the warm-up running start point, and the efficiency aspect of the model.

Reference:

(1) Hu, S.J., *Theory of Railway Transportation Organization*, China Railway Press, Beijing, China, February 1993.

(2) Sun, W.H., Study of the Arriving Traffic, *Journal of Beijing Jiaotong University*, Beijing, China, October 1994. pp. 27-33

(3) Tan, L.G.., Yang, Z.X., Method of Arriving Traffic Flow Simulation, *Journal of Beijing Jiaotong University*, Beijing, China, December 1998.

(4) Xiong,G.L., *Continuous System and Discrete System Simulation*, Tsinghua University Press, Beijing, China, February 1991.

(5) Yang, Z.X., *Computer Simulation and Application*, China Railway Press, Beijing, China, 2001.

(6) Yang, Z.X., Mao, B.H., Sitution and Development of Railway Transportation System Simulators, *Journal of Beijing Jiaotong University Vol.26, No5*, Beijing, China, 2002. pp. 5-13

(7) Yang, H., *Organization of Railway Transportation*, China Railway Press, Beijing, China, 2001.

(8) Zhuang,C.Q., *Application of Statistic*, South China University of Technology Press, Beijing, China, July 2001.

Genetic algorithm based approach for the optimal location of traffic counts sections

E. Cipriani[1] and M. Petrelli[2]

[1]Dipartimento di Scienze dell'Ingegneria Civile, Università di Roma Tre, via Vito Volterra 62, 00146 Roma (I); PH (+39) 0655173406; FAX (+39) 0655173441; email: eciprian@uniroma3.it

[2] Dipartimento di Scienze dell'Ingegneria Civile, Università di Roma Tre, via Vito Volterra 62, 00146 Roma (I); PH (+39) 0655173402; FAX (+39) 0655173441; email: labtras@uniroma3.it

Abstract

In this paper, the authors propose a genetic algorithm approach for solving the problem of optimal location of traffic count sections. This approach aims at assessing the efficacy of a transportation based rule proposed by authors in previous studies. These approaches are tested on a real size extra-urban road network and are compared to a state-of-the art rule that is widely used in practical applications.

Introduction

The availability of an updated and reliable Origin-Destination (O/D) demand matrix is a key requirement in transportation planning and engineering. The direct statistical estimate of the O/D matrix from O/D surveys represents a very expensive method in time and costs because it involves in interviews collection. A different method consists in the estimation of the O/D matrix based on traffic counts collected on a subset of links of the transport network; it implies lower costs and time and provides input data for the so called O/D count based estimation procedure. With reference to the latter method, the location of count sections is strategic to obtain reliable estimation of O/D demand matrix.

To our knowledge, research has focused mainly on the aspects of the O/D matrix estimation accuracy depending on input data and modeling approach rather than the quality of the traffic count sections. The problem of the optimal location of traffic counts sections (OLTCS) has been studied by Lam (1991) who proposed an efficient and systematic sampling procedure in absence of prior information (previous matrices or traffic counts) in which reliability of estimated matrix is associated to location and number of count sections. Yim and Lam (1998) introduced the Maximal O/D Selection method based on the maximization of both the total and the net O/D flows captured. In the same context, Yang and Zhou (1998) suggested to adopt the O/D coverage rule and recommended to locate count sections so as to capture independent flows. Two new rules for solving the OLTCS problem have been proposed by Cipriani et al. (2004) that aim at enhancing the reliability of the estimated O/D matrices in case of limited availability of sections to be located. Last

years, new stocastic techinques as Tabu Search, Genetic Algorithm etc. have been tested with positive results to solve optimization problems. The notion of Genetic Algorithms (GA) was introduced by Holland (1975) and extended by Goldberg (1989). For the past fifteen years, the increasing interest in GA for solving large scale problems has shown their powerful suitability for optimization, design and control applications. Also in the transportation field the use of GA is growing for solving network design and signal control problems.

In this paper, authors' aim is to assess the efficacy of the demand fractions maximization rule introduced in Cipriani et al. (2004) by comparing it with the one of a GA based heuristic, due to its suitability for solving combinatorial problems. The paper is organized as follows: next section describes the state-of-the-art rules and the demand fractions maximization one for solving the OLTCS problem. Then, the GA based approach is introduced and the results of numerical experiments on a real size road network are presented.

State-of-the-art rules for the location of traffic count sections

By following Yang and Zhou (1998) and as reported in Cipriani et al. (2004), a reliability index of the estimated matrix can be defined. Let:

- $G(N,A)$ be a road network, with N the set of nodes and A the set of links;
- $\hat{A}$ be the subset of links where traffic count sections are located;
- W be the set of O/D pairs;
- t^*_w be the actual demand between the w-th O/D pair;
- t_w be the estimated demand between the w-th O/D pair;
- p_{aw} be the fraction of demand between pair w using link $a \in \hat{A}$;
- v_a be the observed flow on link $a \in \hat{A}$.

Both the actual and the estimated O/D matrices must obey the following conditions:

$$\sum_{w \in W} p_{aw} t_w = v_a, \; a \in \hat{A} \tag{1}$$

$$\sum_{w \in W} p_{aw} t^*_w = v_a, \; a \in \hat{A} \tag{2}$$

So that their difference is:

$$\sum_{w \in W} p_{aw} (t^*_w - t_w) = 0, \; a \in \hat{A} \tag{3}$$

If the relative deviation between the estimated and the true matrix is denoted by $\lambda_w = \left(t^*_w - t_w\right) \Big/ t_w$, with $\lambda_w \geq -1$, expression (3) becomes:

$$\sum_{w \in W} p_{aw} t_w \lambda_w = 0, \; a \in \hat{A} \tag{4}$$

Reliability index of the estimated matrix with m O/D pairs can be defined as follows:

$$G(\lambda) = \sqrt{\sum_{w \in W} \lambda_w^2 \Big/ m} \tag{5}$$

The smaller is the value of $G(\lambda)$ the more accurate is the estimate of the matrix. Thus, the maximal possible relative error can be computed as the maximum $G(\lambda)$ complying with constraint (4).
Based on the previous statements, if an O/D pair has p_{aw}=0 for all $a \in \hat{A}$, that is to say if no fraction of trips is collected at any count location, expression (4) is always satisfied whatever is the value of λ_w. This implies that the maximal possible relative error may be equal to infinite. To avoid this, according to the previous studies analyzed, the following rules for the location of the traffic count sections can be stated:
1) *O/D Covering Rule: traffic count sections should be located so as to intercept every O/D pair, even in a small fraction of trips*;
2) *Maximal Flow Fraction Rule: for an O/D pair, traffic count sections should be located in order to maximize the fraction of intercepted flow*;
3) *Link Independence Rule: traffic count sections should be located on those links that involve linearly independent flows*;
4) *Maximal Flow Intercepting Rule (Max-Flow): fixed the number of traffic count sections, they should be located in order to maximize the intercepted flows.*
Financial and time constraints acting in practical applications require that only a limited number of traffic count sections can be located thus implying that the O/D covering rule cannot always be adopted. This occurrence makes information relative to any additional O/D pair very attractive. This consideration is the basis for the rule introduced in Cipriani et al. (2004) and reported below:
5) *Maximal O/D Demand Fraction Rule (MAX-DF): traffic count sections should be located along those links that maximize the sum of intercepted O/D demand fractions.*
According to the results illustrated in Cipriani et al. (2004), MAX-DF rule revealed to be very effective in improving the estimation of the O/D matrix capturing more O/D pairs than the usual methods. This rule is similar to the "Maximal Flow Fraction" one but, for any O/D pair, the flow fraction is computed as the ratio between the path flow t_{aw} against the number of trips of the O/D pair itself, rather than against the flow on link a. This modification has been introduced in order to restrain the occurrence that, being limited the number of count locations, a link with few path flow components (i.e.: traveled by few O/D pairs) may be selected.

In this paper, the authors propose a GA based approach for the OLCTS problem solution. Aiming at exploiting the efficacy of both the 2 previous rules (4 and 5), this approach maximizes both the intercepted flows and intercepted demand fractions.

Overview of GA

GA are random search optimization heuristics inspired by biological evolution and based on concepts of natural selection and genetics. They are population based techniques that preserve the set of potential solutions, called population, during the search. Differently from traditional methods, GA work with a coding of potential solutions rather than with solution values themselves; besides, in GAs each potential solution, called chromosome, evolves through successive

iterations adopting a probabilistic transition scheme instead of gradient information. Each chromosome is evaluated assigning a fitness value on the basis of which genetic operators are utilized to generate the next population. "Best" chromosomes of any population tend to reproduce and survive to the next iteration, thus improving successive generations. GA explore all regions of the solution space adopting a searching mechanism based on the genetic operators of selection, crossover and mutation. These specific features make this technique suitable for a wide range of applications without any of the limitations imposed by local searching methods (i.e. continuity, derivative existence, presence of discrete variables, etc...). To our knowledge a widespread analysis on the performance of GA on the optimal location of traffic count sections has not been carried out yet.

Description and implementation of GA based heuristic for the OLTCS problem

The adopted GA based algorithm is composed by the following steps:

1) *A chromosomal representation of the solutions.* The adopted representation is a n - multiple string of length k where n is the number of candidate links and k is the number of counting sections. This implies that the location of 10 counting sections requires that any solution is represented by a string composed by 10 genes each one chosen among 1,700 candidate links (as in our case). The adopted representation is equivalent to an n – length binary string obeying the constraint that the total number of 1s (counting sections) is always fixed, and equal to 10 in our example. This type of representation yields that the number of possible solutions is equal to $n! / k!(n\text{-}k)!$.

2) *An estimation of the population size.* The estimation of the population size is based on the following consideration: in the n – length binary representation with 2^n possible solutions suggested values of the population size range from n to $2n$. Therefore we assumed the following rule (choosing the lower value): *size* = $\log_2$ (*possible solutions*); in our case we obtain: *size* = $\log_2$ ($n! / k!(n\text{-}k)!$) that, for high values of n can be solved adopting the Stirling approximation which shows that $n! \cong (2\pi n)^{0.5} (n / e)^n$.

3) *An evaluation function of the solutions.* Any solution is evaluated by performing an assignment of the O/D matrix and computing the sum of the intercepted O/D pairs and intercepted O/D demand; specifically the evaluation function consists in the maximization of the sum of the 2 percentages.

Formally, it can be defined as:

$$\max\left\{\sum_a p_{aw} + \left|\hat{W}\right| \Big/ |W|\right\}, \forall a \in \hat{A}, \forall w \in \hat{W} \qquad (6)$$

where p_{aw} is the fraction of demand between pair w using link a and $\hat{W}$ is the set of O/D pairs intercepted by all the links a belonging to the subset $\hat{A}$.

Then the adoption of a fitness function allows to measure the goodness of the solution in terms of reproduction opportunities; that is to say, while the evaluation function shows the performance of the single solution independently from other

solutions, the fitness function provides an assessment of the performance of the single solution as related to those of the other ones. The adopted fitness function simply assigns to solutions values linearly ranging from 0 to 1 corresponding to the worst and best values of the evaluation function, respectively.

4) *Genetic operators that generates a new population of solutions*. The generation of a new population at next iteration is given by genetic operators that perform selection and recombination of solutions. Selection operator selects solutions with highest fitness values for the recombination phase where crossover and mutation execute mixing. An elitist selection type has been considered that allows better solutions to pass to the next generation without any recombination. The selected solutions are randomly chosen two by two for the first step of the recombination phase where the multipoint crossover acts. In the second step of the recombination phase, the offspring solutions resulting from crossover are randomly chosen for the mutation operator that allows to introduce new genes.

The GA based algorithm described above has been implemented in the EMME/2 macro language and it stops if the fixed number of iterations is reached or if the improvement of the best value of evaluation function is lower than 2.5% in the last 25 iterations.

Numerical experiments

Numerical experiments have been conducted on the road network of the Campania region, Italy, illustrated in Figure 1. It has been schematized by a graph composed of about 560 zones, 1,500 nodes and 5,000 links. Estimated peak hour demand amounts to 475,000 veh/h.

Figure 1. Real size test network.

The percentage of intercepted O/D pairs and demand levels has been computed by applying the proposed GA based heuristic as the number of counted sections ranges from 1% to 5% of candidate links. Results have been utilized to assess the efficacy of the MAX-DF rule; analogous results obtained by the

application of the Max-Flow rule are reported as reference. Preliminary tests have been carried out to calibrate the GA parameters improving the solution quality.

Table 1 summarizes the results; as it is possible to see the percentage of O/D pairs intercepted by the GA based heuristic is always higher than that captured by the Max-Flow rule while the percentage of O/D demand caught by GA heuristic is nearly always the same as that resulting from the application of the Max-Flow rule.

The comparison with the MAX-DF rule shows that the GA based heuristic intercepts similar percentages of O/D pairs while better results are obtained in terms of percentage of intercepted O/D demand.

As expected, on the whole the application of the GA based heuristic allows to maximize the sum of the 2 percentages, except in the case of 3% and 4% of candidate links.

Table 1. Results of the three OLTCS techniques.

	% O/D Pairs					% O/D Demand					SUM				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
GA	19,91	34,89	32,59	36,35	40,44	10,47	29,13	27,32	34,78	51,99	30,38	64,02	59,91	71,13	92,43
MAX-DF	20,00	32,65	38,36	39,95	40,83	5,18	12,89	22,90	30,07	33,50	25,18	45,54	61,25	70,02	74,33
Max-Flow	11,61	21,23	28,31	33,75	36,43	16,57	32,01	41,55	48,37	53,33	28,18	53,24	69,86	82,12	89,76

Conclusions

In this paper, a GA based heuristic for solving the OLTCS problem has been proposed combining the features of 2 transportation based rules (MAX-DF and Max-Flow). The comparison of these techniques has been carried out on a real size network. Results show the efficacy of the GA based heuristic in maximizing both the intercepted flows and intercepted demand fractions. The proposed approach seems to be promising for reliable estimation of demand matrix. Further developments will be focused on this topic.

References

Cipriani E., Fusco G., Gori S. (2004). "Optimal location of traffic count sections". *ICTTS 2004 Conference Proceedings*, ASCE.

Goldberg D.E. (1989). "Genetic algorithms in search, optimization and machine-learning". Addison-Wesley Publishing Co., Reading, Mass.

Holland J.H. (1975). "Adaptation in natural and artificial systems". The University of Michigan Press, Ann Arbor, MI.

Lam W. H. K. (1991). "Estimation of O/D matrix from traffic counts: How to determine the count location?". *HKIE Transactions*, Vol. 2, N° 1, pp. 9-16.

Yang H., Zhou, J. (1998). "Optimal traffic counting locations for original-destination matrix estimation". *Transportation Research Part B*, Vol. 32, pp 109-126.

Yim P. K. N., Lam, W.H. K. (1998). "Evaluation of count location selection methods for estimation of O/D matrices". *Journal of transportation engineering*. July-August 1998, pp.376-383.

AN APPLICATION OF NEURAL NETWORK IN CORRIDOR TRAVEL TIME PREDICTION IN THE PRESENCE OF TRAFFIC INCIDENTS

Yang Tao, Ph.D.[1] and Bin Ran, Ph.D.[2]

[1]Traffic Engineering Division, City of Madison, Wisconsin; 215 Martin Luther King, Jr. Blvd., Madison, WI 53701; PH (608) 266-4815; FAX (608) 267-1158; email: ytao@cityofmadison.com

[2]Professor, Department of Civil and Environmental Engineering, University of Wisconsin-Madison, 1415 Engineering Drive, Madison, WI 53706; PH (608) 262-0052; FAX (608) 262-5199; email: bran@engr.wisc.edu

Abstract

Incident related traffic congestion leads to enormous economic loss each year in the world. To predict the traffic situation when an incident occurs and disseminate the information to the traveling public can alleviate the traffic congestion caused by the incident. This research collected traffic incident data and traffic condition data from a highway corridor (Interstate Highway 66 Eastbound) in Northern Virginia. After pre-processing, data fusion between the two different data sources was successfully conducted and the cross reference between the incident and traffic condition data sets was set up. Based on the fused data sets, a neural network model for corridor travel time prediction in the presence of traffic incidents was developed. After the model was trained and optimized, randomly selected new data was used to test the performances of the proposed model under three different scenarios: (1) the input variables were incident related information only; (2) the input variables were current traffic condition information only; (3) the input variables included both the incident related information and the current traffic condition information.

The performance indicators of the model were calculated under the three scenarios, and the statistics were compared. The results demonstrate that it is possible to accurately predict the future travel time within a corridor in the presence of traffic incidents when given sufficient amount of data. With exceptional learning ability, neural network is proven to be an effective tool in modeling this travel time prediction problem. The developed neural network delivers a good fit in most cases, indicating that it is a successful model. It is also observed that incident related information roughly dictates the trend of the impact on traffic, while current traffic condition provides a dynamic environment where the incident occurs. Addition of current traffic condition information can further improve the prediction accuracy. The predicted travel time information is valuable for traveler information systems and traffic incident management systems.

1. Introduction

Incident related traffic congestion leads to enormous economic loss each year in the world. It cost the United States $25.6 billion in year 2000. To predict the traffic situation when an

incident occurs and disseminate the information to the traveling public can alleviate the traffic congestion caused by the incident. However, relatively little research has been conducted on this topic due to the lack of reliable data sources, poor data fusion tools, and modeling difficulties.

Artificial Neural Network (ANN) is also referred as neuromorphic system, connectionist architecture or parallel distributed processing (Mitechell, 1997; Haykin, 1999). Its basic rationale is to emulate the densely interconnected, parallel structure of the human brain information processing system. The neural network performs its computation through the process of learning from "experiences", namely the data samples. Since its birth in 1940's, the structure of the neural network has evolved from a simple neuron to the complicated multi-layer network, and its learning algorithm has developed from the naïve adaptive learning to the elegant back-propagation learning. The neural network is noted for its exceptional learning capability and robustness. Due to this reason, the technique has been welcomed by transportation researchers (Dougherty et al., 1993; Smith and Demetsky, 1994; Park and Rilett, 1998; McFadden et al., 2001; Lint et al., 2002). In this research, we explore the application of neural network on a new topic: corridor travel time prediction in the presence of traffic incidents.

2. Experiment Corridor and Data

A corridor in Fairfax County, northern Virginia was selected as the experiment route. The corridor is a segment of East Bound of Interstate Highway 66 (milepost 46.9 to 64.7). It is a major corridor connecting Virginia and Washington D.C. The modeling problem is that if there is a specific incident at a specific time on the route, what the travel time will be.

Both the traffic condition and incident data were collected from Archived Data Management System (ADMS) Virginia. ADMS Virginia is a data archiving system funded by the Federal Highway Administration (FHWA) and the Virginia Department of Transportation (VDOT). It archives traffic condition and incident related data collected from Hampton Roads Region and Northern Virginia Area. The main purpose of the system is to provide information services to help measure the operation and performance of the transportation system and Transportation Management Centers (TMCs).

The traffic condition data and incident data were collected from May 1st, 2004 to August 1st, 2004 for consecutive three month. The traffic condition data included corridor travel time, flow and occupancy. The flow and occupancy was one-minute aggregated data and all the three values are updated on one-minute basis. The incident data included occurrence time, type, duration and related lane closures. During the three-month period, totally 971 traffic incidents were collected from the corridor.

The mapping between the incident data set and the traffic condition data set is called data fusion. It is to develop the cross reference relationship between the two data sources. The current flow and occupancy information and the travel time 15 minutes later were mapped to each incident record. The incident records which cannot be mapped were deleted from the data set. After the process, 944 records were left.

3. Model Establishment

Figure 1 indicates the neural network architecture developed. The input variables to this model could be categorized into two groups: incident related information and traffic condition information. The output of this model was the predicted travel time of the corridor 15 minutes after the incident happened.

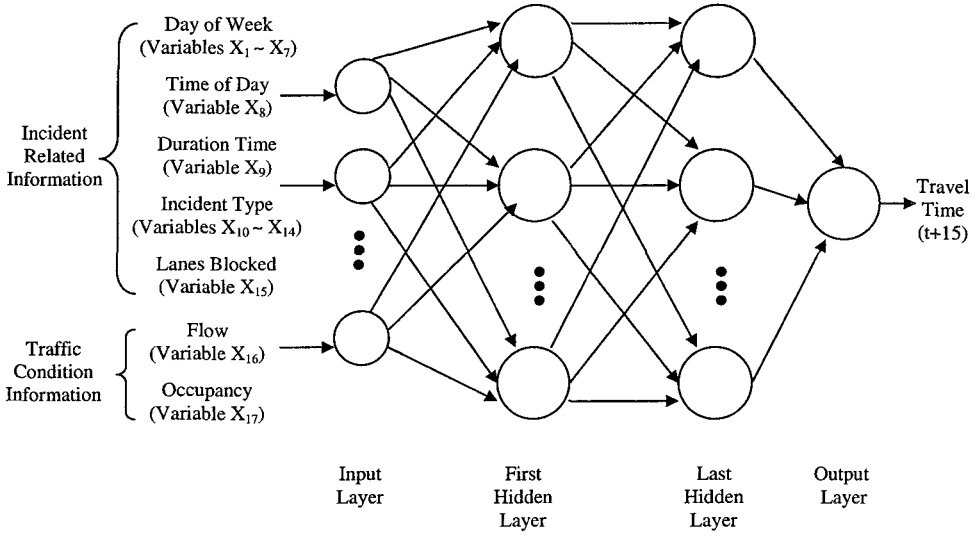


FIGURE 1. A multilayer perceptron architecture developed.

Incident related information was the information that could be associated to a particular traffic incident. It included:

- Day of the week: the day when the incident happened.
- Time of the day: the time when the incident happened. It was a consecutive time value within a 24-hour frame.
- Duration time: how long did the incident exist on the roadway.
- Incident type: the type of incident, which was obtained from the incident's descriptions. There were 5 types: Traffic Collision, Disabled Vehicle, Debris, Towing Activity, and Other Incidents.
- Lanes blocked: the lane closure information associated with the incident. It ranged from 1 lane blocked to 4 lanes blocked.

Traffic condition information indicated what the traffic was like when a particular incident happened. Included in the model, traffic flow and occupancy were two variables used to capture the dynamic traffic situation when the incidents happened. Traffic speed was not included since it could be estimated from the traffic flow and occupancy at a specific time point. It would become redundant if it was included.

- Traffic flow: the average number of vehicles that crossed a sensor station in the one-minute interval.
- Occupancy: the percentage of time there was a vehicle above the sensor station during the one-minute interval. This was also an average number across the whole corridor.

The processed data were further divided into three data sets: training (708 records), cross validation (142 records), and the testing (94 records). The training data set was fed to the neural network model, so the model could recognize the pattern and adjust its weights for the parameters. The cross validation data set was used to calibrate and tune the neural network model. With it, the error with respect to this validation data set was kept track of, while the training data set was used to drive the gradient descent search. When the minimal error of the cross validation data set was achieved, the training process was terminated. This is called "early stopping", which helps to prevent over-fitting.

4. Experiment Results

After the model was trained and optimized, the new testing data set was used to examine the performance of the proposed model. Three scenarios were studied and their results were compared:

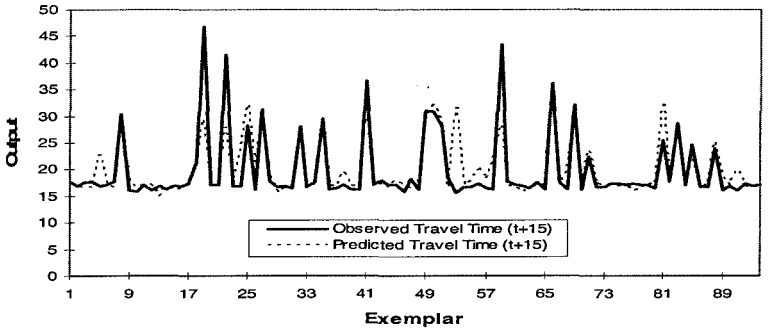
- Scenario 1: the input variables were incident related information only;
- Scenario 2: the input variables were current traffic condition information only;
- Scenario 3: the input variables included both the incident related information and the current traffic condition information.

Figure 2 shows the comparison between the predicted travel time and observed true travel time (or desired value) for each of the scenarios. The data points on each curve represent 94 new incident records used in the testing data set. The unit of the travel time is minute. The performance of the model under each scenario was also calculated, and the statistics are compared in Table 1.

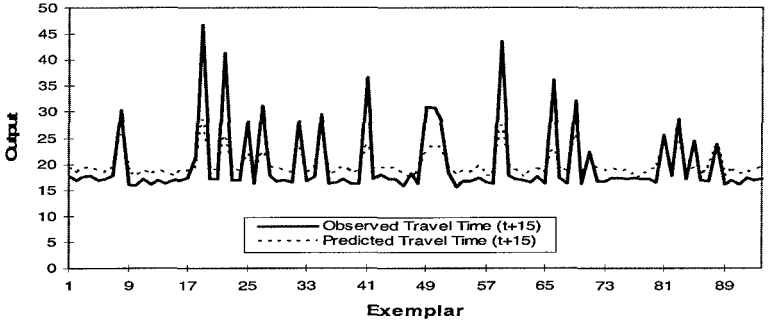
Table 1. Performance of the Model under Different Inputs

Performance Indicator	Scenario 1 Incident Related Information Only	Scenario 2 Traffic Condition Information Only	Scenario 3 Both Information
MSE (min^2)	14.288	19.536	7.440
NMSE	0.332	0.454	0.173
MAE (min)	1.913	2.933	1.618
Min Abs Error (min)	0.054	0.037	0.025
Max Abs Error (min)	17.275	18.328	10.240

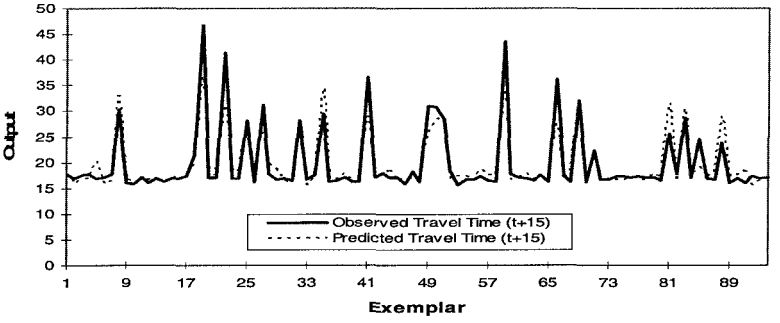
Note: MSE stands for Mean Squared Error. NMSE stands for Normalized Mean Squared Error. MAE stands for Mean Absolute Error. Min Abs Error stands for Minimum Absolute Error. Max Abs Error stands for Maximum Absolute Error.



Scenario 1: Incident related information only



Scenario 2: Current traffic condition information only



Scenario 3: Both incident and traffic condition information

FIGURE 2. Predicted travel time versus observed travel time.

The result indicated that the incident related information was very helpful in the travel time prediction in the presence of incidents. By using the incident related information alone, an

acceptable prediction result could be obtained (MAE = 1.9 minutes. For the desired value of the testing data set, the standard deviation was 4.4 minutes, and the value ranged from 15.5 minutes to 41.1 minutes). However, with the supplement of the current traffic condition data, the accuracy of the prediction could be further improved (MAE = 1.6 minutes).

5. Conclusion

The experiment results demonstrate that it is possible to accurately predict the future travel time within a corridor in the presence of traffic incidents when given sufficient amount of data. With exceptional learning ability, neural network is proven to be an effective tool in modeling this travel time prediction problem. The developed neural network delivers a good fit in most cases, indicating that it is a successful model. It is also observed that incident related information roughly dictates the trend of the impact on traffic, while current traffic condition provides a dynamic environment where the incident occurs. Addition of current traffic condition information can further improve the prediction accuracy. The predicted travel time information is valuable for traveler information systems and traffic incident management systems. However, the limitations of link-based models determine that the parameters in the model shall be recalibrated using new local data when applied to other highway corridors.

References

Dougherty, M. S., Kirby H. R., and Boyle R. D. (1993). "The Use of Neural Networks to Recognize and Predict Traffic Congestion." *Traffic Engineering & Control*, 34(6), 311-314.

Haykin, S. (1999). *Neural Networks: A Comprehensive Foundation*. Prentice Hall. Upper Saddle River, NJ.

Lint, J.W.C., Hoogendoorn, S. P., and van Zuylen, H. J. (2002). "Freeway Travel Time Prediction with State-Space Neural Networks: Modeling State-Space Dynamics With Recurrent Neural Networks." *Transportation Research Record* 1811, TRB, National Research Council, Washington, D.C.

McFadden, J., Yang, W-T, Durrans, S. R. (2001). "Application of Artificial Neural Networks to Predict Speeds on Two-Lane Rural Highways." *Transportation Research Record* 1453, TRB, National Research Council, Washington, D.C., 98-104.

Mitechell T. (1997). *Machine Learning*. McGraw-Hill Press. New York, NY.

Park, D. and Rilett, L. (1998). "Forecasting Multiple-Period Freeway Link Travel Times Using Modular Neural Networks." *Transportation Research Record* 1751, TRB, National Research Council, Washington, D.C., 9-17.

Smith, B. L. and Demetsky, M. J. (1994). "Short-Term Traffic Flow Prediction: Neural Network Approach." *Transportation Research Record* 1453, TRB, National Research Council, Washington, D.C., 98-104.

Multi-layer Traffic Signal Control Model based on Fuzzy Control and Genetic Algorithm

(Li Ruimin[1]. Lu Jiangang[2], Lu Huapu[3])

[1]Institute of Transportation engineering, Tsinghua University, Room 312, Heshanheng Building, Tsinghua University, Beijing , China, 100084; PH(86-10)62772615; FAX (86-10)62795339; email: lrmin@tsinghua.edu.cn
[2]Department of Civil and Environmental Engineering, University of Wisconsin at Madison, Madison, WI 53706; PH(1-608-262-2524); email: jiangang@cae.wisc.edu
[3]Institute of Transportation engineering, Tsinghua University, Room 312, Heshanheng Building, Tsinghua University, Beijing , China, 100084; PH(86-10)62795339; FAX (86-10)62795339; email: luhp@tsinghua.edu.cn

Abstract

In this paper, we propose a multi-layer traffic signal fuzzy control model based on genetic algorithm (GA) for isolated intersections. The model includes three layers. The first layer is traffic demand prediction. By using a comprehensive index, traffic demand intensities (TDI), we estimate the traffic demand of each approach lane during green time. The second layer, called phase sequence fuzzy controller, is implemented to optimize signal phases according to traffic flow conditions. The third layer is green time fuzzy controller. TDI and the phase sequence are used to determine whether the current signal phase will be extended or terminated. In our research, generic algorithm will be adopted to determine the membership function of this multi-layer fuzzy control model. The performance of this control model will be compared to the model without membership function optimization at a simulated four-approach intersection, which will show the control model presented outperforms the traditional model in reducing average total delay at an intersection.

Keyword: traffic signal control, fuzzy control, genetic algorithm

Introduction

In recent years, many researchers began to apply the fuzzy control theory in the traffic signal control field, and expanded the model from the isolated intersection two-phase control to the arterial coordinated control (Stephen and Sujeet, 1993; Niittymäki and Kikuchi, 1998; Mohamed et al., 1999; Ella, 2001; Niittymäki, 2001). In fuzzy control models, some factors, such as the fuzzy input parameters, fuzzy control rules and optimization of membership function, will be critical in traffic control. Optimization of the membership function parameter is the main focus of our research. We propose a traffic signal multi-layer fuzzy control model based on genetic algorithm (GA) to optimize the membership function parameters and finally improve the efficiency of traffic control.

The paper is organized as follows. In section 2, we will have a discussion for the isolated intersection multi-layer fuzzy control models. The membership function parameter optimization model based on genetic algorithm will be proposed in section 3. In sections 4, the performance of this control model is compared to the model without membership function optimization at a simulated four-approach intersection. Finally, section 5 contains some conclusions and future works.

Multi-layer fuzzy control model

Our research is focusing on how to optimize the phase sequence and green time. To improve the response speed of traffic control devices, control rules should be simplified if possible. The multi-layer traffic signal control model we proposed for isolated intersection will effectively reduce the number of control rules, and then improve the efficiency.

The first layer of fuzzy controller

The function of the first layer is fuzzy judgment and reasoning of traffic flow conditions, that is, to compute the value of Traffic Demand Intensities for every phase according to traffic flow parameters.

Signal control is generally considered as a competitive problem. One of the methods used to solve this problem by assigning the green time is to calculate demand intensities from different approaches, which we define here as "Traffic Demand Intensities"(TDI) of lanes from different approaches.

TDI indicates the demand degree of lane's green time for each phase in future, and there are two things we need to pay attention to. First, it is about the TDI of each phase at each time. It depends on the queue length of vehicles on different lanes under this phase; Second, we need to focus on the TDI of this phase in a short time in future, which depends on the arrival traffic volume in the coming period, such as 10s. Without considering factors of special control, such as bus priority, TDI is now the function of arrival traffic volume in Δt and length of queues.

In the process of fuzzy reasoning of TDI, there are two input factors: $p_{\Delta t}$, the number of vehicles which pass the detector in Δt from every approach under a phase, and q_t, the length of queue at the time of t. The output is the value of TDI, tdi. Trapezoidal membership functions are used for computation.

We may establish 25 items of reasoning rules according to the experience of experts, shown as follows.

if q_t is very small and $p_{\Delta t}$ is very small, then TDI will be very small;

if q_t is very small and $p_{\Delta t}$ is small, then TDI will be small;

etc;

The TDI of each phase will be used as the input of traffic signal fuzzy control for one intersection in the following section.

The Second layer of fuzzy controller

The function of the second layer is to optimize phases sequence according to traffic flow conditions, which we generally call phase sequence fuzzy controller.

The four-phase control intersections generally adopt the phase sequence as A—B—C—D, shown in Figure 1. The optimization for control policy will mainly used to decide which one of B、E、F should run after A, and which one of D、G、H should run after C. The input of phase optimization fuzzy controller is the TDI of every phase, which is determined by the first layer of fuzzy controller. General principles of phase optimization are determined by TDI under every phase. When A or C are running, the next phase is judged and obtained. For example, if A is running, the next phase of A will be B, E, or F. During the judge phases, we need to calculate TDI for B, E, or F separately, and then obtain $TDI(B)$, $TDI(E)$, $TDI(F)$, which will be used to judge the next phase according to the volumes. The rule can be described as follows:
If $TDI(B)$ is largest, then next phase will be B; Etc.

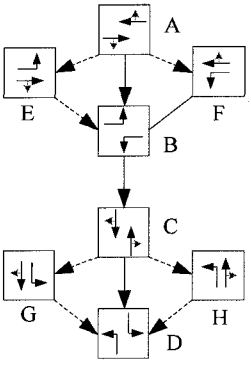


Figure 1. Phase sequence

The Third layer of fuzzy controller

The function of the third layer, the green time fuzzy controller, is to decide whether the green time of each phase should be extended and the volume of extension.

The input of this layer depends on the output of the first two layers. One of input is $TDI(g)$, which represents the value of TDI under current green phase; The other is $TDI(n)$, which represents the value of TDI for the next phase, determined by the second layer of fuzzy controller. The membership function of $TDI(g)$、$TDI(n)$ adopts Trapezoidal membership functions and it has similar output as the first layer.

The output of green time fuzzy controller is the volume of time Δg which the current phase should be extended, and the minimum volume is 0 second (0 means that do not extend the green time), with the maximum volume 18 second.

Optimization of the fuzzy control parameters

Model Framework

As shown in the above fuzzy control model, there are many membership function parameters which are difficult to select manually. In practice, in a middle sized city, there are hundreds of intersections which have various traffic flow, control objective and etc. So it is necessary for the control model to have the real-time learning function to improve the system performance. In our research, we use GA to optimize the membership function parameter on-line. The control model based on GA is shown in Figure 2.

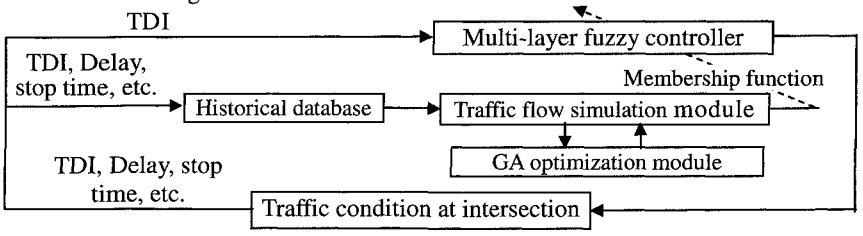


Figure 2. Multi-layer fuzzy control model structure based on genetic algorithm

In practical control, first we set up the membership function parameters manually based on expert experience, at the same time the traffic flow data and corresponding estimation indexes data have been achieved in the historical database. After a specific period, for example three weeks, the traffic flow simulation module will start to use the traffic flow data and control parameters to simulate the traffic condition at a intersection recurrence, at the same time to optimize the membership function parameter using GA optimization module. It will find the most suitable membership function parameter for the Multi-layer fuzzy controller to update the former parameters. With this circulation, the continuous learning of the fuzzy control membership function parameter will be implemented to adapt to the change of traffic condition at intersection.

Design of genetic algorithm

In the above multi-layer fuzzy control model, the membership function is Trapezoidal membership functions. Each membership function can be fully described by four points of trapezoid P_1、P_2、P_3、P_4, so in the fuzzy controller, the sequence $X=[x_1^a \quad x_2^a \quad \ldots \quad x_n^a]$ which consist of P_1、P_2、P_3、P_4 of each membership function expresses one answer of the controller membership function set, $n=20$.

Coding method. We use real number coding method, and each parameter of the set is defined within initial limited range, that is $x_i^a \in [x_i^-, x_i^+]$, $i=1, 2, \ldots, 20$. Through the following transform, the initial range may be mapped to [0,1], $f_i : [x_i^-, x_i^+] \to [0,1]$, where $f_i(x_i) = (x_i - x_i^-)/(x_i^+ - x_i^-)$

Fitness function. In the multi-layer fuzzy control model, our objective is to minimize the average delay of each vehicle in the intersection, that is to minimize $d(t)$. The fitness function of GA is $f_{normal}(x) = 1/(1+d(t))$.

Selection strategy. We use the roulette wheel selection strategy. The fitness function of each individual is f_i, then the selection probability of a individual in population is $f_i/\sum f_i$ for next generation.

Mutation function. We use the nonconforming mutation method. This method builds a relationship between mutation operator and evolvement generation number, which will make the mutation range larger at the beginning of evolvement. With the development of the evolvement, the mutation range will be reduced.

Assume $s=(v_1,v_2,...v_n)$ is a parent and the variable v_k is selected to be mutated, the definition range is $[a_k,b_k]$, after mutation the child is $s'=(v_1,...,v_{k-1},v_k',...v_n)$, where

$$v_k' = \begin{cases} v_k + \Delta(t,b_k - a_k) & , \ if \quad rnd(2)=0 \\ v_k - \Delta(t,v_k - a_k) & , \ if \quad rnd(2)=1 \end{cases}$$

Here t is the generation number, the range of $\Delta(t,y)$ is $[0,y]$, with t increasing, the probability of $\Delta(t,y)$ close to 0 increases. $\Delta(t,y)$ may be expressed as $\Delta(t,y)=y\bullet(1-r^{(1-t/T)^\lambda})$, here r is a random value between[0,1], T is the maximum of generation, $\lambda\in[2,5]$。

Crossover function. We use whole arithmetic crossover method. Assume that the following two parents, $s_1=(v_1^{(1)},v_2^{(1)},...,v_n^{(1)})$、$s_2=(v_1^{(2)},v_2^{(2)},...,v_n^{(2)})$, and two children $s_z=(z_1,z_2,...,z_n)$、$s_w=(w_1,w_2,...,w_n)$ from the crossover of s_1、s_2. $a_1,a_2,...,a_n$ are n random values between (0,1), then $z_i=a_iv_i^{(1)}+(1-a_i)v_i^{(2)}$, $w_i=a_iv_i^{(2)}+(1-a_i)v_i^{(1)}$, here $i=1,2,...,n$.

Control parameters. Initial population scale is 50, generation number is 100, crossover probability is 0.6, and mutation probability is 0.2.

Simulation

We compare the performance of multi-layer fuzzy control model based on GA to the same multi-layer fuzzy control model without membership function optimization in a simulated four- approach intersection under the same traffic flow condition. The simulation will test the extent of adoption under different degrees of saturation, so we need a longer simulation time about 2 hours. We assume that the

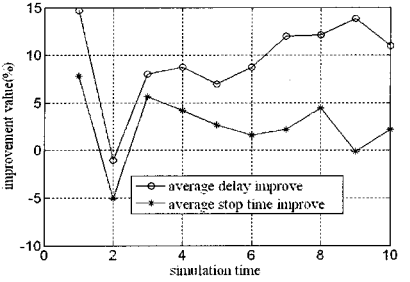


Figure 3. Improvement of average delay and stop time of model based on GA to model without GA under different traffic condition

arrival traffic flow is Poisson distributed, and also binomial distribution is used to simulate the change process of degree of saturation from low to large. The traffic flow ranges are shown as follows, through lane (TL): 350PCU~431PCU, left turn lane (LTL): 280PCU~352PCU, the first condition is, TL: 350PCU, LTL: 280PCU, the second, TL: 359PCU, LTL: 288PCU, the rest may be deduced by analogy. The simulation results are shown as Figure 3. Figure 4 is the membership function after GA optimization and Figure 5 is the membership function set manually during simulation.

From Figure 3 we see that the control model presented in this paper outperforms the traditional model in reducing average total delay and stop time at an intersection. Under ten different simulation initial conditions, there are nine times in which the delay and stop time is reduced using the model presented in the paper. With GA, the membership function parameters can be optimized to adapt to the actual traffic condition and the average delay is reduced with the model based on GA.

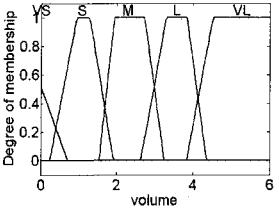


Figure 4. Membership function after GA optimization

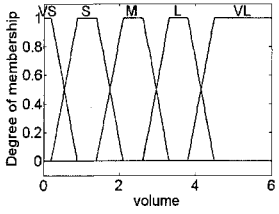


Figure 5. Manually set membership function

Conclusion

The research indicates the application of GA in traffic signal fuzzy control model will have some benefits. In the above fuzzy control model, the membership function can be optimized by GA on-line, but the reasoning rules are all decided by experts and cannot be modified according to the traffic flow change, so it has some limitation in the field applications. In future research, the emphasis will be given to the optimization of the fuzzy reasoning rules on-line, and the GA and enforcement learning theory may be a possible method.

References

Ella Bingham (2001). "Reinforcement learning in neurofuzzy traffic signal control." *European Journal of Operational Research* 131, 232-241.

Mohamed B.Trabia, Mohamed S.Kaseko, Murali Ande (1999). "A two-stage fuzzy logic controller for traffic signals." *Transportation Research Part C* 7, 353-367.

Niittymäki J., Kikuchi S. (1998). "Application of Fuzzy Logic to the Control of a Pedestrian Crossing Signal." *Transportation Research Record* No. 1651. Washington D.C. 30-38.

Niittymäki J. (2001). "Installation and Experiences of Field Testing a Fuzzy Signal Controller." *European Journal of Operational Research,* 131/2, 45-53.

Stephen Chiu. Sujeet Chard (1993). "Self-Organizing Traffic Control via Fuzzy Logic." *Proceedings of the 32nd Conference on Decision and Control,* San Antonic, Texas, December.

Estimating a transit route OD matrix from on-off data through an artificial neural network method

Jie Yu[1] and Xiao-guang Yang[1]

[1]Department of Traffic Engineering, Tongji University, 1239 Siping Rd., Shanghai, 200092 China; PH 86-21-65988372; FAX 86-21-65988372; email: lovelyfish0731@gmail.com

Abstract

A route O-D matrix is the base data for transit line designing and optimizing. This paper presents an OD estimation model, which applies a modified BP artificial neural network (ANN) to track the weights of OD pairs within a transit route using the number of boarding and alighting passengers at each stop. The paper details the key issues associated the estimation process, including constraints for OD estimation models, design of ANN, the learning and stopping rules, and the choosing of model parameters through computational experiment. The results of a rigorous validation with on/off data from a real bus route reveal that the proposed model is quite effective and reliable in estimating the OD matrix for identification of the underlying demand pattern of a transit route.

Introduction

A route O-D matrix is the base data for transit planning and operating, which reflects the concrete flow direction of transit passenger volumes between stops. Conventionally, an accurate route O-D matrix must be obtained through a thorough survey, which is time-consuming and labor-intensive. Besides it is difficult to conclude through a single survey from which a representative pattern of traffic is found. If a consistent pattern does not exist, the O-D table obtained by survey would be of little value. These difficulties arouse researchers' attention on developing efficient and accurate methods to estimate the O-D matrix.

There have been a number of studies proposed to estimate the highway O-D matrix (Bell M.G.H., 1983; Kikuchi et al. 1999; Lin et al. 2004) by using cross-section volume or on/off ramp volume, and to estimate transit route O-D matrix by using on/off passenger volume (Simon et al. 1985; Wong et al. 2005). However, the shortcomings of the above methods are their static property and data-sensitivity. They often require one "representative" dataset of the underneath demand pattern, and hence the dataset, if not well chosen will yield unstable estimation results. So the above models cannot take full advantage of repeated or multiple datasets for better estimation.

Hence, there exists a need to estimate the O-D matrix through a method that not only identifies the underneath pattern but also allows utilizing repeated daily operational datasets without much efforts, such as the number of boarding/alighting passenger of the transit vehicles or the entering/leaving volumes in the highway system.

Based on previous research, this paper intends to present an artificial neural network model to estimate the elements of an O-D matrix for a transit route. The input to the ANN model is the total number of boarding passengers at each stop and the desired output is the total number of passengers alighting at each stop. The model can identify the O-D pattern underlying in a transit route through the training process by repeatedly applying input-and-output data pairs. As long as the stability of the travel behavior is assured and the data required for training are suitable, the method can estimate the O-D matrix with fair accuracy.

O-D Weight Matrix

A transit route with n stops is illustrated in Fig 1, where each node represents the stop. The counts $x_i(i=1...N)$ and $y_j(j=1...N)$, respectively represent the total boarding passengers at stop i and the total alighting passengers at stop j.

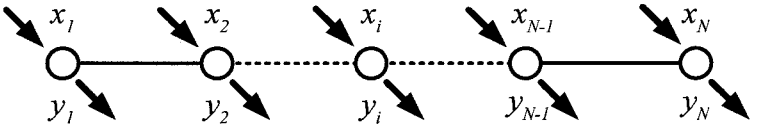


Figure 1. Illustration of a transit route

The value w_{ij} represents the proportion of passengers from i to j (OD_{ij}), to the total number of boarding passengers at stop i (x_i), that is: $w_{ij} = OD_{ij} / x_i$ .Note that three laws of conservation of flow must be kept during analysis: (1) $\sum_j w_{ij} = 1 \ \forall i$, (2) $w_{ij} = 0 \ \forall i = j$,and (3) $\sum_i w_{ij} \cdot x_i = y_j \ \forall j$. The form of O-D weight matrix is shown in Table 1.

Table 1. O-D weight matrix

O \ D	1	2	…	N
1	w_{11}	w_{12}	…	w_{1N}
2	w_{21}	w_{22}		w_{2N}
⋮	⋮			⋮
N	w_{N1}	w_{N2}	…	w_{NN}

Estimation Model

Assumptions

The estimation model by ANN developed in this paper is based on the following assumptions: (1) The boarding/alighting counts at each stop are available, i.e. the counts $x_i(i=1...N)$ and $y_j(j=1...N)$ are known; (2) There does exist an O-D pattern under the datasets provided. Actually, the ANN model verifies the existence of a travel pattern during computation, because finding the weights of an ANN is the process of searching for a pattern of the system behavior. If the pattern does not exist, the ANN model cannot generate an acceptable weight matrix. In other words, the ANN model cannot find a pattern if it does not exist; (3) The integration periods for passenger counts are identical, and the noise of data can be ignored, e.g. no accident is considered.

Formulation

The O-D estimation model can be formulated as follows:

$$Min: \sum_k \sum_j \left| y_j^k - \sum_i w_{ij} \cdot x_i^k \right| \tag{1}$$

s.t.

$$\sum_j w_{ij} = 1, \ \forall i \tag{2}$$

$$0 \le w_{ij} \le 1, \forall i, j \tag{3}$$

$$w_{ij} = 0, \ \forall i = j \tag{4}$$

Where: w_{ij} are the decision variables and k is the index of data set. The objective of the model is to find a set of w_{ij} which can fit all the datasets with minimal errors between desired and actual outputs.

ANN Design

This section is to develop an ANN model that can identify the underlying pattern for the system through the training process by repeatedly applying input-and-output data pairs. Here, the boarding/alighting number of passengers is the input of the model, and the output is the O-D matrix weights. The training idea is that for all the datasets, the number of boarding passengers from each stop is multiplied by the weight matrix and compared with the desired output. If the sum of error is under the tolerance level, then a set of suitable weights is found. Otherwise, an improvement direction of each weight is calculated and this adjustment can be added to the current weight. The updated weights are applied to the inputs for the next iteration and the procedures continue sequentially until the stopping criteria are reached. Here below are some notations used in the training model:

(1) The input vector is: $X^k = [x_1^k, x_2^k, ..., x_N^k]^T, \forall k$;

(2) The desired output vector is: $Y^k = [y_1^k, y_2^k, ..., y_N^k]^T, \forall k$;

(3) The weight matrix corresponding to Table 1 is: $W(t) = \begin{bmatrix} 0 & w_{12} & \cdots & w_{1N} \\ w_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \\ w_{N1} & \cdots & & 0 \end{bmatrix}$;

(4) The estimated output after linear transformed by W(t) is: $Z(t) = W^T X = [z_1, z_2, ..., z_N]^T$, which satisfies the OD conservation laws, where t is the index of training iterations.

Learning Rule

Among various learning rules, the Delta Rule, also referred as the Least Mean Square (LMS) Learning Rule is one of the most commonly used and is based on the simple idea of continuously modifying the strengths of the input connections to reduce the difference (the delta) between the desired output value and the actual output of a processing element.

Let F(W) represent the objective function of the Delta Rule, i.e. $F(W) = (Y - Z)^T (Y - Z)$. To minimize the objective, the solution has to move along the direction of improvement. The decent direction of this problem is:

$$-\nabla F = -\frac{\partial F}{\partial w_{ij}} = 2(y_j - z_j)\frac{\partial z_j}{\partial w_{ij}} = 2(y_j - z_j)x_i \tag{5}$$

These adjustments are then propagated back through the system, causing the system to adjust the weights which control the network. The incremental adjustment of each element in the weight matrix is then defined as:

$$\Delta w_{ij} = \frac{1}{2}\varepsilon\left(-\frac{\partial F}{\partial w_{ij}}\right) = \varepsilon(y_j - z_j)x_i \tag{6}$$

Where, ε is the learning coefficient, which assumed to be a constant value between 0 and 1. It should be noted that the value of ε must be selected considering the trade-offs between accuracy and computation time. Here, the ε value is set at 10^{-8} based on our computational experiments during this study. Therefore, at the (t+1) iteration, we have $w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t)$. However, according to the attributes of an O-D matrix, the weights should be subject to the following constraints: (1) $0 \le w_{ij} \le 1$ and (2) $\sum_j w_{ij} = 1 \ \forall i$. The above Delta Rule will not assure to satisfy these constraints. Hence, we modify the learning rule as follows:

$$w_{ij}(t+1) = \begin{cases} w_{ij}(t) - \alpha \cdot \Delta w'_{ij}(t), & \text{if } w_{ij}(t) + \Delta w'_{ij}(t) < 0 \\ w_{ij}(t) + \Delta w'_{ij}(t), & \text{if } w_{ij}(t) + \Delta w'_{ij}(t) \ge 0 \end{cases} \tag{7}$$

Where $\alpha = \exp(\Delta w'_{ij}(t))$, and

$$\Delta w'_{ij}(t) = \frac{1}{2}\varepsilon\left(-\frac{\partial F}{\partial w_{ij}(t)}\right) + \frac{1-\sum_j w_{ij}(t)}{(N-1)} = \varepsilon[y_j - z_j(t)]x_i + \frac{1-\sum_j w_{ij}(t)}{(N-1)} \tag{8}$$

By Eq. (7) and (8), after each iteration, the new adjusted weights will surely satisfy the above weight constraints.

Stopping Criteria

1. Tolerance level of error

According to the training mechanism, the values of the weight matrix in the last iteration should fit in all the training datasets with tolerable estimation error. That is:

$$\sum_k \sum_j \left| y_j^k - \sum_i w_{ij} \cdot x_i^k \right| \le e \tag{9}$$

The value of e is important because it determines the time the training process will take. Therefore, it should be set by computational experiments.

2. Maximum iteration limitation

As is mentioned before, the convergence value e cannot be pre-determined. A direct method to get it is to run the model under different training iterations and set e at the value where no improvement occurs with the increase of training iterations.

Case Study

Briefing Introduction

A comprehensive set of daily volume counts from Route 41 in Suzhou City, China is available for case study. It contains 6 sets of daily counts of on/off passengers for 28 stops. The route is shown in Figure 2. The on/off volume of two directions is summed together for the input and desired output of proposed ANN model. One thing should be mentioned for the case study. For the ANN model, a large amount of data is needed for training, however, this case study contains only 6 sets of daily counts for the training, which seems not enough. The data limitation problem is due to the survey noise and data filtering procedure, which

eliminate many sets of unqualified data. With more datasets available, this issue will be addressed in future studies.

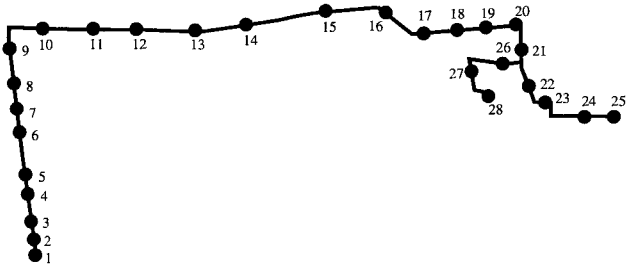


Figure 2. Route 41 in Suzhou City for case study

Estimation Results of the Model

As stated above, different e values will significantly affect the results of the estimation model. Since e cannot be pre-determined, we run the model under different training iterations and the estimation results are shown in Table 2.

Table 2. Results of the model under different training iterations

Iterations	10^6	$2*10^5$	10^5	50000	10000	5000	1000	500	100	10
Estimation Error(trips)*	184	184	187	187	418	932	3119	3664	15193	16661

[1] The estimation error is the cumulative error between the desired and actual outputs computed by $\sum_k \sum_j \left| y_j^k - \sum_i w_{ij} \cdot x_i^k \right|$.

It should be noted that when the training iterations are exceeding 50000, there is little space for the estimation results to be improved. In other words, if we set a convergence threshold $e < 184$, the training process will take a huge amount of time to converge, or even can't converge. Based on the above analysis, we set the value of e at 187 and training iteration limit at 10^5.

Estimated O-D & Surveyed O-D

Given 6 sets of daily volume counts, the proposed ANN is applied to find the weights of OD matrix. After the weights have been found, a real OD data from the same survey period is used to examine the validity of the model. As is shown in Figure 3, a 45-degree line is drawn as a reference line to compare the estimated and the actual O-D values. It is seen that the points are generally distributed along the 45-degree line. As is shown in figure, most of the estimated values are less than 400 trips per day. The maximum value of difference between the estimated and the surveyed is 111 trips per day and the majority of the differences are within 30 trips per day. The RMSE for the estimation is 25.2 trips per day.

Conclusion

This study applies a neural network method to estimate transit route OD matrix using on/off data. The result shows that the proposed model can effectively estimate the elements in the OD matrix without loss of much accuracy. The proposed model in this paper can be applied further in the following aspects. First, if enough datasets are available and the transit route is not often subject to changes, the model can be used to estimate the stable O-D pattern for the

route. Second, the model is also suitable for checking and updating the O-D pattern to see the effects of route or service changes. Finally, the model can also be extended to estimate the O-D matrix for a transit network.

Estimation accuracy of the proposed model could be improved in the further research through adding modification of the historical adjustable steps in the learning rule, because the current learning rule can result in local optimal rather than global optimal for the weights estimation.

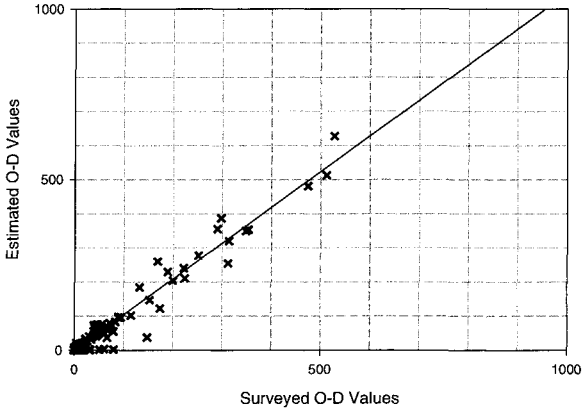


Figure 3. A comparison between surveyed and estimated O-D values

References

Bell. M.G.H. (1983). "The Estimation of an Origin-Destination Matrix from Traffic Counts." *Transportation Science*, 17(2), 198-217

Kikuchi, S. and M. Tanaka. (1999). "Estimating an O-D table under repeated counts of in-out volumes at highway ramps: Use of Artificial Neural Networks." *The 79th TRB Annual Meeting. Washington*, D.C., 59-66

Lin, P. W. and Chang, G. L. (2004). "Robust estimation of the dynamic origin-destination matrix for a freeway network." *The 2004 IEEE International Conference on Networking, Sensing and Control.* Taipei, Taiwan, 862-867

Simon, J. and P.G. Furth. (1985). "Generating a Bus Route O-D Matrix from On-Off Data." *Journal of Transportation Engineering*, ASCE, 111(6), 583-593

Wong K.I., Wong S.C., Tong C.O., Lam W.H.K., Lo H.K., Yang H. and Lo H.P. (2005) "Estimation of origin-destination matrices for a multimodal public transit network." *Journal of Advanced Transportation*, 39, 139-168.

Tests of Simulation Models *FREQ, KRONOS, Integration* and *VISSIM* in Replicating Congested Freeway Flow

Panos D. Prevedouros[1] PhD, James Watson[2] MSCE, and Jerry Ji[1] PhD

[1,3] Department of Civil and Environmental Engineering, University of Hawaii at Manoa; Holmes Hall 383, 2540 Dole Street, Honolulu, HI 96822; Fax: (808) 956-5014; E-mail: pdp@hawaii.edu, xji@hawaii.edu.
[2] Korve Engineering, 155 Grand Avenue, Ste. 400, Oakland, CA 94612, jwatson@korve.com

ABSTRACT

The results from the application of two macroscopic (*FREQ, KRONOS*) and two microscopic (*Integration, VISSIM*) traffic simulation models on the same freeway network are presented. The data were collected contemporaneously on a heavily congested 15-mile section of H-1 Freeway in Honolulu, HI. The models were calibrated so that field-collected speeds at five cross sections were replicated adequately. The models were applied to evaluate a $50 million freeway widening project which was constructed between 2004 and 2006. Model predictions on the savings of the widening varied widely.

BACKGROUND AND OBJECTIVES

The 15-mile section of the westbound H-1 Freeway in Honolulu, Hawaii from the airport to Waikele was used for testing four traffic simulation software in simulating complex and congested conditions. This freeway section is typically four lanes wide, but at one point it reaches a width of eight lanes. The network consists of seven on-ramps and eight off-ramps, including two 2-lane on-ramps, two 2-lane off ramps, two freeway merges (Moanalua Freeway and H-3 Freeway), and one freeway diverge (H-2 Freeway). The westbound H-1 Freeway carries its peak load in the evening peak period, so all simulations were done for the period between 3 PM and 7 PM.

The objectives of this research were to: (**1**) Develop two macroscopic (*FREQ* 12 and *KRONOS* 9) and two microscopic (*Integration* 2.30 and *VISSIM* 4.0 and 4.1) traffic simulation model applications using contemporaneous freeway data. (**2**) Assess speed and travel time model estimates and compare them to each other and to field data. Test combinations of parameter settings that lead to a reasonable replication of speeds at several cross sections. Models should also reflect the propagation and relief of congestion over space and time. (**3**) Apply all four models to a future scenario which is currently under construction in order to receive multiple "second opinions" of a major freeway widening project. (**4**) Begin an investigation of the effect of origin-destination (O-D) estimates from various applications to the output of freeway simulation.

This research was designed to benefit not only the freeway widening project of the Hawaii State DOT by assessing its expected benefits and potential drawbacks such as the generation of downstream bottlenecks, but also the nation's state-of-the-art in freeway simulation by evaluating four traffic simulators. The implementation of the Waimalu widening project presents an extraordinary opportunity to reveal which model produced results closest to reality, when post-implementation volume and speed data become available in 2007.

MODEL TESTS

Three main tests were conducted and results were evaluated: (**1**) Ability to replicate observed speeds at various locations and propagate congestion shockwaves as observed in the field. (**2**) Comparisons of model estimates for a future scenario, in this case a bottleneck removal by lane widening. (**3**) Impacts of origin-destination (O×D) estimations with multiple methods on simulation results.

Representation of Existing Conditions

Comparisons must be made between the real and the simulated freeway system to determine the accuracy of a simulation which is designed to replicate an existing system. This task may be accomplished by visual inspection and statistical comparison. Both *Integration* and *VISSIM* offer a real-time visual component with vehicle speeds represented by colors, therefore the user may be able to observe the propagation of queues and congested areas based on the color and density of vehicles. *FREQ* has static color coded output of density over time and space, and *KRONOS* has a time scanning emulation of speeds over space. If a prior knowledge of the vehicle speeds and congested locations is available, then the animation of micro-simulators or congestion displays of macro-simulators may aid in calibration.

Statistical comparisons between field and simulated data can be conducted by gathering system data with virtual detectors placed at network locations which correspond to vehicle detection stations (VDS) in the field. These comparisons are critical in model parameter calibrations and for determining the overall accuracy of the base model.

The models were calibrated so that field-collected speeds at six locations were replicated adequately for the base case representing February 2002 volume and speed conditions. At each location, speeds from all sources available were compared and goodness of fit was assessed. Field data were collected using three different methods: Surveillance video tapes analyzed with *Autoscope*, moving observer (MO) recordings, and VDS data. *Autoscope* and VDS provided cross-sectional volume and speed data. MO provided travel time data for mainline segments and spot speeds at specific cross-sections. Figure 1 presents speed profiles at four cross sections. The left column of graphs reflect existing conditions: All models after calibration are in agreement with field measured speeds for three locations upstream a major bottleneck. However, *FREQ* and *VISSIM* are sensitive to a major downstream diverge which is reality is only a minor bottleneck with speeds constantly around 50 mph.

The calibration of the macroscopic models was much easier and quicker than that for the microscopic models. Basically under no-to-moderate congestion *FREQ* produces results similar to *KRONOS* when its mainline capacity is set 200 vph below *KRONOS'* setting. *Integration* was the most difficult to calibrate for this application. *VISSIM's* calibration was less tedious, but it requires runs with multiple seeds. The results reflect averages from five seeds.

Future Scenario Estimates

Without any modification in parameters or calibrations, a 5,000 ft. 5-lane section with 6 lanes upstream and downstream of it (e.g., bottleneck) was modified to 6 lanes (e.g., bottleneck removal). This is the roughly $50 million "Waimalu Widening" project of the Hawaii State DOT which was completed in spring 2006. Model predictions on the savings of the widening under construction vary considerably, as shown in Table 1.

The results indicate that the widening should be expected to produce considerable travel time savings ranging from at least 10% and as high as nearly 40%. In past applications of simulation and subsequent verifications with real time deployment both *KRONOS* and *Integration* overestimated freeway savings and *Integration* was about right, and in agreement with *TRANSYT-7F*, in arterial travel time savings (Prevedouros, 2001). Therefore the conclusion delivered to HDOT was that savings of 10% or more should be expected and that

Table 1: Summary Results of Base and Freeway Widening Scenarios with Four Models

	BASE CASE	WIDENING 5→6 LANES	TRAVEL TIME SAVINGS		Simulation Run Time
			Change in Vehicle-Hours	% Change	
FREQ 12	8622	6783	1839	-21%	<2 sec.
KRONOS 9	9814	7446	2369	-24%	23 sec.
INTEGRATION 2.30	9442	5994	3449	-37%	57 min.
VISSIM 4.0 (4/2003)	10945	9713	1232	-11%	110 min.
VISSIM 4.1 (5/2005)	11838	10353	1485	-13%	n.a.

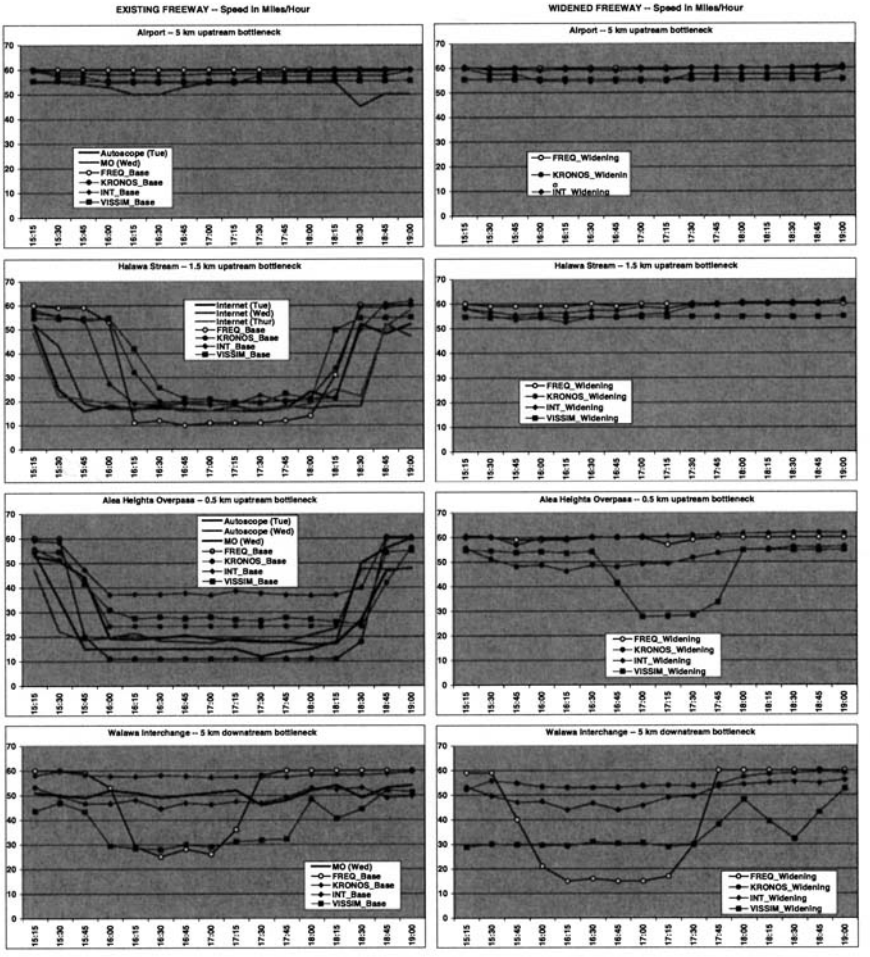


Figure 1. Actual and simulated speed profiles at selected cross sections.

they can be even higher if downstream secondary bottlenecks are relieved. These bottlenecks will worsen as more traffic is relieved from the widening. (Note: *VISSIM* 4.1 was run on a different desktop computer; therefore, its run times are not comparable to the others.)

Implications of O×D Estimates on Output

In order to better represent the motorists in a simulated traffic system, several microsimulation models depend on origin-destination (O×D) data for lane choice, lane changing decisions, and route keeping on the network (Gomes et al. 2003). Theoretically, the most accurate way to determine a network O×D distribution is to track the starting and ending points of all vehicle trips which can be done with a comprehensive license plate survey. Such surveys are too expensive and time consuming for most projects (Chen et al. 2005). Good quality but often dated O×D data may be available from large periodic regional studies conducted by Metropolitan Planning Organizations. The number of possible O×D distribution variations for a network is enormous (Hazelton 2003). On the other hand, the most common demand data available are interval volume counts at various locations on a network such as screen lines, intersection turning amounts and expressway ramp counts.

Based on the data available (detailed ramp and mainline volumes but no origin-destination trips) O×D data from traffic counts had to be developed for the application of *Integration* and *VISSIM* (Willis and May 1981). The variations in distributing the vehicles within a network likely produce different simulation results (Hazelton 2003, Davis et al. 2004). Therefore, the goal for this part of our research was to analyze the implications and effectiveness of different O×D distribution methods in freeway simulation.

O×D distribution matrices were generated using five distribution models which were applied to the same set of input volume data: Deterministic, Proportionate, *QueensOD* with default seed, *QueensOD* with custom seed (Van Aerde 2002), and *FREQ* O×D output.

A custom program, *WatsonOD*, was developed to expedite the generation of O×D matrices (Watson 2004). The program is a MS Excel module which reads tables of traffic counts and produces O×D matrices in three formats: Plain and preformatted as *Integration* and *VISSIM* interval demand data files. The *WatsonOD* module was developed in Visual Basic for Applications (VBA) programming language.

Integration was chosen for testing the generated O×D distributions from the five aforementioned methods. The choice was made because *Integration* requires O×D inputs, it is fairly easy to use, it has been used successfully on similar networks in Honolulu thereby reducing parameter calibration time (Prevedouros and Wang 1999), and it does not use random seeds which would necessitate multiple runs.

The simulated total travel time (TTT) is a network-wide performance outcome. TTT results are summarized in Figure 2. TTT 15-minute profiles for all simulations with *Integration* produced a concave shape, similar to those produced by macroscopic simulation models *FREQ* and *KRONOS*. The TTT from the deterministic O×D method was the most removed from the other four O×D methods.

The regression slope coefficient for each of the ten paired comparisons in Figure 2 shows that TTT for the deterministic method compares poorly with the rest, whereas the other four O×D methods produce similar results in comparison. Every TTT combination of the four is within a 10% range of the ideal 1.000 correlation coefficient. The proportionate and *FREQ* output methods produce a slope of 1.009. Pearson's correlation outcomes are similar to the regression slope outcomes.

The t-test indicates that in general, the null hypothesis that the difference in TTT between two methods is significant is rejected at the 95% confidence level for comparisons among all O×D methods excluding the deterministic method. The TTT results from the de-

terministic O×D method are not similar to any of the other methods. Techniques based on proportional assignment will likely produce flawed results in the presence of congestion when multiple route choices are available Chen et al. (2005). This threat, however, is not applicable to the linear freeway systems examined herein.

The conclusion of this investigation is that depending on the nature of the analysis the user may not be required to determine and apply the most accurate O×D distribution, which in the absence of detailed contemporary license plate survey data cannot be determined. For four of the five O×D methods the simulation results were adequate for common simulation analyses of freeways.

Travel Time Regression	Deterministic	Proportionate	QOD Default	QOD Custom
Proportionate	0.581			
QOD Default	0.465	0.940		
QOD Custom	0.514	1.038	1.085	
FREQ Output	0.546	1.009	1.022	0.944

Travel Time Correlation	Deterministic	Proportionate	QOD Default	QOD Custom
Proportionate	0.792			
QOD Default	0.647	0.961		
QOD Custom	0.655	0.971	0.993	
FREQ Output	0.730	0.992	0.982	0.992

t-Test	Deterministic	Proportionate	QOD Default	QOD Custom
Proportionate	0.001			
QOD Default	0.000	0.257		
QOD Custom	0.000	0.571	0.611	
FREQ Output	0.000	0.719	0.457	0.830

Figure 2. Statistical analysis of total travel time.

CONCLUSIONS

Due to the size of the network and the heavy volume of traffic on it, all simulations with microscopic models took between 60 and 110 minutes to execute. The tests revealed a number of issues which are listed below.

Representation of geometry at complex, unusual merges: Most software cannot represent complex or substandard physical features precisely. Various tricks and auxiliary files or settings were needed.

Representation of weaving (driver behavior) at complex, unusual merges: The level of detail varies widely among software and most require extensive trial-and-error runs in order to approximate the prevailing driver behavior. Unrealistic capacity and/or density settings may be needed at some heavy merges to replicate reality. In turn, these may not work well when conditions are not congested.

Production of O×D matrices from mainline and ramp station counts: There are several alternative methods which may produce widely different results. However, the proportionate and *FREQ* output methods produce O×D data that are reliable for complex freeway networks. This conclusion may not hold for corridor/urban networks.

Model calibration with field-observed speeds and queue lengths: These are among the easiest measures for inspecting simulation output and making necessary calibrations. Microscopic models with virtual detector stations can produce more appropriate speeds samples for comparisons with field values, as opposed to macroscopic models that produce segment averages.

Data quality and day-to-day variability: These are perennial issues which require substantial resources in order to minimize data errors and conduct multi-day runs to account for daily, seasonal and other forms of variability. This research did not address issues of temporal variability.

Duration of execution with microsimulation models for long periods and nearly 100,000 vehicles: Existing desktop PC technology does not provide short enough execution speeds to make real-time scenario analysis possible (e.g., for incident modeling at a traffic control center.) In addition, large networks take a long time to calibrate properly because global parameter settings are insufficient and they usually need to be modified at odd or complex segments.

REFERENCES

Chen, A., P. Chootinan and W. W. Recker (2005) "Examining the Quality of Synthetic Origin-Destination Trip Table Estimated by Path Flow Estimation." *Journal of Transportation Engineering*, Volume 131, Number 7, ASCE, 506-513.

Davis, G. A., D. M. Levinson, M. Davis, P.G. Michalopoulos, and S. Muthuswamy (2004) "Freeway Origin Distribution Matrices, Not as Simple as They Seem." University of Minnesota. Minneapolis, MN, www.ce.umn.edu/~levinson/Papers/ODESSA.pdf.

Gomes, G., A. D. May, and R. Horowitz (2003) "A Microsimulation Model of a Congested Freeway Using *VISSIM*." 2004 Annual Meeting of TRB, Washington, D.C.

Hazelton, M. L. (2003) "Some Comments on Origin-Destination Matrix Estimation" *Trans. Research* Part A: Policy and Practice, Volume 37, Issue 10, 811-822.

Prevedouros, P. and Y. Wang (1999) "Simulation of a Large Freeway/Arterial Network with *Integration*, *TSIS/CORSIM* and *WATSim*." *Transportation Research Record*, No. 1678, 197-207, TRB, Washington, D.C.

Prevedouros, P. (2001) "Freeway Ramp Closure: Simulation, Experimentation, Evaluation and Preparation for Deployment." *ITE Journal*, Vol. 71, No. 6, 40-44.

Van Aerde, M. & Assoc., Ltd. (2002) "*QueensOD* 2.10 – User's Guide: Estimating Origin-Destination Traffic Demands and Link Flow Counts." Kingston, Ontario, Canada.

Watson, J. R. (2004) "Implications of Origin-Destination Distributions in Freeway Simulation." Master's Thesis. Department of Civil and Environmental Engineering, University of Hawaii at Manoa.

Willis, A. E., and A. D. May (1981) "Deriving Origin-Destination Information from Routinely Collected Traffic Counts: Volume 1 – Trip Table Synthesis for Single Path Networks." Report No. UCB-ITS-RR-81-8. Institute of Transportation Studies, U.C.-Berkeley.

Comparison of PASSER V, Synchro, and TRANSYT-7F for Arterial Signal Timing based on CORSIM Simulation

Yunlong Zhang[1] and Yuanchang Xie[2]

[1]Zachry Department of Civil Engineering, Texas A&M University, 3136 TAMU, College Station, TX 77843; PH (979) 845-9902; FAX (979) 845-6481; email:yzhang@civil.tamu.edu

[2]Zachry Department of Civil Engineering, Texas A&M University, 3136 TAMU, College Station, TX 77843; PH (979) 862-3553; FAX (979) 845-6481; email: yx40@tamu.edu

Abstract

Among the many software packages for arterial traffic signal timing, PASSER, Synchro, and TRANSYT-7F are the three most popular ones in the United States. Each of them has its own pros and cons, and it is very important to choose the appropriate software for a traffic study in order to obtain the most suitable signal timing plan. Although several studies have been conducted in the past to compare the performance of these three software packages, a lot of logic changes and new features have been implemented recently and an up-to-date comparison based on the latest versions of these software packages is necessary.

In this paper, we use PASSER V, Synchro, and TRANSYT-7F to develop arterial signal timing plans independently for a four-intersection arterial segment in College Station, TX. Real-world traffic volume data during morning, noon, and afternoon peak hours are used to generate the timing plans using all three packages. The performances of those timing plans are evaluated using CORSIM simulation. Delay, stops, and speed are compared based on the outputs of CORSIM simulation runs. Based on the comparison results under various traffic conditions, an assessment of performance and suitability is provided with recommendations on selection criteria.

Keywords: Traffic signal coordination, Traffic simulation, Signal timing plan.

Introduction

When intersections are closely spaced such that vehicles from the upstream intersections arrive in platoons along the major movement direction, traffic signal coordination should be considered to facilitate vehicles to move through a set of intersections without stopping (Roess, 2004). Many software packages have been developed to help design an optimal arterial signal timing plan, and the most popular packages are PASSER II, Synchro, and TRANSYT-7F. Skabardonis and May (Skabardonis, 1985) compared the arterial timing plans produced by TRANSYT-7F, PASSER II, and MAXBAND. They then used NETSIM to evaluate those timing

plans and found that TRANSYT-7F produced the best timing plan. Wong (1991) compared TRANSYT-7F with PASSER II using field data collected from San Francisco. He concluded that in respect to the whole arterial system including cross streets, TRANSYT-7F was better than PASSER II; If only the arterial street was considered, PASSER II was better than TRANSYT-7F; the field results showed that the performances of TRANSYT-7F and PASSER were approximately the same in terms of travel time and stops along the arterial street. Yang (2001) compared the timing plans optimized by TRANSYT-7F, PASSER II, and Synchro using CORSIM simulation. Five MOEs produced by CORSIM: delay, stops, speed, fuel consumption, and emission were used to evaluate the best timing plans given by each of the three packages. Based on the simulation results, Yang concluded that TRANSYT-7F always gave longer cycle lengths; PASSER II seemed to work better for intersections with unbalanced volumes; and CORSIM was a very good tool to evaluate timing plans. Chaudhary and Chu (2003) did a research to compare Synchro, TRANSYT-7F, and PASSER II. They concluded that Synchro and PASSER II were better than TRANSYT-7F for progression bands optimization. Further investigation of Synchro and PASSER II showed that with an increased arterial size, PASSER II would provide better two-way arterial progression than Synchro.

The purpose of this study is to compare the arterial timing plans optimized by the latest versions of PASSER V, Synchro, and TRANSYT-7F for different traffic demand scenarios and to provide useful guidelines for traffic engineers to choose the proper traffic signal analysis packages. The organization of the rest of this paper is as follows: first a brief summary of the three packages and their new features are given. After that is the test design and test results analysis. Finally, suggestions for software selection and conclusions are presented.

Software Overview

PASSER II and V

PASSER II uses a macroscopic deterministic traffic flow model. It maximizes the arterial progression bandwidth efficiency by varying the cycle lengths, offsets, phasing sequences and splits. PASSER II is one of the most popular arterial signal software. However, it has two major limitations (TTI, 2004). First, PASSER II tends to produce larger optimal cycle length compared with other programs. The reason is that with larger cycle length the split optimization algorithm will assign more green time to the through traffic and consequently larger arterial progression bandwidth efficiency may be obtained. Since maximizing bandwidth efficiency is the primary objective of PASSER II, it tends to choose larger cycle lengths. Another limitation of PASSER II is that although it is good at maximizing the arterial progression bandwidth efficiency, it may produce timing plans that cause more delay to the arterial system and especially to the cross street vehicles compared to other programs such as TRANSYT-7F (Wong, 1991).

PASSER V-03 is the latest version of PASSER family. It corrected many limitations of PASSER II as described before. In addition to optimizing timing plans for maximum arterial progression bandwidth efficiency, PASSER V-03 provides some new features as follows (Chaudhary, 2002):

1. Arterial timing plans that minimize system delay;
2. Bandwidth constrained arterial timing plans that minimize system delay;
3. A new mesoscopic delay model that can be used to model queue spillback or blocking; and
4. A new optimizer based on a genetic algorithm.

Since significant changes have been made in PASSERV-03 compared to PASSER II, in this study, we will use PASSER V-03 instead of PASSER II.

Synchro

Among the three programs of PASSER, Synchro, and TRANSYT-7F, Synchro is the most user-friendly one. Synchro (Husch, 2004) provides various tools such as map window, lane window, volume window, and quick editor that make the network coding very simple. In addition, Synchro can transfer the coded traffic network into formats that can be recognized by CORSIM and HCS, thus some researchers use Synchro to code the initial networks and export them to CORSIM or HCS formats for further analysis.

Synchro provides cycle lengths, offsets, phasing sequences, and splits optimizations. The objective of the optimization is to minimize delay and stops. It can be used for single isolated intersections, arterials, and networks. For arterial signal optimization, Synchro starts from optimizing individual intersections' splits and cycle lengths. It then optimizes system cycle length. Finally it determines optimal offsets and phasing sequence.

The latest Synchro version has one very important new feature called queue interactions. The new feature is designed to analyze the capacity reduction and queue delay due to spillback, starvation, and storage blocking, which are common in a saturated urban surface traffic network. Since Synchro optimizes arterial signal timing based on minimizing delay and stops, this new feature in Synchro 6 is expected to improve its arterial signal timing results, particularly when saturation is an issue.

TRANSYT-7F

TRANSYT-7F (Hale, 2005) is primarily used for arterial and network signal optimization, and it can also be used to analyze stop-controlled and yield-controlled intersections. TRANSYT-7F uses a macroscopic traffic model. It provides two simulation options: link-wise simulation and step-wise simulation. The benefit of using link-wise simulation is its computation efficiency. But unlike the step-wise simulation, it can not effectively analyze queue spillback and platoon dispersion. On the other hand, the drawback of step-wise simulation is that it is computationally less effective. Link-wise simulation can achieve reasonable results only under unsaturated conditions; while step-wise simulation, which is used in this study, is suited for both unsaturated and saturated conditions.

Initially TRANSYT-7F provides only a hill-climbing optimization algorithm for optimizing cycle lengths, splits and offsets. Since version 9.4, a genetic algorithm is added to TRANSYT-7F for the optimization of phasing sequence in addition to the other three timing variables. The most recently three versions of TRANSYT-7F are versions 9.7, 10.1, and 10.2. Since there is no significant model or algorithm difference in the latest three versions, version 9.7 is used in this study.

Data Collection and Test Design

Real world traffic data of four intersections along the Harvey Mitchell Parkway in College Station, Texas were collected for this study. Figure 1 shows the four intersections and the layout of the traffic network.

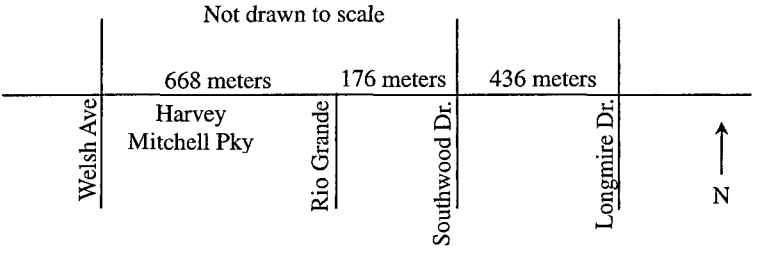


Figure 1. Layout of the study site.

The data collection was carried out on a weekday and hourly traffic volumes were collected during the following three periods: 7:15 am – 8:45 am, 11:45 am – 1:15 pm, and 4:30 pm – 6:00 pm, representing morning, noon, and afternoon peak hour traffics respectively. Three programs, PASSER V-03, Synchro 6, and TRANSYT-7F 9.7, are used to generate arterial signal timing plans for each of those input data sets, and the performance of the timing plans are then compared in terms of three commonly-used MOEs: delay, speed, and stops. Since these software packages use different methods to calculate MOEs, it is possible that they may produce different measurement outputs even for the exactly same timing plan. In view of this, CORSIM is used as a common platform to evaluate the timing plans. For each timing plan, CORSIM simulation is run for 10 times to obtain the mean values of each MOE. As TRANSYT-7F need an initial timing plan input, in this study the optimized timing plans from PASSER V are used as the initial timing plans in TRANSYT-7F.

Test Results Analysis

Table 1 lists the simulated network-wide and arterial MOEs for all the timing plans. It is worth mentioning here that the network-wide MOEs are direct CORSIM outputs and the arterial MOEs are weighted averages of the CORSIM outputs for each arterial link. Two criteria are used to choose the best timing plans in PASSER V. One is bandwidth efficiency and the other is average control delay. For all three traffic demand scenarios, the genetic algorithm in PASSER V does not provide better timing plans than the re-engineered PASSER II's bandwidth efficiency optimization algorithm. TRANSYT-7F provides many optimization objectives. The one used in this study is to minimize the operation cost and all three best timing plans are generated by genetic algorithm. Unlike PASSER V and TRANSYT-7F, Synchro does not need any input for optimization objectives.

The values in Table 1 suggest the following.

1. In general, PASSER V tends to produce larger cycle lengths than TRANSYT-7F and Synchro;

2. Synchro and PASSER V (Delay) produced better timing plans than TRANSYT-7F for all three traffic demand scenarios considered in this study in terms of all network-wide and arterial MOEs;
3. For morning period, PASSER V (Delay) produced better timing plan than Synchro. While for noon and afternoon periods, Synchro produced better timing plans than PASSER V (Delay); and
4. Comparison between the two sets of timing plans produced by PASSER V suggests that timing plans of optimal bandwidth efficiency do not provide better performance in terms of other criteria, delay in particular. For the timing plans with optimal bandwidth efficiency, although the larger cycle lengths for noon and afternoon periods produced better arterial progression and lower percentages of stops, they generated higher average control delay and lower average speed for the network. Surprisingly, even arterial speeds were lower when bandwidth efficiency was optimal. The results seem to suggest that when using PASSER V for signal timing, the average delay criterion should be used.

Table 1. Average MOEs from CORSIM simulation.

Software	Time Period	Cycle Length (sec.)	Network-Wide MOEs			Arterial-Wide MOEs		
			Delay (sec. /veh.)	Stops (%)	Ave. Speed (km /hour)	Delay (sec. /veh.)	Stops (%)	Ave. Speed (km /hour)
PASSER V (Bandwidth Efficiency)	Morning	75	32.8	112.3	40.2	11.2	35.0	41.1
	Noon	115	39.7	104.5	38.4	13.6	33.0	39.3
	Afternoon	120	46.0	108.7	35.6	15.7	34.9	36.6
PASSER V (Delay)	Morning	80	31.7	103.9	40.6	10.1	30.4	42.4
	Noon	65	26.3	109.3	43.5	9.5	35.0	43.8
	Afternoon	75	34.0	115.1	39.5	12.3	37.9	39.1
Synchro 6	Morning	80	35.1	111.8	39.6	11.9	34.6	40.6
	Noon	60	23.7	108.6	44.6	8.3	34.0	45.3
	Afternoon	65	31.0	113.0	40.5	10.1	35.4	42.0
TRANSYT 9.7	Morning	65	35.9	121.3	39.0	13.6	39.6	39.7
	Noon	60	26.5	117.2	43.5	9.8	39.4	43.2
	Afternoon	70	34.3	121.8	39.3	12.7	41.2	38.9

In addition to the results presented in Table 1, we also have the following findings and comments based on our understanding of the packages and experience from this study.

1. Synchro is a very user friendly software, and its performance is comparable with PASSER V and maybe better than TRANSYT-7F in many arterial cases;
2. Compared with traditional optimization methods such as Hill Climb, Genetic algorithms require significantly more computation time, and sometimes they may generate worse timing plans than the traditional algorithms. By

increasing the population size and number of generations, the performance of the genetic algorithm optimization will generally be improved, but this may come with the cost of considerably more computation resources;

3. When using PASSER V for arterial signal timing, bandwidth efficiency should be used cautiously as the criterion to select timing plans;
4. Although TRANSYT-7F is one of the most powerful signal optimization tools, it does not provide an easy interface compared to Synchro or PASSER V. The input and editing of traffic data is not straightforward and prone to errors; and
5. Arterial signal timing is a very complicated problem that is affected by many factors such as traffic volume, turning movement, geometric design, and link length. The findings in this study may not be true if conditions change.

Conclusions

This study compares the performance of optimized arterial signal timing plans from the latest versions of PASSER, Synchro and TRANSYT-7F. Optimal timing plans are generated for a four-intersection arterial network using real-world data. The evaluation and comparison of timing plans are based on CORSIM simulation. The study results suggest that Synchro and PASSER V performed better than TRANSYT-7F in terms of the simulated network-wide and arterial MOEs, and the performance of Synchro and PASSER V are comparable.

References

Chaudhary, N.A., Kovvali, V.G., and Alam, S.M.M. (2002). "Guidelines for selecting signal timing software." *Project report 0-4020-P2*, Texas Transportation Institute, College Station, TX.

Chaudhary, N.A. and Chu, C.L. (2003). "New PASSER program for timing signalized arterials." *Project summary report 4020-S*, Texas Transportation Institute, College Station, TX.

Husch, D. and Albeck John. (2004). *Synchro 6 user guide*, Trafficware, Albany, CA.

Hale,D.K. (2005). *TRANSYT-7F user manual*, Mctrans Center, University of Florida, Gainesville, FL.

Roess, R.P., Prassas, E.S., and Mcshane, W.R. (2004). *Traffic engineering* (3rd Edition). Pearson Education, Inc., New Jersey.

Skabardonis, A. and May, A.D. (1985). "Comparative analysis of computer models for arterial signal timing." *Transportation Research Record, No. 1021*, 45-52.

Texas Transportation Institute (TTI). (2004). *PASSER V user guide*, College Station, TX.

Wong S.Y. (1991). "TRANSYT-7F or PASSER II, Which is better ---- A comparison through field studies." *Transportation Research Record, No. 1324*, 83-97.

Yang, X.K. (2001). "Comparison among computer packages in providing timing plans for Iowa arterial in Lawrence, Kansas." *Journal of Transportation Engineering*, 127(4), 311-318.

Evaluation of ITS alternatives targeting crash mitigation on Freeways using Micro-simulation

Mohamed Abdel-Aty[1] and Albinder Dhindsa[2]

[1] Associate Professor, Department of Civil and Environmental Engineering, University of Central Florida, Orlando, Florida, 32816-2450, Phone: (407) 823-5657, Fax: (407) 823-4676, E-mail: mabdel@mail.ucf.edu

[2]Department of Civil and Environmental Engineering, University of Central Florida, Orlando, Florida, 32816-2450, Phone: (407) 823-4552, Fax: (407) 823-4775, E-mail: albinder@gmail.com

This study provides an analysis of the methodology of using Crash Prediction models in simulation to evaluate different scenarios related to safety in dynamic traffic conditions. The analysis is carried out on a network calibrated according to the needs of the crash prediction algorithm. Multiple scenarios for ITS alternatives viz. ramp metering, variable speed limits, and their different combinations, are tested and a methodology developed for quantifying their safety benefits over multiple locations at different time intervals. The Crash Prediction models used have been developed based on I-4 loop data on a 36 mile (58 km) stretch in Orlando, Florida. A 9 mile (14.5 km) section has been calibrated using PARAMICS micro-simulation. Along with validation of volumes special emphasis has been placed on speeds and queues at the locations of interest in the network (Loop Detector stations). Also, since the crash prediction algorithms use data aggregated over 5 minute intervals, the simulation has been calibrated for Traffic Counts and Speeds at 5 minute intervals. Loop data was used to arrive at 5 minute means and variances of speeds at each of the locations in the field. The speeds in simulation were then compared to the actual scenarios to determine the best cases. The calibrated network was then used to test the safety benefits from using the different ITS strategies mentioned above. This meant testing the strategies implemented in a coordinated manner. An experimental design was used to arrive at the best possible scenario on a whole section of the freeway.

The main conclusions of the study include:

- The Crash Prediction Model used on the network calibrated at 5 minute intervals provided values closer to the actual Crash risk indices at these locations. As such any conclusions from the testing on simulation could be more effectively traced to field implementation.
- Safety Benefits at each location could be quantified and compared across locations and these measures could be used to determine an overall Safety Benefit Index between cases.
- The key effect of the ITS measures especially Ramp Metering and VSL was reducing the Variances in speed leading to lowered risk of crash.

Keywords: Micro simulation, Crash Prediction, Ramp Metering, Variable Speed Limits.

Introduction

Recent research has been focused on quantifying the risk of crash under different traffic conditions on freeways. These measures have varied from real-time (Abdel-Aty et al., 2005) based on traffic parameters, or historical (Hauer, 2001) based on crash rates. This research uses the real time measures to provide the methodology for their implementation in evaluation of ITS alternatives when the final aim is to mitigate the risk of crashes on freeways. The study provides 4 measures which can be used to evaluate the effectiveness of any strategy. These measures are derived from the statistical crash risk measure defined by Abdel-Aty et al. (2005) for congested situations on freeways. In the end, the study illustrates the successful implementation of a coordinated Ramp Metering and Variable Speed Limits strategy in congested situations and results of using each of the defined measures to analyze the results and arrive at the best practice for each of the measures.

Literature Review

Lee et al. (2002) hypothesized that the likelihood of crash occurrence is significantly affected by short-term turbulence of traffic flow. A crash prediction model was developed using log-linear analysis. In a later study (Lee et al., 2003), they continued their work along the same lines and modified this model. Golob et al. (2004, a) also showed that certain traffic flow regimes are more conducive to traffic crashes than the others. Also a number of studies now rely on Simulation to test and prove new theories and enhance the validity of the results. Chu et al. (2004) evaluate the potential benefits of ITS strategies like Local and Coordinated Ramp Metering under incident scenarios. Chu et al. (2002) and Hasan et al.(2002) also make use of simulation to evaluate different Ramp Metering algorithms. Ben-akiva et al. (2001) use MITSIMLab to evaluate Freeway control. Zhang et al. (2001) evaluate different Ramp Metering strategies using PARAMICS. Most of these studies primarily targeted congestion and travel time as their final objectives and did not consider safety as a primary objective. Lee et al. (2004) dealt with the safety benefits of using Variable Speed Limits by using Variable Message Signs. The study found that variable speed limits can reduce average total crash potential 25% by temporarily reducing speed limits during risky traffic conditions. The measure used to define the benefits was the aggregate measures for a single location. Since only a single location was being considered, no network wide measures were provided. Abdel-Aty (2005) starts off from where this study ends by using traffic data from I-4 in Central Florida and tests for the effects of Variable Speed Limits at locations with higher speeds. The testing in this study considered only single stations and there is no network wide coordinated strategy tested and a simple before and after analysis of the measure of crash risk on a time line. Lee et al. (2005) investigated the effects of ramp metering on risk of crash using their models developed in Lee (2004). The study indicated effected of Ramp Metering on the same measure of crash risk as their previous study for a single location. This study will provide the aggregate measure methodology as well as introduce new measures to capture the effects of any strategy efficiently and in a manner that could be compared across locations as well as show the effectiveness of Variable Speed Limits and Ramp Metering implemented together for all these measures when dealing with congested freeways.

Safety Metric

Abdel-Aty et al. (2005) developed a model based on matched case control logistic regression. The purpose of the matched case-control analysis was to explore the effects of independent variables of interest on the binary outcome (crash or no crash) while controlling for other confounding variable (e.g. location, time, etc.). This model is used to assess crash likelihood for the simulated network used in this study. The models were developed for the same segment of Interstate 4 under study here, making them the most appropriate choice. These models were separately developed for a moderate-to-high-speed and low-speed traffic regime and the threshold for separating the two regimes was set at 37.5 mph (60 kph) based on examination of traffic speed distributions. Above this speed, a moderate-to-high speed model, which takes average occupancy and flow as input, is used. Below this speed, a low-speed model, involving average volume, occupancy, and coefficient of variation in speed as inputs, is used. These models may be used to assess the crash potential at any given location in real-time using loop detector data. Since the input parameters to these models were measured 5–15 min before the crash, there would be time to introduce alternative strategies at locations experiencing crash prone conditions before they culminate into a crash. The nature of this study- dealing with congested freeways- means the low speed scenario applicable model should be used. This model is shown in Equation 1.

$$\text{Risk Index} = 2.64827*\text{LogCVSF2} + 0.88842*\text{LogCVSF3} + 1.33966*\text{LogAOE2} + 0.97766*\text{LogAOH3} - 0.43603*\text{SVF2} \quad (1)$$

where LOGCVSF2 is the log of the standard deviation of speed divided by the average speed at the station of interest 5–10 min before the time of interest, LOGCVSF3 the log of the standard deviation of speed divided by the average speed at the station of interest 10–15 min before the time of interest, LogAOE2 the log of average occupancy 0.5 mile (0.8 km) upstream of the station of interest 5–10 min before the time of interest, LogAOH3 the log of average occupancy 1 mile (1.6 km) downstream of the

station of interest 10–15 min before the time of interest, and SVH2 is the standard deviation of volume 1 mile (1.6 km) downstream of the station of interest 5–10 min before the time of interest. The model shown above would provide a measure that may be used to evaluate the impact of our application experiments on the safety situation of the freeway. However, due to the nature of development of this index, it cannot be used for comparison across stations when testing alternative scenarios in Micro-simulation. This is due to the control of parameters used in deriving these parameters (Abdel-Aty et al., 2005). As such, this measure can be used for before and after studies at the same location, the measure as such cannot be used for the analysis across locations.

This overall study aims at identifying successful risk mitigation strategies and tries to optimize them for application. The Measure of effectiveness (MOE) used was the cumulative Risk Index averaged over 20 runs from simulation for each case. Equation 2 shows the measure of Effectiveness for the East Bound Direction.

$$(MOE)_i = 1/n(\sum_{j=1}^{n}\sum_{t=1}^{360}(RiskIndex)) \quad (2)$$

(MOE) = Mean Cumulative Index (MCI)

i= Station Number (33-49 for East Bound Direction); j=Seed Number (20); t= time slice (360 for 3 hours evaluated every 30 seconds). The index would serve as the primary MOE and is called the Mean Cumulative Index (MCI). As stated above, the methodology means that the model would give a unique index for each station. As such no formal conclusion can be made based on the individual safety indices between stations in general. For example, using the index for two stations at the same time, we cannot compare and tell which one is corresponding to a high risk potential. The only effect we can inspect is if the index was sufficiently influenced by an alternative. Thus, the index can be said to be spatially independent and the same value of index at different location does not amount to same risk of crash. The same difference in the before and after index, however means that the risk reduces by the same magnitude for each location. The MCI provides a quicker comparison index instead of comparing each location with risk index. Thus, the safety benefits from any case are measured by the area between the risk index curves for the base case and the test case. When summed over the entire network and divided by the area under the curve for base case, an index is obtained which illustrates the safety benefit for that particular test case over the base case. This calculation of the Safety Benefit Index (SBI) is illustrated below:

Safety Benefits Index= {| (Area under Test Case – Area under Base Case) |/| Area under Base Case |} (3)

Station wise comparisons are done using just the Safety Benefits for just that station and are calculated as illustrated below.

Safety Benefits Index= {| (Area under Test Case – Area under Base Case) |} (4)

Figure 1 shows the same areas graphically. The shaded regions in the figure refer to the benefits over the whole period of simulation at a particular station for the base strategy and a test strategy. These 3 indices (Risk Index, MCI, SBI) when used together for comparison would provide a good basis for determining the risk of crash in base case and the test case over the whole network. These measures will be illustrated later in the testing of the scenarios. These measures for quantifying and comparing the risks however, are not enough and there is a need for making them understandable in terms of traffic parameters and interpret the effect of the strategies. Variation of speeds is looked at as a measure which could be used to determine the simple effects of the alternative strategies for crash mitigation.

Study Section

This study is done by simulating a 9 mile (14.5 km) section (study section) of Interstate-4 in metropolitan Orlando area. The study section consists of 17 loop detectors, 11 on-ramps and 10 off-ramps in the East Bound Corridor of Interstate-4. The loop detectors are placed on three lanes of I-4 and spaced roughly at half a mile (0.8 km). The study section spans the area from just north of I-4 and SR 441 interchange to north of I-4 and Fairbanks Avenue interchange.

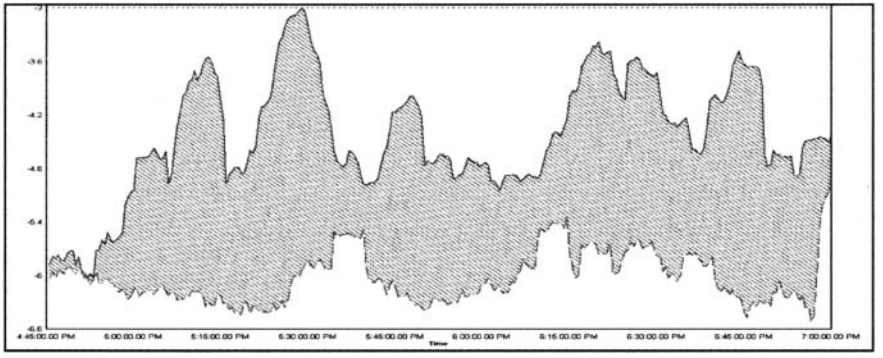


Figure 1 Illustration of Base Case Index (solid line on top), Test Case Index (dotted line below) and the benefits over the period of simulation (shaded area)

Micro-simulation

PARAMICS is used as the simulation software of choice. It is chosen because of the proven background of PARAMICS in studying such cases. Lee et al. (2004; 2005) also used PARAMICS in their study. USDOT (2003) published a report on using Micro-simulation for study of surrogate safety measures. The study reported that PARAMICS, VISSIM and AIMSUN are the simulation suites best serving the purpose when studying surrogate safety measures. Calibration of the network in PARAMICS was done using actual loop data from I-4 to ensure the validity of the results. The loop detectors on I-4 collect data on average vehicle counts, average speed, and lane detector occupancy every 30 seconds. This data was archived at UCF for the period 1998-2004. This Loop data was used from 20 weekdays (Tuesday, Wednesday and Thursday) from September 23 to November 20, 2003. The flows and speeds at 5 minute intervals were compared for 30 intervals from 4:30 to 7:30 PM for simulation and actual loop data over 17 detector stations and reduction in this error kept as the objective function. Research using Simulation has been focused towards calibrating driver behavior parameters and validating using aggregated data requirements (Hourdakis et al., 2002; Cheu et al., 1998). The basic procedure involved is adjusting OD until a reasonable flow on links is obtained and then using the links to adjust driver behavior parameter to adjust flows. However, most of the studies stop at 15 minute resolution of data for validation and almost never even consider such resolution for speeds (Chu et al. 2003). This study focused on insuring that the calibration gave results for both speeds and flows at the loop detector stations at every 5 minutes for the whole stretch of the study section. This is because these parameters - flows and speeds - are used in the statistical measures (Equation 1) at a 5 minute resolution and hence they needed to be calibrated at the same resolution. At the end of calibration process, the overall speeds had an average error of 18.9% calculated every 5 minutes and flows had an average error of 8.26 % calculated every 5 minutes. The details of the calibration process can also be found in Dhindsa (2005).

Results

Feedback Ramp metering strategy ALINEA was used along with a corridor wide Variable Speed Limits strategy on the simulated network and each of the measures for comparing the before and after crash risk conditions for all the cases of the experimental design. In total 24 cases were evaluated for investigating the coordinated ITS strategies, with 3 experimental variables. These were VSL strategy (4 levels, 4 different strategies (Dhindsa, 2005)), Number of metered ramps (3 levels, 0, 3 and 7 ramps), and speed increments (2 levels, 5 mph (8 kph) and 10 mph (1.6 kph)). The reason for the selection of the above levels of variables was to control the size of the experimental design. Although the effect of each addition of ramp meter might provide more information on an operations scale, the total number of cases to be considered would make it unfeasible for a complete experimental design. Figure 2 shows the risk

index for station 43 on I-4 for the simulated network in the base scenario and the case with VSL and Ramp Metering at 3 locations. The inset (Figure 2b) also shows the MCI for all the stations under consideration with some of the best cases from the experiment. The 2 figures show that there is a significant change (indicated by MCI for test case being outside the t-interval for random variance indicated by maximum and minimum lines in Figure 2b) in the risk index if the MCI shows a drop.

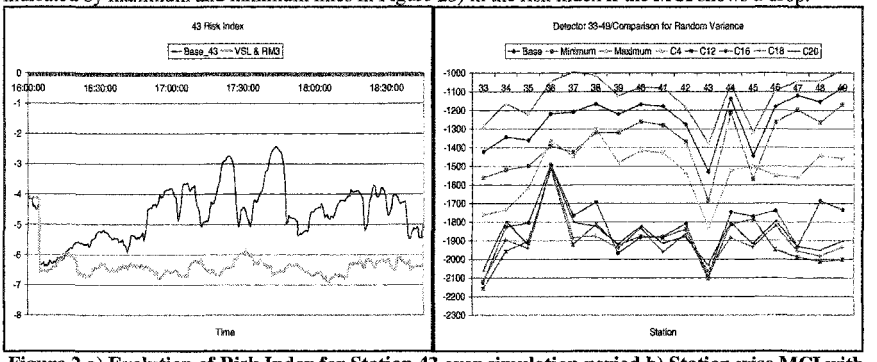


Figure 2 a) Evolution of Risk Index for Station 43 over simulation period b) Station wise MCI with t interval for base case and best cases in design

However, there is still the possibility of the MCI showing a significant difference when actually there might be an undesired peak in the risk index. Figures 3a and 3b show that under such conditions the areas under the risk index would provide the best indicator. Figure 3a and 3b show the case wise and station wise comparisons of the areas. The higher SBI index indicated greater positive area (i.e. risk decreases). The figures show that the cases with lower SBI have lower positive areas at individual stations and vice versa. These comparisons together can help in identifying the best case scenarios. For the better performing cases, the decrease in variance of speeds is indicated by the 30 second speed averages in Figure 4. A lower Risk Index is indicated by a low variance in speeds.

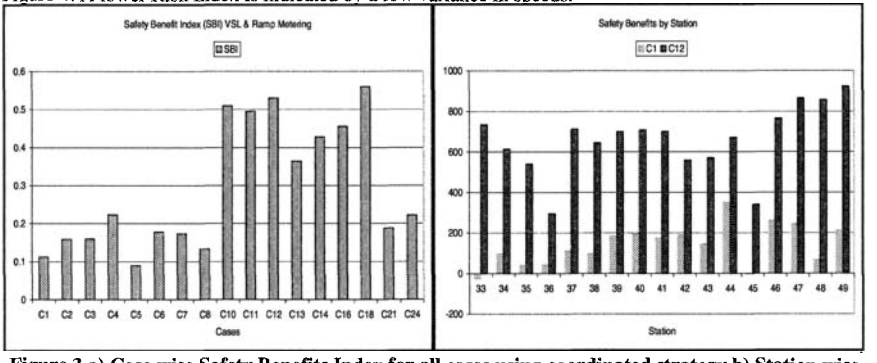


Figure 3 a) Case wise Safety Benefits Index for all cases using coordinated strategy b) Station wise comparison of benefits for case without coordination (only VSL) and in a coordinated ITS setup

Conclusions

This study presents a methodology for evaluating best practice strategies for alleviating the risk of crashes on urban freeways by ITS strategies using micro-simulation. The micro-simulation network is calibrated for the same resolution in time for flows and speeds (5 minutes) as the measures used in the

crash prediction models. The 3 measures of evaluating the strategies were shown and used in the study to evaluate the effectiveness of a coordinated Variable Speed Limits and Ramp Metering strategy and the effects over the whole network. All measures show a significant decrease in Risk of crash and significant benefits of using the strategy in congested situations. Apart from the 3 surrogate measures, the speed profiles also show stabilization in 30 second speeds indicating a decrease in the speed variances.

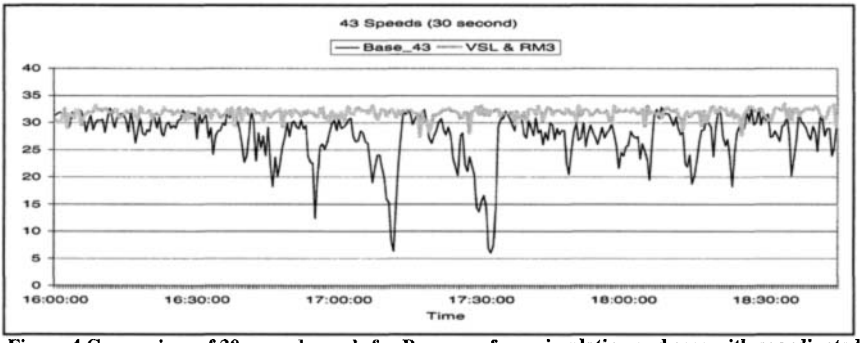


Figure 4 Comparison of 30 second speeds for Base case from simulation and case with coordinated VSL and Ramp Metering in effect

References:

Abdel-Aty, M., Dilmore, J., Dhindsa, A., "Applying Variable Speed Limits And The Potential For Crash Migration." Forthcoming in Journal of Accident, Analysis and Prevention, December, 2005.

Abdel-Aty, M., Uddin, N., and Pande, A. "Split Models for Predicting Multi-Vehicle Crashes during High-Speed and Low-Speed Operating Conditions on Freeways." Transportation Research Board, 84th Annual Meeting, Washington, D.C, 2005.

Ben-Akiva, M., Cuneo, D., Hasan M., Jha, M., and Yang, Q. "Evaluation of Freeway Control using a microscopic Simulation Laboratory." Transportation Research, Part C, vol.11, No.1, pp.29-50, 2003.

Cheu, R. L. , Jin, X., and Srinivasan, D., "Calibration of FRESIM for Singapore expressway using Genetic Algorithm." Journal of Transportation Engineering, Vol. 124, No. 6, pp. 526-535, 1998.

Chu, L., Liu, H., and Recker, W. "Development of A Simulation Laboratory for Evaluating Ramp Metering Algorithms." Transportation Research Board, 81st Annual Meeting, Washington, D.C, 2002.

Chu, L., andYang, X. "Optimization of the ALINEA Ramp-metering Control Using Genetic Algorithm with Micro-simulation" Transportation Research Board 82nd Annual Meeting, Washington, D.C., 2003.

Dhindsa, A., "Evaluating Ramp Metering and Variable Speed Limits to Reduce Crash Potential on Congested Freeways using Micro-Simulation." MS Thesis, University of Central Florida, 2005.

Hasan, M., Jha, M., and Ben-Akiva, M. "Evaluation Of Ramp Control Algorithms Using Microscopic Traffic Simulation." Transportation Research Part C, vol.10, No.3, pp.229-256., 2002.

Hauer, E. "Over-dispersion in Modeling Accidents on Road Sections in Empirical Bayes Estimation.", Accident Analysis and Prevention Vol. 33, pp. 799-808, 2001.

Hourdakis, J., Michalopoulos, P.G., and Kottommannil, J. "A practical procedure for calibrating microscopic traffic simulation models." Transportation Research Board, 82nd Annual Meeting, Washington D.C., 2002.

Lee, C., Hellinga, B., and Saccomanno, F. "Real-time crash prediction model for the application to crash prevention in freeway traffic." 82nd Annual Meeting of Transportation Research Board, 2003.

Lee, C., "Proactive Vehicle Crash Prevention on Instrumented Freeways Using Real-Time Traffic Control." Ph.D. Thesis, University of Waterloo, Ontario, Canada, 2004.

U.S Department of Transportation. "Surrogate Safety Measures from Traffic Simulation Models - Final Report." Publication No: FHWA-RD-03-050, 2003.

Heterogeneous traffic flow modeling and simulation using cellular automata

Tom V. Mathew[1], Pradip Gundaliya[2] and S L Dhingra[3]

[1] Assistant Professor, Department of Civil Engineering, Indian Institute of Technology Bombay, India; PH: (9122) 2257 7349; FAX: (9122) 2257 7302; email: vmtom@iitb.ac.in
[2] Research Scholar, Department of Civil Engineering, Indian Institute of Technology Bombay, India; PH: (9122) 2576 7349; email:gundaliya@iitb.ac.in
[3] Professor, Department of Civil Engineering, Indian Institute of Technology Bombay, India; PH: (9122) 2257 7329; FAX: (9122) 2257 7302; email: dhingra@iitb.ac.in

Abstract

This study presents an attempt to simulate heterogeneous traffic using the principles of cellular automata. A conventional cellular automata model for the homogeneous traffic is modified to handle heterogeneity. The principle adopted in this study is to retain the passenger car equivalent concept of HCM and the basic structure of the CA model. The modifications are done at three levels: First, the cell size is reduced to take the concept of passenger car equivalents. Second, the randomizing rule is modified to suit heterogeneous traffic conditions. Finally, the lane changing rules are modified to handle overtaking maneours and considers different types of vehicles characterized by their dynamic characteristics. The models were calibrated and validated with limited data sets and the performance of the model is presented.

Introduction

Most of the traffic flow models are developed for homogeneous traffic where majority of the vehicles are passenger cars. These models cannot be implemented directly for the arterial of developing countries having vehicles with diverse physical and dynamic characteristics. Due to the complexities involved in developing mathematical models for heterogeneous traffic, simulation based approaches are usually adopted by most of the researchers. Car following models based on Newtonian laws and stimulus-response hypothesis are used to describe movement of vehicles in these models. The position, speed, and acceleration of every vehicle at each time interval need to be computed from the basic equations of motion. Thus, although these approaches provide accurate individual vehicle details like position at every desired time, they require high computational resources making them unsuitable for many real time application. This has motivated the development of computationally efficient model with reasonable accuracy. In this context, cellular automata (CA) has emerged as an efficient tool in modeling traffic flow. CA discretizes continuous variables and uses simple vehicle movement rules (acceleration, deceleration, randomization, and updation) to describe traffic flow. Several studies demonstrated the use of CA in traffic flow modeling and proved its computational efficiency over microscopic models. It may be noted that although these models are not as computationally efficient as macroscopic models, they offer adequate details about the traffic flow phenomenon like shock waves. Therefore,

in this study, heterogeneous traffic flow modeling using cellular automata is attempted for a urban arterial and their potential use is for real time applications. In this approach, CA model for homogeneous traffic is modified for heterogeneous traffic conditions. This modification is done at three levels. First, the cell size is reduced to take the concept of passenger car equivalents. Second, the randomizing rule is modified to suit heterogeneous traffic conditions. Finally, the lane changing rules are modified to handle overtaking maneours and considers different types of vehicles characterized by their dynamic characteristics. The models were calibrated and validated with limited data sets and the performance of the model is presented.

Review

Most of the simulation models were developed for homogeneous traffic. The earliest work on simulation of traffic flow at mid-blocks was done by Gerlough (1956) where a two lane freeway segment of 0.25 miles with an additional end section is simulated. Since then several simulation models have been developed using car-following models (Aycin and Benekohal, 1999). Although these models were studied well and give accurate micro-characteristics results after extensive calibration and validation, they require large computing resources making them unsuitable for many real time applications. Further, many practical problems like traffic controls do not require such accuracy. This has lead to the emergence of cellular automata (CA) models for traffic flow simulation. Nagel and Schreckenberg (1992) used CA for microscopic traffic flow model for single lane and motivated the emergence of several studies using CA. In this model, traffic system is divided into lattice of the equal size of cells typically 7.5 m assuming homogeneous traffic. Nagel et al. (1998) have done extensive review of CA models and summarized different approaches to homogeneous traffic flow modeling. However, heterogeneous traffic poses some unique challenges (Ramanayya, 1988)

Several simulation studies for heterogeneous traffic have been conducted in the past to capture the mixed nature of traffic flow prevailing in developing countries (Popat et al., 1990). Marwah and Bandyopadhyay (1983) have developed a computer simulation model for a road section having four lanes - two in each direction, with signalized intersections in Calcutta city. The five types of vehicles considered in the model are car, mini-bus, single-Decker bus, double-Decker bus and tram. The are several other studies on traffic flow simulation on heterogeneous traffic, either at macroscopic models (Wong and Wong, 2002) or more commonly at microscopic level (Khan and Maini (2000); Koshy and Arasan (2005)) All the above models considered heterogeneous traffic flow at microscopic level. They have modeled heterogeneous traffic for single lane to multi lane road. Many of them have incorporated the effect of the geometric features of the road to the driver behavior for the various situations. A simulation approach for the solution can be explored, particularly when more complexity arises as in the case of heterogeneous traffic. Car-following simulation models are found more accurate than the macroscopic flow models. However, they are found to be considerably slower for any on-line application. Cellular automata found a new computationally efficient way to simulate the vehicular traffic. CA has proved it's capability to predict the behavior of traffic phenomenon and overcome some of the drawbacks of the mathematical modeling. Due to the discrete nature of CA model, it is becoming more

popular for computer implementation and real time application. In the recent past, CA has been successfully applied for the homogeneous traffic flow modeling. CA is computationally efficient and it can handle problems of large size as it works with the cells which represent the vehicles and this seems to be more realistic than the aggregate behavior. Hence, in the present study, an attempt has been made to model the heterogeneous traffic flow using CA for unidirectional traffic stream.

Methodology

The strategy adopted to develop a heterogeneous traffic flow model is by modifying an efficient homogeneous traffic flow model. The model developed by Nagel and Schreckenberg (1992), popularly known as NaSch model is adopted as the homogeneous model for modification. In this, two principles are adopted: First, the computational advantage of CA model arising out of discretisation of variables and adoption of uniform cell size are retained so the proposed model resembles to classical CA. This is different from the approach adopted by Lan and Chang (2005) where some vehicles are represented by multiple cells. Second, the usual concept of converting all the diverse vehicles into passenger car equivalent (PCE) concept proposed by Highway Capacity Manual (2000) is also retained. Accordingly, to incorporate heterogeneity, the NaSch model is modified by changing cell size based on the vehicle types and by considering different types of vehicle depending on the dynamic characteristics like maximum speed, acceleration and deceleration. This has resulted in a lower cell size of 5 meters representing a PCE. Although at a microscopic level, this may appear to be irrational, but for large runs, they are expected to be accurate. It is also proposed to use modified randomization rule for better performance of the model in heterogeneous traffic flow where a diver react not only with the vehicle ahead, but also the density ahead. The algorithm for the single lane model is shown in Figure 1. The model requires information like cell size, which represent the standard vehicle with the required clearance; types of vehicles and maximum velocity of each type of vehicle; the initial density (cells occupied with vehicles); mean arrival rate; classified volume for each type of vehicles; incident details like place and duration of incident; and driver behavior probabilities. The illustration of the working of the CA model and the application of the CA rules are shown in Figure 2. The developed models are first tested with homogeneous traffic. These are validated at microscopic and macroscopic level. The microscopic validation is done with the output from the VISSIM software which is a conventional micro-simulation model. Validation at the macro level is attempted with limited field data. The model provides relations of fundamental parameters, trajectory of individual vehicles, and animation of traffic flow. The output of macroscopic parameters shows that the models are giving satisfactory result. The two lane model is an extension of the single lane model, but permits lane changing and overtaking. Various lane changing rules were studied by giving due consideration to the drivers behavior. These model needs additional inputs like lane changing probability and maximum speed allowed in each lane. At each time step t, the arrangement of all the vehicles on cells are updated simultaneously according to the following rules:

Rule 1 (Acceleration): If speed of the vehicle is lower than v_{max}^k, then the driver's tendency is to reach the maximum speed and at every time step, the

```
Input: Vehicles, lanes and cell details,
       arrival pattern, classified volume,
       and noise probability
Initialization: t=0
   Generate initial vehicles
begin
   Calculate gap
   Application of CA rules
      Acceleration
      Deceleration
      Randomization
      Vehicle position update
      Lane changing
   Vehicle generation
   t=t+1
end stopping criteria
Output: Time-space data
```

Figure 1: Algorithm of single lane CA model

vehicle increases its speed. The rule is given as $v_n \longrightarrow min(v_n + a_k, v_{max}^k)$

Rule 2 (Deceleration): The vehicle reduces its speed if the front gap (spacing) is not sufficient to drive at current speed. The rule is given as $v_n \longrightarrow min(v_n, gap_n^f - 1)$

Rule 3 (Incident Occurrence): If any incident occurs in the stream, the vehicles start accumulating at the place of incident and when the road stretch is cleared then the vehicles start dissipating from the queue. If $(x_n^{t-1} \leq x_{ip}$ and $x_n^t > x_{ip})$ then $v_n = 0 : \quad T_1 \leq t < T_2$ where, x_{ip} is the cell location of the incident, x_n^{t-1} is the position of n^{th} vehicle in time-step $t-1$, x_n^t is the position of n^{th} vehicle at time-step t, T_1 is the start time of the incident, and T_2 is the end time of the incident.

Rule 4 (Randomization): As mentioned earlier, in CA models, the stochastic driver behavior is incorporated by introducing noise probability p_n. The rule is given by $v_n \longrightarrow max(v_n - 1, 0)$ *with* p_n

Rule 4 (Lane Changing Rule): This rule allows the vehicle to move from one lane to the other depending on the trigger and safety criteria at every alternate time step.

Rule 6 (Position updation): Application of the above rules give new speed and position to the vehicles. By using this speed and the position in the current time step, the position at the next time step is updated accordingly. The rule is given as: $x_n \longrightarrow x_n + v_n$

The model is validated with the field data collected at macroscopic level as well as number of lane changing occurred. The developed model is applied to a long two lane

road with the incident occurrence and time space plot is shown in Figure 3.

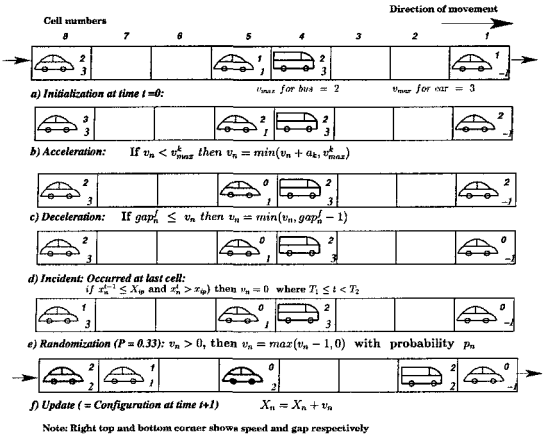


Figure 2: Illustration of modified CA rules for single lane traffic

Conclusion

In the present study a CA based model is developed for heterogeneous traffic flow simulation. The model is developed from a conventional CA models by reducing the cell size and modifying the randomization rule. While addressing heterogeneity, the basic structure of CA approach and the concept of passenger car equivalent is retained. The above model is very generic in the sense that it can handle vehicles of any dynamic characteristics and is quite efficient. It can also simulate many of the traffic situations for instance, free and forced lane changing, and incident. These models need to be tested for diverse traffic and road conditions before any real time applications.

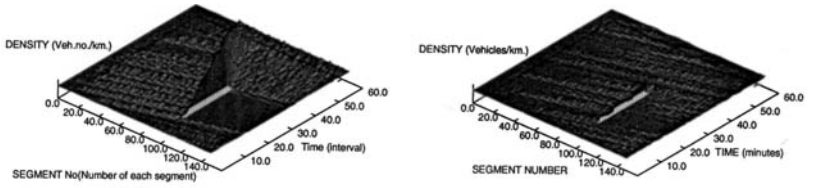


Figure 3: The effect of incident when lane changing is not permitted and when lane changing is permitted

References

Aycin, M. F. and Benekohal, R. F. (1999), "Comparison of car-following models for simulation", *Transportation Research Record* , Vol. 1678, pp. 116–127.

Gerlough, D. L. (1956), Simulation of freeway traffic by an electronic computer, Proc., 35th Annual Meeting, Highway Research Board, pp. 548–557.

Highway Capacity Manual (2000), National research council, Technical report, Transportation Research Board, Washington, D.C.

Khan, S. and Maini, P. (2000), "Modelling heterogeneous traffic flow", *Transportation Research Record* , Vol. 1678, pp. 234–241.

Koshy, R. Z. and Arasan, V. T. (2005), "Methodology for modeling highly heterogeneous traffic flow", *Journal of Transportation Engineering, ASCE* , Vol. 131, pp. 544–521.

Lan, L. W. and Chang, C.-W. (2005), "Inhomogeneous cellular automata modelling for mixed traffic with cars and motorcycles", *Journal of Advanced Transportation* , Vol. 39, pp. 323–349.

Marwah, B. R. and Bandyopadhyay, S. A. (1983), "Development of a traffic simulation model for an Indian city street", *Journal of Indian Roads Congress* , Vol. 46, pp. 655–695.

Nagel, K. and Schreckenberg, M. (1992), "Cellular automaton model for freeway traffic", *Journal de physique I* , Vol. 2, pp. 2221–2229.

Nagel, K., Wolf, D. E., Wagner, P. and Simon, P. (1998), "Two lanes traffic rules for cellular automata: A systematic approach", *Physical Review E* , Vol. 58, pp. 1425–1437.

Popat, T. L., Gupta, A. K. and Khanna, S. K. (1990), "A critical appraisal of simulation of at-grade highway intersection - a global scenario", *Journal of Institute of engineers(India)* , Vol. PartCV2.

Ramanayya, T. V. (1988), "Highway capacity under mixed traffic conditions", *Traffic Engineering and Control* , Vol. 29, pp. 284–287.

Wong, G. C. K. and Wong, S. C. (2002), "A multi-class traffic flow model - an extension of lwr model with heterogeneous drivers", *Transportation Research Part A* , Vol. 36, pp. 827–841.

A Comparison of Microscopic Simulation Models: FRESIM, VISSIM and CELLSIM Based Weaving Model

Shengnan Kan[1] and Ghulam H. Bham[2]

[1] PH (573) 202-0848; email: skrg8@umr.edu
[2] PH (573) 341-6286; email: ghbham@umr.edu
Civil, Arch. & Envr. Engineering, University of Missouri-Rolla, 1870 Miner Circle, Rolla, Missouri 65409

ABSTRACT

The objective of the paper is to compare the performance of VISSIM, FRESIM and CELLSIM based weaving model and provide recommendations for calibration of these models. This paper analyzes a weaving section with these microscopic traffic simulation models and compares their results with the field data. The field data was collected on Baltimore-Washington Parkway, on a Type A ramp weave section, part of a cloverleaf interchange, at every second for individual vehicles using aerial photography. The comparison between field data and the results of simulation models is made for traffic volume, section density and average speed. Linear regression analyses (R-squares) are used to compare output of simulation models. The error tests are performed to evaluate which simulation model better fits the field data for a weaving section. The comparison shows that the CELLSIM based weaving model is the easiest to calibrate and validate and it provides the best fit to the field condition. The results of section density are closer to the field data compared with the results of average speed. The important variables which greatly influence the results of simulation programs are also provided in the paper.

INTRODUCTION

Microscopic traffic simulation models such as FRESIM and VISSIM are better suited to analyze and evaluate freeway facilities compared with traditional approaches. These models can provide detailed information on volume, section density, average speed and the queue length near the ramp under congested conditions. However, more information is needed to calibrate these models and utilize them for evaluating traffic conditions. To address the calibration of some of the important parameters in these models, this paper presents comparisons of these models based on calibration and simulation results from a weaving section. Simulated throughputs were compared with actual volumes, section densities and average speeds.

FRESIM (Halati, Henry and Walker, 1998) is a microscopic stochastic simulation model of freeway traffic. It is based on a link-node network model. The links represent the roadway segments while the nodes mark a change in the roadway, an intersection, or entry points. VISSIM (Wiedemann and Reiter, 1992) models driving perception and

behavior in detail. It combines psychological aspects and physiological restrictions of the driver's perception and uses many stochastic variables to represent different perception, evaluation and decision of the driver population. The weaving model based on CELLSIM (called the Weaving Model in this paper) (Bham, 2003) represents simplified driving behavior. The most important driver behavior parameters are used in the model as it is a simplified microscopic traffic simulation model. The purpose of the Weaving Model is to reduce the computational effort, simulate a bigger traffic network and minimize the number of calibration parameters.

FIELD DATA

The field data (Smith, 1985) was collected on the Baltimore-Washington Parkway (I-95 NB), a Type A ramp weaving section on a cloverleaf interchange. The weaving section has two freeway lanes and an auxiliary lane. The section is 1606 feet long, with the entrance gore and exit gore at 643 feet and 1335 feet from the start of the section, respectively. The configuration is shown in Figure 1. The data was collected every second for every vehicle for an hour using aerial photography. The first 1800 seconds of data is used in this study.

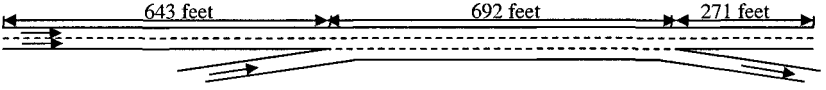


Figure 1. Weaving Section Configuration

When applying simulation models, the field data should be entered as much as possible, referred to as input parameters. The different input parameters for FRESIM, VISISM and the Weaving Model and are listed in Table 1.

Table 1. Input Parameters for FRESIM, VISSIM and the Weaving Model

Input Data	FRESIM	VISSIM	Weaving Model
Volume for every minute	√	√	√
O-D for whole half an hour	√	√	√
Lane Distribution of Entering Vehicles	√		√
Initial speeds for every vehicle			√
Initial positions for every vehicle			√
Destination lane for individual vehicle			√

CALIBRATION PARAMETERS

Calibration parameters need to be modified in a manner that the parameters represent the characteristics of the field data. First, the parameters with the greatest impact on simulation results are selected. Then, the field data are analyzed. The mean values and the standard deviations are calculated for field speeds, headways and accelerations and stochastic distributions are used to fit these data sets.

FRESIM

In FRESIM, four parameters are calibrated including speed, headway and lane-changing aspects. The calibrated values are based on the field data and the closer to the field data, the better. The sequence that the parameters are calibrated is as follows.

- Vehicle Entry Headway Distribution

In FRESIM, three distributions are used to specify the vehicle entry headway. They are Uniform, Normal and Erlang Distribution. The normal distribution is assumed and used in the simulation model to represent conditions close to field data.

- Average Free Flow Speed

The default value of average free flow speed is 65 mi/h. The calibrated values are 60 mi/h for freeway links and 35 mi/h for ramps.

- Distribution of Free Flow Speed by Driver Type

Distribution of free flow speed by driver type is based on the field speed distribution, and modified to make the total of ten driver types equal to 1000. The default values and calibrated values are shown in Table 2.

Table 2. Default and Calibrated Values of Free Flow Speed Distribution

Driver Type	1	2	3	4	5	6	7	8	9	10
Default	88	91	94	97	99	101	103	106	109	112
Calibrated	80	90	95	98	100	100	102	105	110	120

- Off-Ramp Reaction Point

This specifies the distance from the off-ramp gore to an information sign at which drivers begin to react to the off-ramp from this link. The default value is 1000 ft. In this model, the length from the start of the section to the off-ramp is 1335 ft and is used as off-ramp reaction point.

VISSIM

VISSIM is a sophisticated model as it uses various driver behavior parameters for both car-following and lane-changing. It also uses driving behavior parameters for speed, acceleration and deceleration. Among these, desired speed distribution, headway time, and acceleration at 50 mi/h are calibrated with the field speeds, headways and accelerations. However parameters such as maximum acceleration and deceleration, desired acceleration and deceleration are difficult to calibrate with field data. Therefore, default values are used. The simulation speed selected in this model is two steps per simulation second, because simulation speed of one step per second cannot input all vehicles into the model and a significant difference is found between simulation volumes and field volumes. The parameters and the calibration sequence are shown below.

- Desired Speed distribution

The desired speed distribution of entering vehicles is normal distributed with the mean value of 60 mi/h and the standard deviation of 5 mi/h.

- Car-Following Behavior

Two car-following parameters are calibrated in this model: the headway time (CC1) and the acceleration at 50 mi/h (CC9). The default headway time is 0.9 second; the calibrated value is 0.8 second. The default acceleration at 50 mi/h is 4.9 ft/sec^2. Values from the dual-regime acceleration model (Bham, 2003) are used to calibrate the parameter. The acceleration is 2 ft/sec^2 when speeds are greater than 40 ft/sec.

- Number of Observed Preceding Vehicles

The default number of observed preceding vehicles is two. The calibrated value changes to one vehicle.

- Lane Change Distance

Lane change distance is similar to off-Ramp Reaction Point in FRESIM. This parameter has significant impact on lane-changing behavior and simulation results. The default value is 656 ft and is too small for freeway sections with medium or high volume. Vehicles will jam around the off-ramp gore and try to make a lane change in order to exit the freeway. So a value of 1331 ft is used and is near the beginning of the section.

The Weaving Model

There are four main calibration parameters in the Weaving Model: desired speed distribution, preferred headway distribution, acceleration and deceleration (shown in Table 3). Desired speed distribution and preferred headway distribution are gathered from field data. The acceleration model in the Weaving Model (Bham, 2003) is the dual-regime model and obtained from field data by observing and analyzing vehicle accelerations. The deceleration model uses the same values as the INTRAS (FHWA, 1980) model. As these parameters can be found from the field data, they are easy to be calibrated. The simulation step is one per simulation second to keep the model simple.

Table 3. Calibrated Parameters for the Weaving Model

Desired Speed Distribution	Normal (60mi/h, $25mi^2/h^2$) Truncated below 55 mi/h and above 75 mi/h
Preferred Headway Distribution	Normal (0.96sec, $0.25sec^2$) Truncated below 0.4 sec and above 1.5 sec
Acceleration	5.5 ft/sec^2, speed < 40 ft/sec; 2 ft/sec^2, speed ≥ 40 ft/sec
Deceleration	10 ft/sec^2, normal condition; 21 ft/sec^2, emergency condition

COMPARISON OF MOES

To compare the simulation results with the field data, three MOEs are selected: entering volume, section density and average speed over time. There are several reasons for this selection. First, they represent the fundamental measures of traffic flow. Additionally, the comparison of entering volumes can help ensure that the simulation conditions are similar to actual traffic conditions. Moreover, density and speed can be used to represent the level-of-service of weaving sections. In this section, the overall performances of the simulation models are compared with the field data. Regression analysis and error tests are also used to compare the fitness of simulation results to the field data.

Volume, Section Density & Average Speed

The entering volumes for every minute are compared with the total volumes of left and right lane. For the entering volume of left lane and right lane, there is a very good relationship between the three simulation models and the field data. In the Weaving Model, the initial speeds, positions and entering times are specified for each vehicle. For each lane, the Weaving Model's volume matches the field data very closely. In VISSIM, the volume is entered for the whole link and the percentage of left and right lane

volumes cannot be entered. There is a little difference of volume ratio of left and right lanes between VISSIM and the field data. For the field data, the average ratio of left to right lane volume is 0.935, while the average ratio for VISSIM is 1.012, which means the distribution between the two lanes is nearly equal. In FRESIM, the volume percentage for each lane of the link can be specified, but the entering volume cannot be output for each lane separately.

Section density and average speed for every minute are compared for the average of two freeway lanes. The results for every one minute are presented in Figure 2. For section density, the Weaving Model's results show better fit to the field data; both FRESIM and VISSIM results show similar transience compared to the field data but are not as good as the Weaving Model's results. Similarly, for average speed, the Weaving Model's results show better fit to the field data. VISSIM results can reflect some properties of speed but the relationship is not very good. FRESIM results do not fit the field data well and are devoid of transience. The average speeds are nearly constant during the time period. Further discussion about the FRESIM model and its behavior is beyond the scope of this paper.

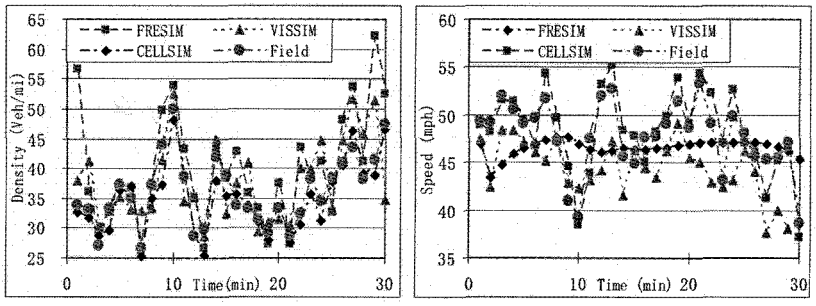


Figure 2. Section Density and Average Speed

Statistical Analysis

Regression analysis, root mean squared percent error (RMSP), and Theil's inequality coefficients (U) (Bham, 2004) are used to analyze the relationship between simulation results and field data. The aim of first two analyses is to find how well the field data can be explained by simulation. Theil's inequality coefficient, bias proportion (U^M), variance proportion (U^S) and covariance proportion (U^C) are used to provide specific information about the nature of the error. The results are presented in Table 4.

The Weaving Model's densities and speeds fit the field data very well. R-square is very high and RMSP is low. The error due to incomplete covariation is the largest source of error. For VISSIM results, the relationship between simulation and the field data is not very good. R-square is lower and RMSP is higher than the Weaving Model's value. For section density, the error due to incomplete covariation is the largest source of error, while for average speed, the largest source of error is the error due to bias. For FRESIM, the relationship between simulation and the field data is not very good. For section

density, the error due to incomplete covariation is the largest source of error, while for average speed, the largest source of error is the error due to variance. FRESIM doesn't show good fit to the field speeds.

Table 4. Statistical Analysis Results

Statistical Measure /Error Tests	FRESIM		VISSIM		the Weaving Model	
	Section Density	Average Speed	Section Density	Average Speed	Section Density	Average Speed
R^2	0.58	0.04	0.53	0.27	0.86	0.71
RMSP (%)	20.09	8.71	14.98	9.68	6.55	4.76
U	0.0942	0.0427	0.0704	0.0510	0.034	0.0229
U^M	0.2967	0.0898	0.1844	0.5821	0.2491	0.0725
U^S	0.2335	0.5117	0.0721	0.0080	0.0002	0.0840
U^C	0.4697	0.3985	0.7435	0.4098	0.7507	0.8435

CONCLUSIONS

From the calibration of three microscopic simulation models, the CELLSIM based Weaving Model was found to be the easiest to calibrate. Nearly all calibrated values can be obtained from the field data. FRESIM and VISSIM are more complicated and are developed to handle different scales of roadway networks. Calibrating VISSIM was the most complicated among the three models because its traffic model is a psycho-physical model and it combines several driving behavior parameters. Compared with the field data, the Weaving Model's results show the best fit to the field data and for all of the three models, the results of section density are closer to the field data than the results of average speed. The fit conditions of section density are similar for FRESIM and VISSIM. FRESIM and VISSIM are also similar in some aspects; such as Free-Flow Speed Percentages in FRESIM versus Desired Speed Distributions in VISSIM, and Off-ramp Reaction Point versus Lane Change Distance. VISSIM is more flexible to generate output data compared with FRESIM. Users can specify where to collect the data and what type of data is needed. A more detailed calibration procedure will be presented in a future paper.

REFERENCES

Bham, G. H. (2003). "Development and Validation of a Simplified Microscopic Multilane Model for Traffic Flow Simulation", Ph.D. Dissertation, University of Illinois at Urbana-Champaign.

Bham, G. H. (2004). "Simplified Weaving Model for Microscopic Traffic Flow Simulation", 83th Annual Meeting of the Transportation Research Board, National Research Council, Washington, D.C.

FHWA, "Development and Testing of INTRAS, A Microscopic Freeway Simulation Model", Vol. I, Report No. FHWA/RD-80/106. U.S. Department of Transportation, Oct. 1980.

Halati, A., L. Henry and S.Walker (1998). "CORSIM-Corridor Traffic Simulation Model", Traffic Congestion and Traffic Safety Conference, Chicago, ASCE.

Smith, S. A. (1985). "Freeway Data Collection for Studying Vehicle Interaction", Technical Report *FHWA/RD-85/108, Department of Transportation, U.S.*.

Wiedemann, R. and Reiter, U. (1992). "Microscopic Traffic Simulation: the Simulation System MISSION, Background and Actual State", Project ICARUS (V1052) Final Report, Brussels.

Examining the Operational Benefits of Freeway Ramp Metering Control Using CORSIM (CORridor SIMulator)

Md. Shoaib Chowdhury[1] (PhD; Member, ASCE)

[1]Professional Associate and Senior Transportation Engineer, Parsons Brinckerhoff, Quade and Douglas Inc. One Penn Plaza, New York, NY 10119, USA; PH. (212)-465-5672, FAX: 212-465-5584; email: chowdhury@pbworld.com

Abstract

A series of ramps could be metered to avoid flow breakdown or to lessen the extent of congestion at freeway downstream bottleneck locations. Although ramp metering might improve the traffic flow on the freeway but it increases the delays to vehicles behind metered ramps and induces ramp vehicles to divert if alternative routes can be found. The benefits to the overall system from metering ramps depend on factors including freeway demand flow rate immediately upstream of metered ramps, demand flow rates at metered ramps, ramp storage length and metering rate, freeway bottleneck section capacity (both normal and breakdown conditions) and traffic diverting rate around the bottleneck. The focus of this paper is to examine how above factors affect the operational benefits of metering ramps. In a simple transport network, various cases are designed and simulated using CORSIM micro-simulation model, while ramp-metering benefits are investigated. It is found that benefits from ramp metering control can be achieved the most if stable operations can be maintained on both freeway and diverting route. Benefits also depend on the demand flow rate that passes through the bottleneck sections under ramp metering control condition and this study suggest that selection of metering rate to minimize the probability of flow breakdown is important in achieving maximum benefits. It is also found that ramp metering may not be beneficial if diverting route operates poorly and congestion likely to propagates onto freeway.

Introduction

As part of freeway systems management strategy, ramp-metering control has been in practice for quite a long time with mixed degrees of success. If properly implemented, ramp-metering control may reduce the extent or prevent congestion on freeways and maximize flow through bottlenecks. The argument is that by preventing or delaying flow breakdown, ramp metering can process more vehicles at any active bottleneck locations on the freeway mainline (Banks, 1990). Many previous studies supported this argument that maximum flow rate decreases after flow breakdown occurs (Hall and Agyemang-Duah, 1991; Banks, 1991; Cassidy and Bertini, 1999). The range of drop reported varies from one percent to as high as twenty five percent (Hall and Agyemang-Duah, 1991; Cassidy and Bertini, 1999; Lorenz and Elefteriadou, 2001; Zhang and Levinson, 2004).

In addition to maximizing throughput at bottleneck section, ramp metering may also be potential for increasing freeway speed and thus reducing travel time. Piotrowicz

and Robinson (1995) reported that highway speed may increase as high as 56% by implementing ramp metering, while Hellinga and Aerde (1995) found that ramp metering may achieve total travel time reductions between 14 to 26 percent simply by avoiding the flow breakdown. However, the benefits of ramp metering may diminish if diversion creates disruption to local street traffic operations (Nsour and et al., 1992). Nevertheless, it is found that the relationship between ramp metering control and its benefits to overall system under various demand and roadway operations conditions has not been adequately analyzed.

The objective of this study is to examine the potential operational benefits of freeway ramp metering control under alternative diverting route condition. The study is based on the assumption that vehicles of metered ramp will be diverted to alternative route (in this study frontage road) if vehicle queue exceeds ramp storage length and will then be allowed to re-enter the freeway after bypassing the bottleneck section. The benefits of ramp metering control under various demands, roadway operations, ramp storage lengths and metering conditions are examined. CORSIM (FHWA, 2003), a traffic micro-simulation model developed by the Federal Highway Administration (FHWA), was utilized to conduct the experiments. CORSIM has the capability to simulate traffic operations on integrated networks including freeway and surface-street. Over the years, the model has been extensively calibrated and validated (Vadakpat and Dixon, 1999) for wide verity of traffic operations evaluation studies.

Description of the Network and Experimental Design

A hypothetical network (see Figure 1) consisting of a 3-lane freeway and a 2-lane parallel frontage road is designed as a test bed for conducting the simulation experiments. The freeway connects the frontage road via an on-ramp and two successive weaving areas, both of which are created by introducing a full auxiliary lane between on and off ramps, while all ramps are assumed to be one lane. For simplifying the work efforts, intersections between frontage road and cross streets are not modeled. The free flow speed of freeway and frontage roads are 60 mph and 30 mph, respectively.

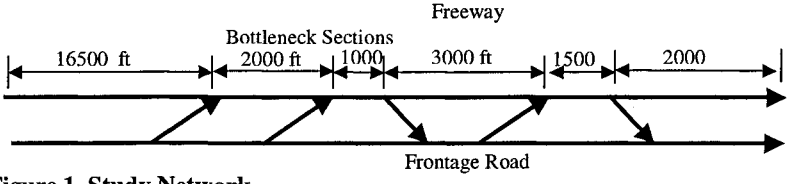


Figure 1. Study Network

To achieve the objective of this study, five cases (see Table 1) are designed. Cases 1 through 4 are designed to reflect the network volumes with different degrees of saturations, while the last case keeps the identical network volumes as Case 2, but alters the capacity of the frontage road at its last on-ramp junction area. All volumes represent hourly passenger car volumes (vph). Given the volumes for each of the five cases, two physical bottlenecks are identified along the freeway segment. The first

bottleneck section with capacity 6300 vph is the first on-ramp junction area, while the second one with capacity 7300 vph represents the first weaving area. The critical section along frontage road is the last on-ramp junction area with capacities 3100 vph under Case 5 and 4100 vph under all other cases.

Table 1. Case Number Showing Respective Network Volumes (vph)

Case #	Frontage Rd (FR) Entry Volume	Freeway On-ramp 1 (from FR)	Freeway On-ramp 2 (from FR)	Freeway On-ramp 3 (from FR)	End of FR	Freeway Entry Volume	Freeway Off-ramp 1	Freeway Off-ramp 2	End of Freeway
1	3000	600	900	300	1200	6175	1544	926	3705
2	4000	800	1200	400	1600	6175	1544	926	3705
3	3000	600	900	300	1200	5850	1463	878	3510
4	4000	800	1200	400	1600	5850	1463	878	3510
5	4000	800	1200	400	1600	6175	1544	926	3705

Given the bottleneck locations, first two freeway on-ramps are selected for metering purpose. To examine how ramp storage lengths behind metered ramps influence the benefits of metering ramps, three different ramp storage lengths including 250, 500 and 1000 ft are considered. Therefore, each of the five cases is simulated with (M) and without (N) ramp metering control for each of the three ramp storage lengths. Each simulation run extends for two time periods, the first time period with the duration of one hour, generates and simulates traffic simultaneously, while the second time period with variable lengths, simulates only network traffic, if there remains any from the pervious time period due to congestion.

While determining ramp-metering rates, it is important to consider the stochastic nature of capacities at bottleneck sections as the number of vehicles passes through the bottleneck sections relative to the capacities may impact the benefits of metering ramps. Simulation experiments showed that the capacity is a stochastic event and flow breakdown can occurs under wide verity of volume range. Moreover, capacities of on-ramp junctions and weaving areas may further depend on the distributions of mainline and ramp volumes (Chien and Chowdhury, 1998; Chowdhury, 2005).

The ramp metering rates under different Cases are selected in such a way that different flow levels at bottleneck sections can be achieved. Metering rates under Cases 1, 2, and 5 govern the flow level near capacities, while rates under Cases 3 and 4 govern the flow level fairly below capacities. Generally, there are high probabilities of occasional flow breakdowns as flow approaches near capacity, and as flow decreases and further deviates from capacity, probabilities of occasional flow breakdowns decrease. The metering rates (pre-timed and fixed) under Cases 1, 2 and 5 are set to 60 seconds per vehicles (equivalent to 60 vph) and 5 seconds per vehicle (equivalent to 720 vph) for the first and the second ramps, respectively, while under Cases 3 and 4, those rates are 15 sec per vehicle (equivalent to 240 vph) and 10 seconds per vehicle (equivalent to 360 vph), respectively. It is important to note that the metering rate of 60 seconds per vehicle applied at first on-ramp under Cases 1, 2 and 5 is extremely conservative. However, in reality to avoid metering violations, additional ramps could be metered in such a way that a practical minimum discharge rate can be maintained. Table 2 shows the estimated volumes and capacities of the two freeway bottleneck sections with and without ramp-metering controls.

Table 2. Demands (with and without Metering control) and capacities of freeway bottleneck sections

Case #	Bottleneck section 1			Bottleneck section 2		
	Demand (vph)		Capacity (vph) (before breakdown)	Demand (vph)		Capacity (vph) (before breakdown)
	Without metering	With metering		Without Metering	With metering	
1	6775	6235	6300	7650	6955	7300
2	6975	6235	6300	8175	6955	7300
3	6450	6090	6300	7350	6450	7300
4	6650	6090	6300	7850	6450	7300
5	6975	6235	6300	8150	6955	7300

Analysis of Results

After performing the simulation runs with different cases designed in the previous section, results are summarized and presented through series of figures. Figure 2 shows that system-wide benefits under ramp metering control condition can be achieved the most if stable operation can be maintained on both freeway and frontage road (Case 3). Since under Cases 1 and 2 with metering in effect, flow levels at bottleneck sections are relatively higher than Cases 3 and 4, freeway speed under Cases 1 and 2 are found to be lower than those of Cases 3 and 4 (see Fig 2).

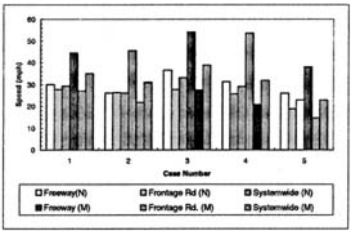


Figure 2. Speed Profile with and without Ramp Metering Control

Animation views of the simulation run under Cases 1 and 2 reveal the fact that occasional flow breakdowns mainly contribute to lower the speeds. Although under Cases 3 and 4 with metering in effect, freeway speed can be achieved very much the same, however, system-wide benefits under Case 3 is higher than Case 4 because of better operational performance of frontage road. System-wide benefits under Cases 1 and 2 are also lower than Case 3 mainly due to lower freeway speeds. Noticeably under Case 5, the system-wide speed does not improve with ramp metering control. The reason is that under Case 5 with and without metering controls, frontage road operates at congested environment due to capacity failure at its last on-ramp junction area. The congestion becomes more severe under metering condition due to diversion of traffic onto frontage road; which also results in accelerating the queue propagation process onto freeway via first freeway off-ramp. As a result, despite metering control, overall freeway speed does not increase much, while at the same time frontage road operates relatively poorly with diverted traffic. The degradation of freeway speed due

to queue propagation from frontage road can further be analyzed using time-series speed plot (see Figure 3). Figure 3 shows that under metering condition, the segment of freeway containing two bottleneck sections maintains speed at or above 40 mph until 40 minutes of simulation before it hit by off ramp queue that propagates from frontage road. The speed then gradually decreases as congestion continues to grow and reaches below no metering speed after around 50 minutes. The congestion also propagates onto freeway under no metering condition, but at a later time, at around 65 minutes. Nevertheless, the overall freeway speed with metering condition is found to be higher (see Fig. 2, case 5).

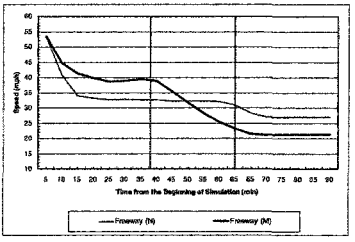


Figure 3. Time-Series Speed Plot (Case 5)

Figure 4 shows that as ramp storage length increases, vehicle delays behind metered ramp also increase. This is due to the fact that vehicles are diverted as soon as queue begins to exceed ramp storage length, thus the shorter the ramp storage length; the fewer the vehicle can be accommodated. The Figure also reveals that under identical ramp storage length condition, vehicle delays will increase as the metering rate (vph) decrease. This analysis leads to the conclusion that if vehicle diversion is possible without allowing vehicles to backup beyond ramp storage length, both the shorter ramp and the higher metering rate (vph) will lessen the vehicle delays behind metered ramps and thus will also increase overall system benefits.

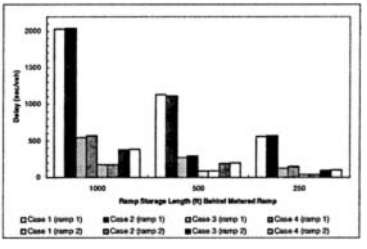


Figure 4. Vehicle Delays Behind Metered Ramps

Conclusions

In this study ramp metering benefits under various freeway demand flow rates immediately upstream of metered ramps, demand flow rates at metered ramps,

metering rates and ramp storage length are examined. Major findings include the following:

- The overall system benefits can be achieved the most under ramp metering control condition if stable operation can be maintained on both freeways and frontage road.
- Freeway speed can be increased the most under ramp metering control condition, if metering rates are selected in such a way that probability of flow breakdown minimizes
- The overall system benefits unlikely to be achieved under ramp metering control condition, if diverting road (in this case frontage road) operates poorly and congestion likely to propagates onto freeway.
- If vehicle diversion is possible without allowing queue to building up behind storage length of metered ramps, both the shorter ramps and the higher metering flow rates (vph) will lessen the vehicle delays behind metered ramps.

References

Banks, J.H. (1990) "Flow processes at a freeway bottleneck." *Transportation Research Record,* 1287, 20-28.

Banks, J.H. (1991) "Two-capacity phenomenon at freeway bottlenecks: a basis for ramp metering?" *Transportation Research Record,* 1320, 83-90.

Cassidy, M.J., and Bertini, R.L. (1999) "Some traffic features at freeway bottlenecks." *Transportation Research,* 33(B), 25-42.

Chien, S. and Chowdhury, M. (1998) "Freeway Capacity Analysis with Microscopic Simulation Model (CORSIM)." Phases I&II, FHWA-RD-97, USDOT, Final Report.

Chowdhury M. S. (2005) " Capacity Analysis of Freeway On/Off-Ramps and Weaving Areas using CORSIM." working paper # 2005-1

FHWA. (2003) "Traffic Software Integrated System." User's Guide, Version 5.1

Hall F. and Agyemang-Duah, K. (1991) "Freeway capacity drop and the definition of capacity." *Transportation Research Record,*1320, 91-98.

Hellinga, B., and Aerde, V. M. (1995) "Examining the Potential of Using Ramp Metering as a Component of an ATMS." *Transportation Research Record,* 1494, 75-83.

Lorenz, R.M., and Elefteriadon, L. (2001) "Defining Freeway Capacity as Function of the Breakdown Probability." *Transportation Research Record,* 1776, 43-51

Nsour, A. S., Cohen, L.S., Clark, E., and Santiago, J.A. (1992) "Investigation of the Impacts of Ramp Metering on Traffic Flow With and Without Diversion." *Transportation Research Record,* 1365, 116-124.

Piotrowicz, G. and J. Robinson. (1995) "Ramp Metering Status in North America." U.S. Department of Transportation.

Vadakpat, G; Dixon, K. (1999) "Calibration and validation of CORSIM for workzones." *ITE Journal*, 69(2), 12-14.

Zhang L., and Levinson D. (2004) "Some Properties of Flows at Freeway Bottlenecks." *Transportation Research Record,* 1883, 122-131

Automatic Calibration of Two-Lane Highway Traffic Simulation Models Using a Genetic Algorithm

C. Y. Egami[1], M. L. Mon-Ma[2], J. R. Setti[3] and L. R. Rilett[4]

[1]Dept. of Transportation Engineering, Universidade de São Paulo, 400 Trabalhador São-carlense Ave., S. Carlos, SP, Brazil, 13560-590; email: cintiaye@sc.usp.br
[2]Dept. of Transportation Engineering, Universidade de São Paulo, 400 Trabalhador São-carlense Ave., S. Carlos, SP, Brazil, 13560-590; email: lika@sc.usp.br
[3]Dept. of Transportation Engineering, Universidade de São Paulo, 400 Trabalhador São-carlense Ave., S. Carlos, SP, Brazil, 13560-590; PH.: (55-16)3373-9596, FAX: 3373-9602; e-mail: jrasetti@usp.br
[4]Dept. of Civil Engineering, University of Nebraska, Lincoln, W339 Nebraska Hall, P.O. Box 880531, Lincoln, NE, 68588-0531; PH.: (402) 472-1992, FAX: (402) 472-0859; email: lrilett2@unl.edu

Abstract

This paper proposes a system for automatic calibration of two-lane traffic simulation models based on a genetic algorithm. The calibration process requires searching for values of selected model parameters, as to minimize the differences between simulation results and the observed data. The calibration parameters are usually related to driver behavior and because of their large number and the fact that their affects are often highly correlated they are difficult to calibrate for specific applications. In this paper the differences between the simulation results and the observed data are assessed based on performance measures that are related to the application of the model (e.g., average travel speed, percent vehicles in platoons, etc.). The proposed procedure calibrates each model using several different highway sections to find the set of calibrated parameters that can best represent a typical two-lane highway in Brazil. Data for the calibration were collected in five different locations in the state of São Paulo which were considered representative of roads in the region in terms of terrain and traffic mix. Two data sets were collected, one for model calibration and the other for model validation. The results indicate that the proposed approach is highly efficient, albeit computationally intensive; the average difference between simulated and observed traffic data was less than five percent.

Introduction

Simulation models are very useful for the operational analysis of traffic streams. One of their greatest advantages, as opposed to observation of actual traffic streams, is the ability to control the all the conditions related to the traffic stream (McLean, 1989), which explains their extensive use in the development of the HCM 2000. Any simulation model must, however, be calibrated to properly reproduce the behavior of the traffic stream being simulated. Due to the complexity of traffic simulation models, traditional manual calibration procedures tend to be complex, time-consuming, tedious and not very efficient. This paper reports the development of an automatic system to calibrate two-lane highway traffic simulation models that is based on a genetic algorithm (GA) and easily obtainable traffic data. The applicability of the procedure is demonstrated using two models, the Australian-developed TRARR and the U.S.-developed TWOPAS. Even though the use of GA in the calibration of traffic simulation models has already been demonstrated for traffic simulation models (Cheu et al., 1998; Lee et al., 2001; Ma and Abdulhai, 2002; Rilett and Kim, 2001), the authors are not aware of any previous use of GA for the calibration of two-lane highway traffic simulation models.

Calibration of traffic simulation models

The calibration of a simulation model is the procedure through which the values of the model internal parameters are adjusted so that it is able to realistically represent the operation of the real system. Due to the complexity of traffic simulation models and the number of parameters involved, the calibration is an essentially heuristic procedure that searches for parameter values that optimize one or more fitness measures (Rilett and Kim, 2001). Thus, parameter calibration can be viewed as an optimization problem in which a parameter set maximizing an objective function is searched (Rilett and Kim, 2001).

The proposed approach is based on a GA, which is a search algorithm based on genetics and natural selection mechanisms that differ from traditional procedures in four major aspects (Goldberg, 1989). The first aspect is that, instead of parameter values, GA work with a binary coding of the parameter set. Parameters are transformed into chromosomes that are evaluated to determine how fit they are to a given problem. Genetic operators (reproduction, gene crossover and mutation) are used to create new chromosomes (e.g. micro-simulation parameter sets), thus passing beneficial and survival-enhancing traits to offspring. Chromosomes that result in simulation output that is closer to the observed data have more opportunity to reproduce thereby leading to better solutions. The second aspect is that the GA are robust in they perform searches from multiple points as opposed to single point (Gen and Cheng, 2000). This increases the likelihood of finding the global, rather than a local, optimum, especially in large, complex and poorly understood search spaces, typical of traffic simulation models. The third aspect is that GA do not use derivatives in the search. Because it is hard to identify gradients in traffic simulation models, the ability GA have to use a measure of fitness as an objective function makes them particularly attractive to calibration procedure over traditional search methods. Lastly, GA use probabilistic transition rules to guide the search to portions of the search space that will likely improve the solution, instead of simple random search (Goldberg, 1989).

GA have been used before in the calibration of traffic simulation models such as FRESIM (Cheu et al., 1998), CORSIM (Rilett and Kim, 2001), TRANSIMS (Rilett and Kim, 2001) and PARAMICS (Lee et al., 2001; Ma and Abdulhai, 2002), usually for one section of roadway. The proposed approach uses data collected on several highway sections, in order to obtain a new set of default values for the model parameters that can best represent typical two-lane highways in the region of interest, as discussed in the next section.

Proposed procedure

The proposed approach is based on the GA developed for the calibration of CORSIM and TRANSIM by Kim and Rilett (2001) and utilizes data from five different sites simultaneously to obtain new default values for model parameters. Figure 1 shows the system overall structure and its three modules: control, simulation and genetic algorithm. The control module handles input/output data used or generated by the other modules; simulations are run by the simulation module; and the genetic algorithm module applies genetic operators to chromosomes.

The automatic system works as follows. An initial population is required for the procedure and it may be created by the user or be randomly generated by the control module. This initial population is a set of P individuals, each being a set of values for the parameters subjected to the calibration. Table 1 shows the parameters used for each model, which were chosen through a sensibility analysis which identified parameters that had greatest impact on simulation results.

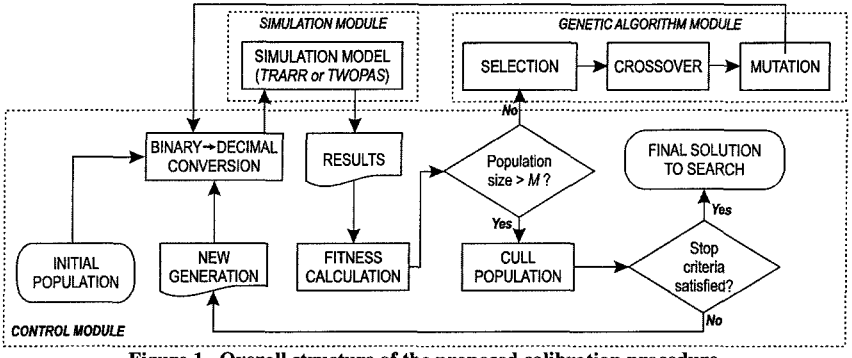


Figure 1. Overall structure of the proposed calibration procedure

Chromosomes are binary strings containing the values of the calibration parameters (genes). Each gene is string of length B given by $B \geq \log_2[(x_{max} - x_{min})/D + 1]$, where D is the desired precision and x_{max} and x_{min} are the parameter maximum and minimum values. These strings are then concatenated to create a chromosome. The bits of each gene and the total number of bits in the chromosome are also shown in Table 1. The second step in the procedure is to execute the simulations, which requires the conversion of the chromosome into decimals values for each parameter. The model runs a simulation for each of the five sites. In the third step, the fitness $F(I)$ of each chromosome I, a measure of how well the chromosome can reproduce the observed traffic flow, is calculated using Equation 1.

$$F(I) = A \cdot e^{-B[AMAER(I)]} \tag{1}$$

where A = 100 and B = 5 and $AMAER(I)$ is the average mean absolute error ratio found for the five sites simulated with the parameters contained in chromosome I:

$$AMAER(I) = k^{-1} \cdot \sum MAER(I,i) \tag{2}$$

where k is the number of sites used (k = 5) and $MAER(I,i)$ is the mean absolute error ratio for site i using chromosome I. $MAER(I,i)$ is calculated by:

$$MAER(I,i) = N^{-1} \cdot \sum_{j=1}^{N} \left[\left| OBS(j) - SIM(j) \right| / OBS(j) \right], \tag{3}$$

where N is the number of measures of performance used to compare the simulated and observed streams, $OBS(j)$ is the observed value of measure of performance j and $SIM(j)$ is the value of measure of performance j obtained from the simulation.

Table 1(a). Calibration parameters for the TRARR model

Parameter	*Definition*	*Total bits*
VOSFN	VOSFN × desired speed = desired speed when overtaking	70
VHSFN	VHSFN × desired speed = speed at which vehicle is happy to stay in platoon	98
VNP	Maximum power/weight in non overtaking situations (for trucks only)	24
VXP	Maximum power/weight in overtaking (for trucks only)	36
LAG	Driver aggression factor	56
VMF	Mean basic desired truck speed for each one of the 5 sites	160
	Total string length (bits)	444

Table 1(b). Calibration parameters for the TWOPAS model

Parameter	*Definition*	*Total bits*
PREC	Probability of passing reconsidered in review period	4
VEAN	Mean desired speed for each one of the 5 sites	35
VSIG	Standard deviation of desired speed (varies by travel direction and site)	100
VBI	Bias added to desired speed (varies by travel direction and vehicle type)	120
ZKCOR	Car-following sensitivity factor	4
BKPM	Stochastic driver type factor (varies by types of driver)	80
WOHP	Weight/net horsepower ratio (varies for each vehicle type)	36
	Total string length (bits)	379

The fourth step is the selection of chromosomes that will reproduce and create the next generation. The chromosomes with highest fitness values are more likely to be chosen with each chromosome's probability of being selected equals to the ratio between the fitness of that individual and the sum of the fitness values for the whole population. Following this, the crossover and mutation genetic operators are applied to the reproducing individuals. A crossover probability of 50% was adopted and up to $C \leq P/2$ descendents can be created, P being the number of individuals in the population. The mutation probability adopted was 30%. If a chromosome is selected to mutate, a digit is randomly selected in that string and its value is changed from 0 to 1 or vice-versa. M mutated chromosomes will be obtained and $P + C + M$ chromosomes will be obtained. The sixth step consists of executing simulations for this gene pool and of calculating fitness values. In the seventh step, the population is culled back to P individuals because the GA works with a constant size population. This is achieved by sorting the chromosomes in descending order of their respective fitness values. The chromosomes with the "best" fit will become 80% of the new generation; the remaining 20% will be composed of the individuals with the "worst" fit. This was designed to reduce chances of quick convergence to a local optimum of the fitness function. This new generation is then subjected to the same process, until one of the stop criteria is achieved. Two stop criteria are used: the maximum number of generations (350) or the minimum difference between simulated and observed traffic flows (1%), as measured by the *AMAER*. The maximum number of generations was chosen after a preliminary study showed little or no improvement when more than 350 generations were used.

Data used in the calibration procedure

Data for the calibration were collected in the state of São Paulo, Brazil. Given that the aim of the project was to replace the models default parameter set with another, better adjusted to the local conditions, the sites selected were representative of two-lane roads in the region. The five sites varied from sections in flat terrain (0% no-passing zones and 0.5% average grade) to sections in rolling terrain (42% non-passing zones and grades varying between 3% and 5%) with truck percents varying between 26% and 64% and average hourly traffic ranging from 100 to 500 veh/h. Two performance measures were used: average travel speed and percent of vehicles traveling in platoons at the end of the section. They were chosen because: (a) they are easily observed in the field; and (b) they are closely related to the intended model use, namely the adaptation of the HCM 2000 procedures to Brazilian two-lane highways. For each site, two data sets were collected in two different occasions, one for model calibration and the other for validation. Data collected in each site were related to geometric design (vertical and horizontal profiles; no-passing zones; and visibility distances for both directions) and traffic (traffic composition and percent vehicles traveling in platoons). Travel time for the calculation of average travel speed was obtained using digital camcorders with synchronized internal clocks at the beginning and end of the sections. Travel distance was obtained from design plans and GPS and varied from 4 to 8 km. Observed travel speed distributions were thus obtained for all sites and vehicle classes.

Results

The proposed GA was coded in Perl and run for the two models. The calibration of TRARR required over 340 thousand simulation runs and the best set of default parameters produced an average error of 2.55%, as shown in Table 2, which summarizes the results obtained with the proposed procedure. The maximum difference found for the TRARR model was slight above 4%. The search for the new default values for TWOPAS required 175 thousand simulations and the best solution found was associated with an average error of 2.27%, whereas the largest difference was slight less than 4%.

Table 2. Summary of the calibration and validation results for the two models

	TRARR		*TWOPAS*	
Site	*Calibration*	*Validation*	*Calibration*	*Validation*
1	1.66%	4.41%	3.78%	
2	2.23%	4.59%	1.70%	
3	4.28%	5.81%	3.67%	
4	2.30%	5.11%	0.03%	
5	2.29%	4.80%	2.18%	
Average	2.55%	4.94%	2.27%	

Figure 2 compares the quality of the results obtained with the TRARR model to observed data. It can be noticed that the recalibrated simulation model is capable of adequately reproducing the real system with respect to average speed. Additional comparisons and statistical tests were also carried out to verify the quality of the calibration, but are shown in this paper due to length restrictions.

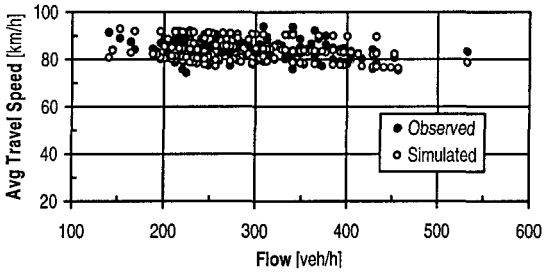


Figure 2. Simulated vs. observed average travel speeds for the TRARR model

A second set of data, collected in the same sites in a different occasion, was used to validate the results obtained with the proposed procedure. In the validation procedure, each site was simulated five times, with different seeds for the random number generation. The comparisons were carried out using the average travel speed and the percent vehicles traveling in platoons as performance measures. The averages of the five replications were used to calculate the differences between the observed and simulated traffic streams. Table 2 shows the *MAER* and fitness values found for each site andfor the calibration and the validation of the model. It can be seen that the average error found is less than 5%, ranging between 4.41% and 5.81%.

Concluding remarks

The results show that the proposed procedure is efficient, albeit computationally intensive. However, as no direct interaction with the user is required this is not an important factor. The proposed approach can easily find values for parameters that would be difficult to be obtained in any other way, such as driver aggression and even power/mass ratios.

References

Cheu, R.-L., Jin, X., Ng, K.-C., Ng, Y.-L., and Srinivasan, D. (1998). "Calibration of FRESIM for Singapore Expressway Using Genetic Algorithm." *Journal of Transportation Engineering of the ASCE*, 124(6), 526–535.

Gen, M., and Cheng, R. (2000). *Genetic Algorithms and Engineering Optimization*, Wiley-Interscience, New York.

Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Publishing Co., New York.

Lee, D., Yang, X., and Chandrasekar, P. "Parameters Calibration for PARAMICS Using Genetic Algorithm." *80th Annual Meeting of the Transportation Research Board*, Washington, DC.

Ma, T., and Abdulhai, B. (2002). "Genetic Algorithm-Based Optimization Approach and Generic Tool for Calibrating Traffic Microscopic Simulation Parameters." *Transportation Research Record*(1800), 6–15.

McLean, J. R. (1989). *Two-Lane Highway Traffic Operations – Theory and Practice*, Gordon and Breach Science Publishers, New York.

Rilett, L. R., and Kim, K. (2001) "Automatic Calibration of the Surface Transportation and Supply Models Using ITS Data." *9th World Congress on Transportation Research*, Seoul, Korea.

Acknowledgments The research reported in this paper was financed by FAPESP (grants 01/02779-8 and 03/06094-5) and CNPq (grant 303857/02-1), which the authors gratefully acknowledge.

Advanced Simulation Technology Applied to Port Safety and Security

C. Berkowitz, FASCE, Ph.D., PE, AICP[1] and C. Bragdon, Ph.D., FASA, AICP[2]

[1]Professor and Director of the Center for Intermodal Transportation Safety and Security at Florida Atlantic University, 5353 Parkside Drive. Jupiter, FL 33458; PH (561) 799-8726; FAX (561) 799-8535; email: cmberkowitz@hotmail.com
[2]Dean and Distinguished Professor, University College, Florida Institute of Technology, 150 West University Blvd., Melbourne, FL 32901-6975, PH (321) 674-8821, FAX (321) 674-7597, email: cbragdon@fit.edu

Abstract

The US maritime ports are vital multi-use facilities both for commercial and military transportation activities and their safety and security is essential to the nation's economic well-being; it is, therefore, critical to have the ability to conduct comprehensive safety and security analysis of the seaports. Approximately 95 percent of our nation's trade, valued at nearly $1 trillion, enters or leaves our 361 seaports annually. A key analysis tool for safety and security is advanced simulation technology that depicts (in four dimensions: 4-D) the air-land-seaport access and potential vulnerabilities in a virtual real-time format. This technology allows for the development of surface and underwater scenes in order to evaluate incident response training, conduct tabletop exercises, train management personnel and more effectively introduce transportation security systems.

Introduction

Since September 11, 2001, the nation's seaports have been increasingly viewed as potential targets for terrorist attacks. Security experts are very concerned that the ports can be an entry point for the smuggling of weapons and other dangerous materials into the USA and cargo and cruise ships could present potentially desirable terrorist targets as well. The ports are gateways for the movement of people and goods, and are industrial hubs located very close to population centers, presenting additional opportunities for terrorists. (Lake 05, GAO 04) A coordinated port security program is critical to protecting the American people. The 9/11 Commission stated "...while commercial aviation remains vulnerable it appears that ports are even a greater risk." (Kean 04)

There are many port security stakeholders including law enforcement, transportation agencies, vessel operators, railroads and trucking companies, port authorities, support services, businesses, and people working at the port or live nearby.

Currently, port security efforts are focused on using radar, sonar, cameras, and direct observations to track vessels, trucks, containers, and people entering or leaving the facility. Four-D simulation overlays these traditional techniques and is the next major step in the planning, designing and managing the safety and security of America's seaports. It is an important tool that can provide guidance to prevent threats and dangerous conditions. Through virtual simulation, an analytical framework is developed that enhances the understanding of a complex environment and provides guidance to facilitate security policies and practices, including: access and control, operations and management, preventing cargo theft, training personnel, evaluating regulations, risk assessment and contingency planning.

Our 361 U.S. based marine ports currently receive 8,500 foreign flag vessels, 9 million containers of cargo and approximately 200 million passengers every year. (Collins 06) This advance simulation technology provides a baseline structure to evaluate complex situations by focusing on important policies and by providing guidance for prioritizing future actions. The University Consortium for Intermodal Transportation Safety and Security (UCITSS), which represents 12 universities in the State of Florida, have focused their transport research on several modes, including maritime ports as mandated by Congress (Public 108-199). Port Everglades, has become the prototype for modeling and creating virtual real-time 4-D simulation for this 2,190 acre facility that is an economic engine, with operating revenues exceeding $66 million annually. Located in South Florida it ranks as the second largest cruise line based port in the world and is nationally ranked 12th in terms of cargo activity (TEUs) among U.S. ports. (IANA 06)

The basis for developing and applying intermodal simulation to assist in planning for all modes is based on the Intermodal Transportation Simulation System (ITSS) patented January 26, 1999 (U.S. Patent 5,863,203) and invented by Dr. Bragdon. This computer based visualization allows the individual using a mouse to drive through generated scenes from almost any perspective or eye-point, including time of day, multiple weather conditions and spatially below, on the surface or in the air. (Bragdon 06) There are many elements involved in the virtual scene generation as described below:

Simulation Software Technology (MultinGen 05)

Creator Terrain Studio Software addresses the challenges of constructing open, reusable and layered terrain databases. This software provides a unique workflow process, geospatial data management infrastructure and user interface which allows the modeler to deliver large-area terrain databases that satisfy varying simulation and training requirements (see Figures 1 and 4).

Creator Software builds high-fidelity real-time 4-D content for use in visual simulation and the integrated and extensible toolset puts more interactive real-time modeling power in the modeler's hand. Creator simulates real-time rendering techniques for true WYSIWYG (what you see is what you get) interactive modeling and changes appear in real-time. Creator, also, offers the flexibility of modeling from multiple points of view (see Figures 2 and 4).

Vega Prime Software creates visual simulations, general visualization applications and delivers the infrastructure needed to rapidly develop and deploy accurate real-time 4-D software applications; module FX delivers the functionality required to simulate a wide variety of special effects in real-time applications. Effects can be predefined and adjusted by a number of visual attributes to customize the display, timing, triggering and performance characteristics of the effects within scenes. The marine module utilizing water surfaces such as: oceans or lakes for real-time simulation and provides the capability to easily add realistic and dynamic water surfaces. Marine also combines the realism of accurately synthesizing a dynamic ocean with the performance required for interactive real-time simulation and training (see Figures 3 and 4).

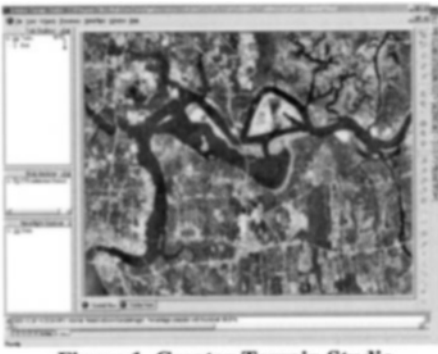


Figure 1. Creator Terrain Studio

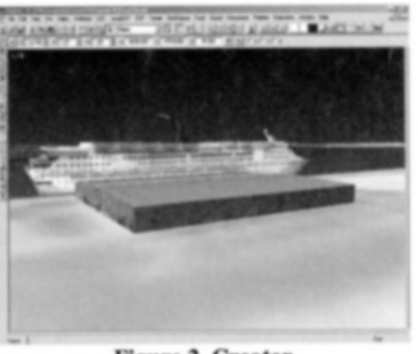


Figure 2. Creator

Clockwise: Landside, Port Gate, Underwater View, Night Scene

Figure 3. Vega Prime

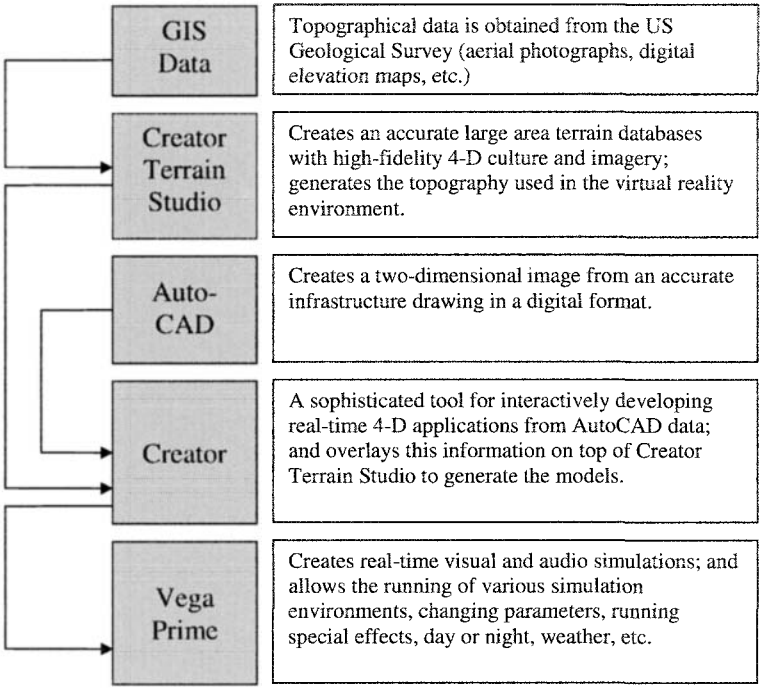


Figure 4. Simulation work flow

Applications to Maritime Port Safety and Security

The need to apply practical solutions for assuring safe and secure seaports in the U.S. is based on both social and economic considerations, and regulatory requirements. Our deepwater ports are a vital economic transportation link in the supply chain; and both national and international regulations (e.g., IMO, SOLAS/ISPS) require that safety and security standards, programs, implementation strategies and port security plans (approved by the Department of Homeland Security-U.S. Coast Guard) be in-place. This advanced simulation technology has many applications that make it an indispensable tool for this process. These practical tools include:

Scenario Development - The financial cost and logistical requirements for conducting live exercises to play out potential safety and security scenarios involving all the role players in a port environment places severe constraints on this approach. Electronic table top exercises that can be modified and replicated for training and certification purposes are very compatible with this advanced simulation; and this approach is becoming increasingly popular since all variables of weather, time-of-day, location, port activity; methods of terrorism and prevention are possible in a timely manner. Changes and modifications can be easily facilitated, including the introduction of unanticipated events

to examine personnel responses, related to emergency or incident response scenarios (see Figure 5).

Logistical Optimization - All ports are in a competitive environment with the goal of optimizing their logistical operation to maximize their revenue potential, while also addressing safety and security. This technology can accurately visualize the entire operation of the port and assists management in optimizing critical logistical issues such as: queuing, headways, dwell times, passenger and commodity flows, port scheduling, docking and berth assignments, and access and control. Timing, safety and security is of the essence when for example on Saturdays at Port Everglades five cruise ships must dock discharge passengers, crew, luggage, supplies, and clean, maintenance and restock, board and prepare for departure.

Planning and Engineering - Master planning and engineering for a maritime facility requires a new approach that departs from physical printed documents and maps. The desirable alternative is electronic interactive visual plans that allow all parties to fly-through the scenes, and venues thereby making modifications and producing a consensus-based virtual approach. Site planning and design must be carefully executed to maintain the safe and efficient flow of people, goods and services utilizing all the transport modes and infrastructure. To properly design and operate gates (handling nearly 50,000 vehicle movements a day) requires Port Everglades to examine alternative designs. Methods of primary and secondary inspection of trucks can be simulated and designed.

Environmental - Ports need to continuously analyze environmental impact, assessment and control. The interface of air, land and sea based activities require that the ecosystem and the environmental setting be compatible with the local conditions and standards in terms of health, safety and general welfare. Fuel storage, container protection and surveillance, lighting, acoustics, visibility, all must be considered and reviewed continuously. All eye-points, views and perspectives need to be environmentally examined as well as at intermodal connecting points coupled with safety and security. This environmental perspective involves aerial, surface and subsurface assessments for identifying problems and creating effective solutions.

Technology Applications - Every port is continuously investigating methods and technologies that have the potential for addressing safety and security issues. Virtual simulation is a method whereby some of these potential technologies (e.g., laser scanning, computer activated cameras, geo-fencing and optical arrays) can be put into the ports digital scene to assess their possible benefit, including line of sight, communications integration, and response time. Several U.S. ports are participating in a PILOT (Performance Integration Logistical Operations Test) where a parallel resource for intelligence surveillance monitoring (PRISM) is being used to invasively monitor containers using Immarsat, GPS, and IFR. Monitored on a 24/7 basis whether at sea, in a port or traveling to a distribution warehouse by road or rail. The container is under constant systematic passive surveillance evaluating such parameters as: temperature, shock, O_2 and CO, and container penetration. Once the field test, (involving approximately 15 containers and five ships serving Caribbean and Latin American ports)

is completed and evaluated, there is a world-wide application to further enhance intermodal safety and security. Simulation is being used for assisting the assessment process of this trial.

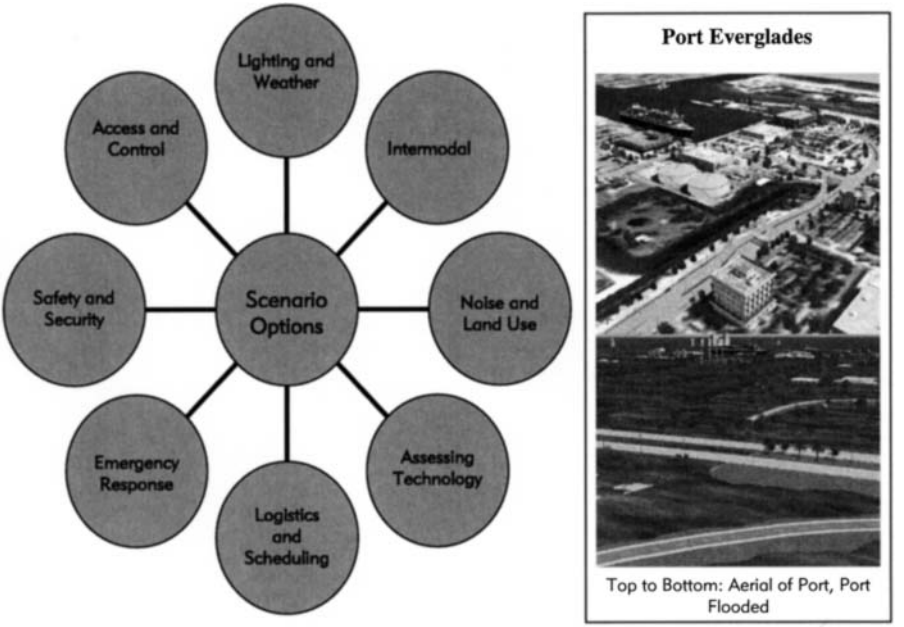


Figure 5. Scenario Options

References

1. Bragdon, C., and Morgan, S. (2004). *New visualization technologies for port security*, Cargo Security International, Vol. 2, No. 6.
2. Collins, S. (2006). *Proceedings of the American Association of Port Authorities*, Spring Conference, Alexandria, VA.
3. GAO. (2004). *Maritime security, substantial work remains to translate new planning requirements into effective port security*, Washington, DC.
4. Intermodal Association of North America. (2006). *International traffic activity, industry statistics*, Washington D.C.
5. Kean, T. (2004). *9/11 Commission report: Final report of the National Commission on Terrorist Attacks Upon the United States*, Washington D.C.
6. Lake, J., Robinson, W., and Seshetti, L. (2005). *Border and transportation security*, Congress Research Service, Library of Congress, Washington, DC.
7. MultiGen-Paradigm. (2005). *Real-time application development*, San Jose, CA.

Parameter Calibration for VISSIM Using A Hybrid Heuristic Algorithm: A Case Study of a Congested Traffic Network in China

Haode Liu, Xiaoguang Yang, Jian Sun

Haode Liu, Ph.D. Candidate, Research Interest is traffic simulation. School of Transportation Engineering, Tongji University, Siping Road 1239, Shanghai, P. R. China, 200092; Email: beuture@hotmail.com. Tel (Fax): (86-21) 6598-1409;

Xiaoguang Yang, Professor, Tongji University. Main research is transportation systemic engineering and theory of Intelligent Transportation Systems (ITS).

Jian Sun, Ph.D, Candidate, Tongji University. Research Interest is traffic system analysis.

ABSTRACT

This paper presents an application of a hybrid heuristic algorithm as an optimization method to find a suitable combination of micro-simulation parameters. The calibration is based on field data collected on weekday peak time of an urban downtown road network in Hefei, China. The tests indicated that the hybrid algorithm can well combine the advantage of Genetic Algorithm (GA) and Simulated Annealing (SA) algorithm. It can enhance the efficiency of convergence, and effectively reduce the likelihood of premature and converge to local optima. Meanwhile, eight important parameters and their proposed values were found which will affect the simulation accuracy specially when modeling the local peak time traffic condition.

INTRODUCTION

Now more and more people treat simulation as an important evaluation tool, and the essential issue is its reliability. Microscopic simulation models contain numerous independent parameters describing the traffic-flow characteristics: driver behavior and traffic-control parameters. These models provide a default value for each parameter, and also allow users to change the values according to the real traffic conditions.

Several researchers have discussed the general requirements of a simulation model calibration procedure (Hellinga, 1998; Lianyu Chu etc., 2003; Cohen, 2004). A formal procedure and an automated optimized method for conducting calibration are proposed in these literatures. Park and Schneeberger (2003) introduced a regression method to find candidate combinations of parameters to find reasonable sets through evaluation of travel time and visualization. As far as heuristic algorithm is concerned, Cheu etc., (1998), Lee (2001) and Ma (2002) have established integrated procedure and algorithm for parameters calibration. Sun etc. (2005) has developed Simulate Annealing (SA) algorithm integrated with VISSIM to conduct parameters calibration. GA and SA are virtual algorithm for global optimum searching. However, when using GA we had to confront the problem of premature (Ma, 2002); meanwhile, a drawback of SA is its dependence on initial population (Gao, 2003). For GA is a search strategy well suited for parallel computing while SA has weak

power to exploit parallelism (Wang, 2004).
The main contribution of this paper is to present an application of a hybrid heuristic algorithm combined GA and SA as an optimization method to find suitable combination of VISSIM parameter values. The proposed algorithm was evaluated and demonstrated via a case study of a congested traffic network in china.

TEST SITE AND DATA COLLECTION

The site and time period chosen for the study is the round range of Big East Gate in Hefei, China, between 7:30 am and 8:30 am. This is a 1-mile urban road segment sustaining heavy congestion during the morning peak hour.
Original traffic data were collected through License Plate Survey and loop detector. Note that data from License Plate Survey may not be integrated and accurate, we should integrate them with loop detector data. Subsequently, these data were fused via flow fuzzy model in VISSUM. Accurate OD matrices were estimated as inputs of VISSIM.

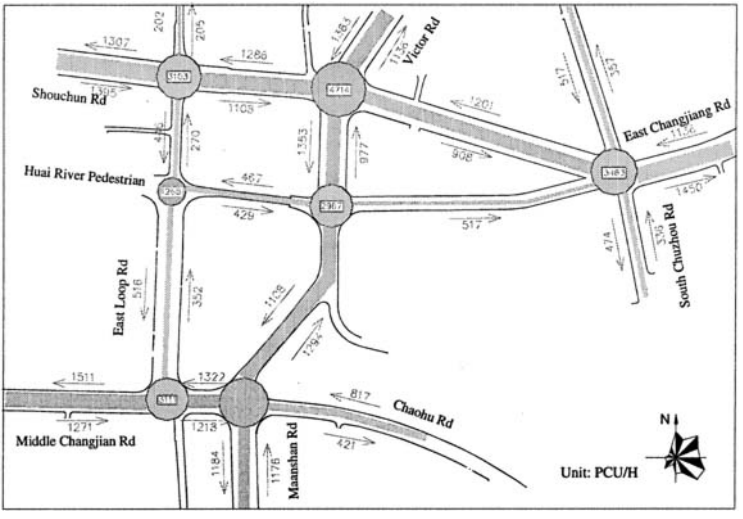


FIGURE 1 Distribution of traffic volume.

There are seven intersections in the test area. Control and management scheme of each intersection is required for micro-simulation, limited to the paper size here we didn't list the figures of control and management scheme.
To conduct this research, two groups data of different weekday were collected: one group is used for the parameter calibration, the other was used for validation.

GENETIC SIMULATED ANNEALING (GSA) ALGORITHM

Parameter calibration of a microscopic simulation model may be regarded as a combined optimization problem. To improve the performance of GA and SA, we apply a hybrid heuristic algorithm named GSA algorithm in our paper, which combines the local stochastic hill climbing SA and the global crossover and mutation operating GA.
GSA algorithm use SA to create new solutions while iterating: not only the best solution was

accepted, but also the "bad" solutions were accepted with a defined probability. This method can reduce the chances of premature, meanwhile, the fast convergent ability of GA and the jump-out from local optima ability of SA were preserved, so GSA is more efficient in optima searching.

PARAMETER CALIBRATION USING GSA ALGORITHM

VISSIM Simulation Model

VISSIM was chosen in this study, which is a discrete, stochastic, microscopic, time step and behavior based simulation model. What influence the accuracy of a traffic simulation model is the quality of the actual modeling of vehicles and the methodology of moving vehicles through the network. VISSIM uses the psycho-physical driver behavior model developed by Wiedemann.

Identification of Calibration Parameters

In micro-simulation models, the magnitude of effect on driving behavior and network performance coming from different parameters are significantly different. Two criteria potentially help users to select the parameters candidates: (1) conduct trial and error test for each parameter; (2) empirical judgment from a traffic engineer's point of view. In this research, through lots of simulation tests we select the eight most sensitive parameters for calibration: ESD, LCD, WTBD, MH, NOV, ASD, APSD, MPSD.
The final acceptable ranges of the selected parameters are listed in Table 1, for each control parameter we select four typical different values, so, there are 4^8(i.e., 65,536) possible combinations. We list the four typical value of each parameter in Table1.

TABLE 1 Candidate and Default Parameter Sets

Parameter	Value 1	Value 2	Value 3	Value 4	Default
Emergency Stop Distance (ESD)	2	3	4	5	5
Lane Change Distance (LCD)	150	175	200	225	200
Waiting Time Before Diffusion (WTBD)	45	60	75	90	60
Min Headway (MH)	0.5	0.75	1.0	1.25	0.5
Number of Observed Vehicles (NOV)	1	2	3	4	2
Average Standstill Distance (ASD)	1.0	1.5	2.0	2.5	2.
Additive Part of Safety Distance (APSD)	1.0	1.5	2.0	2.5	2.0
Multipliable Part of Safety Distance (MPSD)	2	2.5	3	3.5	3

Multiple Runs

VISSIM runs with identical input files and random seeds generate identical results. Using a different random seed changes the profile of the traffic arriving (stochastic variation of input flow arrival times) and therefore may come up different results. In this research, five random seed runs were conducted in VISSIM for each of the iteration cases, the average values of each output were calculated, and then were input into an objective function to check their convergence.

Objective Functions

The objective functions are to minimize the deviation between the observed and the

corresponding simulated values. They can be called travel time match. In this paper, "Mean Absolute Error Ratio" was adopted as objective function as follows:

$$g(X_i^g) = AAER(X_i^g) = \frac{\sum_{l=1}^{M}\left(\frac{\left|T_l^{Obs} - T_{il}^{Sim}\right|}{T_l^{Obs}}\right)}{M} \quad \text{(Eq.1)}$$

Where g = Number of generation,

i = Chromosome,

X_i^g =Parameter set of chromosome i at generation g,

$g(X_i^g)$ = Fitness function for a parameter set of chromosome i and generation g,

T_l^{Obs} =Observed travel time for link l,

T_{il}^{Sim} =Simulated travel time (from simulation run i) for link l, and

M =Number of links in network.

The dependent variable is the MAER as shown in Equation 2.

$$F(X_i^g) = f(g(X_i^g)) = C\exp(-Dg(X_i^g)) \quad \text{(Eq.2)}$$

Where F =Fitness function values,

C, D =Constants, and $g(X_i^g)$ =Mean Absolute Error Ratio for chromosome i calculated using Equation 1.

While conduct simulated annealing, new populations were accepted with the probability of $\min\{1, e^{-\Delta f/T}\}$, in which Δf is the cost reduction, it is the difference of f_i and f_{i-1}. f is calculated with Equation *1*.

Stopping Rule

The whole process is repeated until the convergence criterion is satisfied. In the case, when the difference between each simulated travel time of a link and its corresponding observed travel time are smaller than 15 percent of the latter, we consider that the solution is the global best-so-far solution. In addition, a maximum number of iterations or generations G, is identified as stop rule before we attain a satisfying solutions.

MODEL CALIBRATION RESULTS AND DISCUSSION

Visualization and Parameter Set Selection

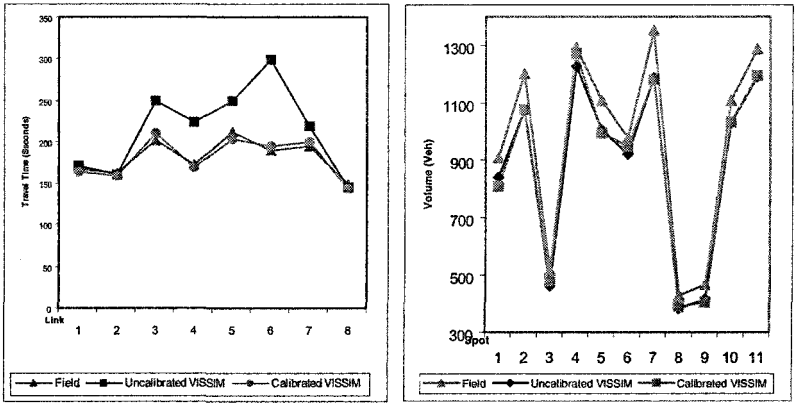
The purpose of the microscopic simulation is to represent the real condition close as possible. A model cannot be considered calibrated if the animations are not real. For each set of calibration parameter solution, we run VISSIM again to observe its animation, and discarded those parameter combinations whose VISSIM animation was not realistic.

In this research, the best combination value of ESD, LCD, WTBD, MH, NOV, ASD, APSD and MPSD for local traffic condition is 3.0, 175, 45, 0.5, 3, 1.0, 1.5 and 3.0 respectively.

Validation With Data

After calibration, we use validation data to validate the model with the above parameter value. The average run results of travel time and volume was used for the un-calibrated (default) and calibrated VISSIM, and the statistical result was compared with the field data. As shown in the figure 2, travel time of the calibrated model show much better with the field data than the un-calibrated model. The comparison of the field data, the un-calibrated VISSIM model, and the calibrated VISSIM model shows the necessary of calibration and validation for microscopic simulation models.

The un-calibrated model has longer travel time than that observed in the field. Which indicates that the driving behavior of China are more aggressive than the default VISSIM parameters, and for more than 80 percent of travelers during this peak time are commuters, they are more familiar with the network than the VISSIM model and usually choose optimal route and driving behavior to reach their destination. Figure 2 (b)



(a)Travel times comparision (b)Volume Comparison

FIGURE 2 Travel times comparison of field data, un-calibrated and calibrated VISSIM.

verified this conclusion: while the volume of the calibrated model is close to the field data, the volume of the un-calibrated model matches the field data much better than the calibrated data.

CONCLUSIONS

This paper proposed an application of a hybrid heuristic algorithm with GA and SA as an optimization method to find a suitable combination of micro-simulation parameters value. In this research area, we find that the main parameters affecting simulation accuracy are ESD, DLCD, WTBD, MH, NOV, ASD, APSD and MPSD, and the proposed value of them are 3.0, 175, 45, 0.5, 3, 1.0, 1.5 and 3.0 respectively. The important results are summarized as follows:

(1) The calibrated VISSIM model reasonably reproduced the real traffic flow on the test

site. Through validation, the measured values are in closely matched with the predicted values obtained from calibrated VISSIM model by GSA algorithm.

(2) In general, simulation results and the calibrated parameter values indicate that drivers in China are more aggressive in urban road during peak time compared with the default values in VISSIM.

(3) The hybrid algorithm can combine the advantage of GA and SA: it can not only enhance the efficiency of convergence but also effectively reduce the probability of premature and converge to local optimum.

In the network level model calibration process, the inter-relationship between route choice and OD estimation make the problem more complicated. This will be a topic for further study in the micro-simulation calibration process.

REFERENCES

Cheu, R.L., X. Jin, K.C. Ng, Y.L. Ng, and D. Srinivasan. Calibration of FRESIM for Singapore Expressway Using Genetic Algorithm. ASCE Journal of Transportation Engineering. 1998, 526-535.

Cohen, S. L. An Approach to Calibration and Validation of Traffic Simulation Models. 83rd Annual Meeting Preprint CD-ROM, Transportation Research Board, Washington, D.C., 2004

Gao Shang. Solving TSP with Simulated Annealing Algorithm. Journal of East China Shipbuilding Institute, v 17, n 3, June, 2003, 13-18.

Hellinga, B. R. Requirement for the Calibration of Traffic Simulation Models. Department of Civil Engineering, University of Waterloo. http://www.civil.uwaterloo.ca/bhellinga/Publications%20Page/Publications/CSCE-1998-Calibration.PDF. Accessed July 2005.

Jian Sun, Zhizhou Wu. Calibration of VISSIM for Shanghai Expressway Weaving Sections Using Simulated Annealing Algorithm. 2005. International Conference on Computing in Civil Engineering. Ducan Mexico. 2005.

Lee, D., Xu Yang, P. Chandrasekar. Parameter Calibration for PARAMICS Using Genetic Algorithm. Presented at 80th Annual Meeting of the Transportation Research Board, Washington, D.C., 2001.

Ma Tao, Abdulhai, Baher. Genetic algorithm-based optimization approach and genetic tool for calibrating traffic microscopic simulation parameters. Transportation Research Record 1800, 2002, 6-15.

Lianyu Chu, Liu, H.X., Jun-Seok Oh, and Recker W. A calibration Procedure for Microscopic Traffic Simulation. 83rd Annual Meeting Preprint CD-ROM, Transportation Research Board, Washington, D.C., 2004

Park, B., and J. D. Schneeberger. Microscopic Simulation Model Calibration and Validation: A Case Study of VISSIM for a Coordinated Actuated Signal System. Transportation Research Record 1856, TRB, National Research Council, Washington, D.C., 2003, 185-192.

Wang, Z.G., Wong, Y.S., Rahman, M. International Journal of Advanced Manufacturing Technology, v 24, n 9-10, 2004, 727-32.

Topological and Operational Improvements to a Cell-Transmission-Based Simulation Model

Sherif Ishak[1], Ciprian Alecsandru[2], and Dan Seedah[3]

[1]Assistant Professor, [2]Ph.D. Candidate, [3]Undergraduate Student, Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA, 70803 sishak@lsu.edu, Ph: (225) 578-4846, Fax: (225) 578-8652

ABSTRACT

The cell transmission model (CTM) developed by Daganzo in 1994, represents a reliable and efficient simulation environment mostly for transportation planning applications. This research study demonstrates that specific improvements can convert CTM into an operations model for large-scale traffic networks. These improvements include modification to allow for a variable cell length and adjustments to the flow advancing algorithms for a better representation of traffic flow. Evaluation and comparison of the improved version of CTM versus CORSIM has been performed using a freeway network of I-10 corridor in Baton Rouge. In addition, a sensitivity analysis demonstrates the model's good performance under the improvements developed in this study.

Keywords
Macroscopic Simulation, Mesoscopic Simulation, Cell Transmission

INTRODUCTION AND OBJECTVES

Over the last two decades, there has been continuous development and enhancement of traffic network modeling/simulation tools. One of the most recent models, which provide reasonable approximation of traffic evolution in realistic networks, is the cell-transmission model (CTM) developed by Daganzo (*1*, *2*). CTM is a macroscopic model derived from an approximation of the Lighthill and Whitham (3) and Richards (4) (LWR) model. CTM was adequately validated with field data by Lin and Daganzo (5), and Lin and Ahanotu (*6*). Several authors used CTM for dynamic traffic assignment (see for instance, *7, 8, 9*), network design applications (*10*, *11*), and travel time prediction (*12*). As stated by its author, CTM in its original form is more suitable for planning than operational analysis of traffic networks. Because of its macroscopic / mesoscopic characteristics CTM might be successfully use in operational analysis of large-scale traffic networks, while retaining computational and calibration advantages over microscopic models. Nevertheless, to improve its accuracy in traffic flow representation certain improvements or enhancements may be required as presented in this study.

The objective of this study is to present specific improvements to the original form of CTM and investigate their effect on the model performance. Certain assumptions in the original form of CTM limit the applicability of the model at the operational analysis level. The original model requires that the simulation time step be equal to the free-flow travel time of each cell; that is the cell length divided by the free-flow speed of that cell. Consequently, this limits the flexibility of arbitrary selection of cell lengths, which are dictated by the roadway geometry. One solution to overcome this limitation is to split the network links into shorter cells to reduce the approximation errors. However, this may also result in increasing the number of cells in the network unnecessarily. Ziliaskopoulos and Lee (*13*) proposed that long homogeneous cells could be divided into small imaginary cells (subcells) to maintain a constant simulation time step without increasing the number of cells. However, the approach requires that large cells be divided into equal-size subcells. Therefore, in this study more flexibility has been added to relax the constraint on the selection of cell sizes.

A second improvement to the original CTM is to modify the flow advancing algorithm to account for a better representation of traffic flow at network junctions level. Ziliaskopoulos and Lee (*13*) showed that the original CTM might violate the FIFO (first-in, first-out) principle at the network diverging junctions, when traffic congestion propagates upstream the diverging location. Therefore, special treatment of flow advancing has been implemented such that accounts for FIFO constraint under all geometric configurations. These add-ons have been integrated into a newly developed software module for evaluating the model performance.

BACKGROUND

CTM offers a solution for the LWR model using a finite difference equations (FDE) method. The building block of CTM is the *cell*, which represents a homogeneous segment of the highway. Each link in a traffic network is divided into one or more interconnected cells. Two adjacent cells may be interlinked with one connector only, which has no physical dimensions, but is a simple representation of the interface between cells. The cell length determines the simulation time step, which is equal to the free-flow travel time to traverse the cell. This ensures that within each simulation time step, vehicles cannot skip cells. Each cell/connector is classified into one of five distinct types: ordinary, merging, diverging, source, or sink. Detailed description of CTM network topology may be found in (*1*, *2* or *7*).

CTM IMPLEMENTATION

CTM software module was developed using an object-oriented programming language (Visual Basic .NET). For each time step, the simulation is executed in three stages. In stage 1, the number of vehicles to advance is determined and vehicle platoons proceed out of each cell onto the outgoing connectors. In stage 2, the simulation proceeds with movements of vehicle platoons between subcells within large cells. Finally, stage 3 advances vehicles from the connectors into the receiving cells.

Treatment of Ordinary Cells

The original CTM flow advancing equation (1) does not enforce the FIFO principle under all geometric and operational conditions. Therefore, a modified formulation is expressed by:

$$y_{ij} = \min\{x_i^s, Q_i, Q_j, \delta_j x_{j,n_j}^r\} \quad (2)$$

This equation differs from its original CTM counterpart by the first and the last term. x_i^s is the number of vehicles eligible to advance out of cell i. x_i^s ensures that vehicles must spend at least τ_i time units, the free-flow travel time in cell i, before proceeding to the next cell. Q_i and Q_j the flow capacities with respect to the simulation time step out of cell i and into cell j, respectively. δ_j is set to W_j / V_j for congested conditions (if $x_i > \min\{Q_i, Q_j\}$) and to 1 otherwise. V_j and W_j are the free-flow speed and backward moving wave speed in cell j, respectively. x_{j,n_j}^r is the usable space capacity in the last subcell of cell j.

The FIFO principle is guaranteed by storing the order in which the vehicles entered the cell and using this information when they are processed out of the cell. FIGURE 1 shows an example of how vehicle platoons are represented inside the cells.

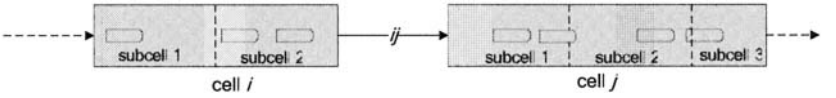


FIGURE 1: An example of vehicle platoons representation inside the cells

The simulation time step, τ, may be arbitrarily selected such that $\tau \le \tau_i \forall i \in \mathcal{C}$; where $\mathcal{C}$ is the set of all the cells in the network, and τ_i is the cell i free-flow travel time. Because τ also determines the size of subcells in large cells, the number of subcells, n_j, is determined by $ceiling(\tau_j / \tau)$. Since n_j is rounded up to the nearest integer, the size of the *front* subcell, τ_j^*, may be $\le \tau$, assuming all other subcells of equal size τ. This allows arbitrary selection of cell length. For instance, a 700-ft cell may be divided into two 300-ft subcells and a front 100-ft subcell. Special adjustment for shorter front subcells is implemented with the internal movements in stage 2. x_{j,n_j}^r is the usable space capacity in the last subcell of cell j. y_{ij} is the number of vehicles advancing in one platoon onto connector (i, j).

Treatment of Merging Cells

Advancing vehicles from two or more cells into one merging cell must account for the maximum to be sent $S_{\max} = \sum_{i \in \Gamma^{-1}(j)} \min\{x_i^s, Q_i\}$, where $\Gamma^{-1}(j)$ is the set of predecessor cells of cell j, as well as the maximum that can be received $R_{\max} = \min\{Q_j, \delta_j x_j^r\}$ by

the merging cell j. This flow advancing logic insures that the actual merging ratios under light traffic conditions ($S_{max} \leq R_{max}$) are determined by the flows out of each sending cell. For $S_{max} > R_{max}$, congestion is expected on at least one of the merging connectors, and thus, merging priorities P_{ij} must be clearly specified depending on the hierarchies of the sending cells. Such ratios are only effective if all merging connectors are congested. If only some merging connectors are congested then any unused capacity in the merging cell may be allocated to the congested connectors. This implementation for the merging logic allows for maximizing the utility of the available spacing in the merging cells with respect to the merging priorities. Therefore, a more realistic representation of network traffic flow is achieved, as opposed to the original CTM.

Treatment of Diverging Cells

Special treatment when moving flows out of the diverging cells is needed, because vehicles must decide on which diverging connector to choose on the path to their final destination. A more realistic approach for selection of a target connector should consider some route choice model (e.g. shortest path), but this is beyond the scope of this paper. At this implementation phase, simple diverging ratios are assumed to be known a priori. The diverging logic was improved to account for possible blockage on one or more of the diverging connectors, and the subsequent restraint of flow in the corresponding lanes of the diverging cell. Therefore, as vehicles advance onto diverging connectors, the first connector that gets blocked triggers a temporary blockage of as many lanes in the diverging cell as in the receiving congested cell. One advantage of this approach is that it allows for blocking of shared lanes leading to a blocked connector. However, lanes that remain open may still be used to advance vehicles to unblocked connectors. This treatment resembles more closely the observed real life driving behavior.

Advancing Vehicles between Subcells

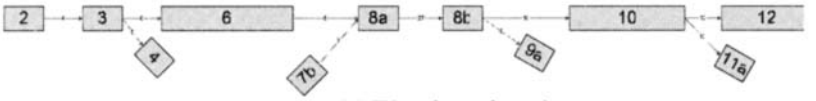
In stage 2, the platoons are allowed to advance internally within large cells from one subcell into another. The internal advancing process applies the basic form of the flow advancing equation for ordinary connectors. However, for diverging cells a special treatment is applied to account for vehicles destination. Unfortunately, due to the limitation on the paper length, details on such treatment are omitted, but can be found in (*14*).

Advancing of Vehicles into Cells

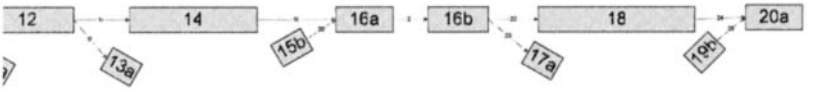
During stage 3 vehicles advance from each connector into the receiving cell in one platoon. For ordinary and diverging cells, the process is straightforward since only one platoon advances into the cell. For merging cells, platoons advance simultaneously into the merging cell from two or more connectors. The developed approach constructs the final platoon by selecting vehicles from merging platoons randomly, weighted by the merging flows priorities. Statistics on selected performance measures such as cell occupancy, travel time, and others is collected at the end of stage 3.

PERFORMANCE EVALUATION FOR A FREEWAY NETWORK

In order to evaluate the improvements introduced in CTM a freeway network was selected from the I-10 corridor in Baton Rouge. The selected network includes 4 off-ramps and 4 on-ramps, as sketched in FIGURE 2 (a, b).



(a) West-bound section



(b) East- bound section

Source Cell	Vehicle Count Time Period = 1	Vehicle Count Time Period = 2
SRC1	1500	750
SRC_7a	300	150
SRC_15a	100	50
SRC_19a	1000	500

(c) Demand

FIGURE 2: Freeway Network Topology for CTM

The tested network consists of 15 ordinary cells, 3 merging cells, and 5 diverging cells. The network also has 4 source cells, where traffic enters, and 6 sink cells, where traffic exits the network. The cell length was chosen to retain geometric homogeneity while minimizing the number of cells. A 2-hour simulation period was divided into three intervals. The traffic demand for the first two periods (30 minutes each) is detailed in FIGURE 2 (c). The network is clearing up all the queues and vehicles during the third period (60 minutes). Congestion conditions have been created by reducing the speed limit in cell 18 to 10 mph throughout the entire simulation period. The demand flow rates were selected such that queues do not spill back to block the source cells. Merging priorities were set to 50% on all merging connectors, while diverging ratios for off-ramps were assumed 10% of the mainline flow. The free-flow and wave speeds for all cells were assumed 60 mph and 30 mph, respectively. The cell flow capacity was assumed 2250 pcphpl and the jam density 214 pcpmpl. Based on the shortest cell (700') in the network, the simulation time step was set to 8 seconds.

SIMULATION RESULTS

A similar network has been simulated using CORSIM. The simulation of each network has been repeated five times using different random seeds, and the average of the five runs was compared. The average total network travel times for CTM and CORSIM were 293.1 and 299.86 vehicle hours, respectively. The 2.25% difference

in travel time may be attributed to the random variations in driving behavior embedded in car-following and lane-changing models used in CORSIM. Given 4350 of simulated vehicles, the average error in travel time is nearly 5.6 seconds per vehicle for an average trip time of 248 seconds per vehicle. Hence, despite its macroscopic nature, CTM provides a good representation of traffic flow when compared to microscopic simulation.

A sensitivity analysis has been performed in order to evaluate the robustness of the modified model. Two performance measures have been investigated, the total network travel time and the simulation run time. The total network travel time has been defined as the total time spent by all the vehicles in the network in vehicle-hours. By adjusting the size of the subcells inside the long cells, one can achieve a certain degree of network granularity. Hence, the more subcells the more interaction is captured by the model, and presumably a more accurate representation of the traffic flow. Therefore, the shortest subcell length, which also yields the shortest time step (2 seconds), is considered the reference (base) case for the analysis depicted in FIGURE 3. As the time step increases the total network travel time also increases (due to less accurate, larger approximations of the vehicle flows). However, the relative error does not exceed 5.6%, as the time step increases to four time times the base case. It was observed that, as the time step increases (i.e. the total number of subcells decreases) the simulation run time drops off, up to 70% less than the reference case. Therefore, depending on the desired level of accuracy one could strike a balance between the simulation run time and the model accuracy, by choosing an appropriate simulation time step

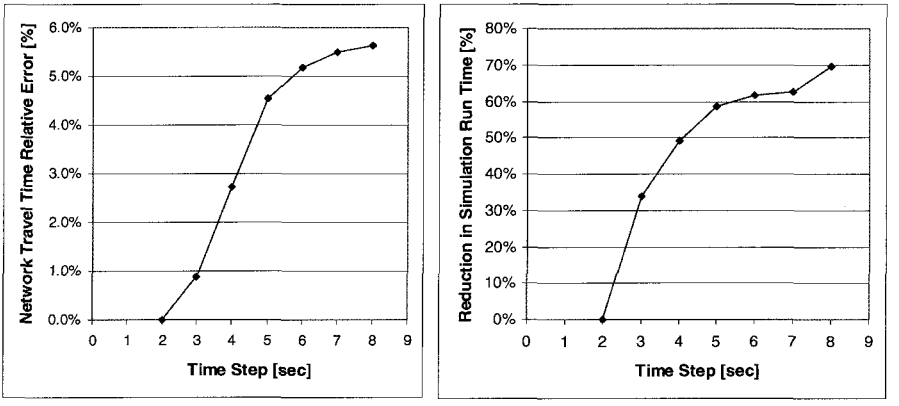


FIGURE 3: Model sensitivity analysis

CONCLUDING REMARKS

This paper summarized specific improvements in the development and implementation of a traffic simulation model derived from the cell transmission model by Daganzo (*1*, *2*). Due to its macroscopic nature, CTM provides a reasonably accurate and computationally efficient tool for modeling large-scale networks. The

developed improvements aim at improving both flexibility and realism of the original CTM in traffic flow and network representations. Such changes enabled (1) arbitrary selection of cell lengths to conform to geometric characteristics and (2) more realistic treatment of traffic behavior in both merging and diverging junctions. All improvements are deemed necessary to enhance CTM's ability to support operational analysis of traffic networks, while retaining its macroscopic characteristics and relative ease of calibration when compared to microscopic simulation models.

Using a segment of I-10 freeway network, the performance of CTM after modifications was compared to that of CORSIM, a microscopic simulation platform. Link by link comparisons showed very similar results for both platforms over time, despite the difference in level of simulation and sources of stochastic variations. While microscopic in nature, variations in CORSIM results may be mostly attributed to random variations in driving behavior, as described by car-following and lane-changing models. On the other hand, variations in CTM results are attributed to random variations in both merging, where platoons merge, and diverging, where platoons split by target cell. Regardless of source of variations, both platforms produced very comparable results, which reflect the potential of CTM for development and implementation as an operations model for large-scale networks. This requires, however, additional modifications to allow more accurate representations of traffic flow, on hybrid networks of both interrupted flows (e.g. signalized intersections) and uninterrupted flows (freeways). Research is currently underway to expand the applicability of CTM to modeling pre-timed and actuated signalized intersections, as well as other network components such as toll plazas and unsignalized intersections.

REFERENCES

1. Daganzo, C. F. (1994). "The Cell Transmission Model: A Simple Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory," Transportation Research 28B, 269–287.
2. Daganzo, C. F. (1995). "The Cell Transmission Model. II. Network Traffic," Transportation Research 29B, 79–93.
3. Lighthill, M. J. and J. B. Whitham (1955). "On Kinematic Waves. I. Flow Modeling in Long Rivers. II. A Theory of Traffic Flow on Long Crowded Roads," Proc. Roy. Soc. A, 229, 281–345.
4. Richards, P.I. (1956). "Shockwaves on the Highway," Operations Research 4, 42–51.
5. Lin, W. H. and C.F. Daganzo (1994). "Technical description of NETCELL: General framework and data structure," PATH Technical Note 94-7. Institute of Transportation Studies and Department of Civil Engineering, University of California, Berkeley, CA.
6. Lin, W. H. and D. Ahanotu (1994). "Validating the basic cell transmission model on a single freeway link," PATH technical report 95-3. Institute of Transportation Studies and Department of Civil Engineering, University of California, Berkeley, CA.

7. Ziliaskopoulos, A. K. (2000). "A Linear Programming Model for the Single Destination System Optimum Dynamic Traffic Assignment Problem," Transportation Science, Vol. 34, No. 1, 37–49.
8. Li, Y., A. K. Ziliaskopoulos and T. Waller (1999). "Linear Programming Formulations for System Optimum Dynamic Traffic Assignment with Arrival Time-Based and Departure Time-Based Demands," Transportation Research Record, No. 1667, 52-59.
9. Golani, H. and S. T. Waller (2004). "A Combinatorial Approach for Multiple-Destination User Optimal Dynamic Traffic Assignment," Paper presented at the 83rd TRB Annual Meeting, Washington, D.C.
10. Waller, S. T. and A. K. Ziliaskopoulos (2001). "Stochastic Dynamic Network Design Problem," Transportation Research Record, No. 1771, 106-113.
11. Ukkusuri, S. and T. Waller (2004). "Linear Programming Models for the User and System Optimal Dynamic Network Design Problem: Formulations, Implementations and Comparisons," Paper presented at the 83rd TRB Annual Meeting, Washington, D.C.
12. Mark, C. D. and A. W. Sadek (2004). "Learning Systems for Predicting Experience Travel Times in The Presence Of Incidents: Insights and Lessons Learned", Paper presented at the 83rd TRB Annual Meeting, Washington, D.C.
13. Ziliaskopoulos, A. K. and S. Lee (1997). "A Cell Transmission Based Assignment Simulation Model for Integrated Freeway/Surface Street Systems," Paper presented at the 76th TRB Annual Meeting, Washington, D.C., 1997.
14. Ishak, S., C. Alecsandru, and D. Seedah (2006) "Improvement and Evaluation of the Cell-Transmission Model for Operational Analysis of Traffic Networks: A Freeway Case Study" Paper to be presented at the 85th TRB Annual Meeting, Washington, D.C.

Application of Wavelet Time Series Analysis in Short-Term Traffic Volume Prediction

Yuanchang Xie[1] and Yunlong Zhang[2]

[1]Zachry Department of Civil Engineering, Texas A&M University, 3136 TAMU, College Station, TX 77843; PH (979) 862-3553; FAX (979) 845-6481; email: yx40@tamu.edu

[2]Zachry Department of Civil Engineering, Texas A&M University, 3136 TAMU, College Station, TX 77843; PH (979) 845-9902; FAX (979) 845-6481; email:yzhang@civil.tamu.edu

Abstract

Short-term traffic volume forecasting is a key component of many intelligent transportation and traffic control systems. The Autoregressive Integrated Moving Average (ARIMA) models have been widely studied and used for short-term traffic volume prediction by many researchers. However, the fluctuations due to local noise make it difficult for ARIMA models to obtain high prediction accuracy. This paper investigates the use of Discrete Wavelet Transform (DWT) to improve the performance of ARIMA models for short-term traffic volume prediction. We first use the DWT to denoise the original traffic volume data such that less fluctuating data are obtained. An ARIMA forecasting model is then fitted based on the denoised data series. Real-world data from Interstate 80 in California are used to test and compare this proposed wavelet ARIMA approach with ARIMA models.

Keywords: ARIMA models, Short-term traffic forecasting, Discrete wavelet transform

Introduction

Since the early 1970s, many approaches have been applied to the short-term traffic volume forecasting problem. These approaches mainly include neural networks (Jiang, 2005), time series models such as ARIMA (Ahmed, 1979; Nihan, 1980; Hamed, 1995; Van Der Voort, 1996; Willams, 1998; Smith, 2002), spectral analysis (Nicholson, 1974), and Kalman filtering theory (Stathopoulos, 2003). In this study we will focus on the ARIMA models and the combination of ARIMA models and Discrete Wavelet Transform (DWT) analysis.

Ahmed and Cook (1979) used an ARIMA(0,1,3) model to analyze the 20-s and 60-s traffic volume data from Los Angeles, Minneapolis, and Detroit areas. Nihan and Holmesland applied an ARIMA(12,1,7) model to forecast the monthly average weekday traffic volume on a freeway segment and reported very good forecasting results. Hamed et al. (1995) fitted an ARIMA(0,1,1) model for the 1-minute traffic volume data from five arterials in one urban area. However, they used the same data sets for both model fitting and evaluation and didn't apply the model to predict any other data. Van Der Voort et al. (1996) developed a KARIMA model that combines the Kohonen neural networks and the ARIMA models. Williams et al. (1998) reported that the seasonal ARIMA model is better than the nearest neighbor,

neural networks, and the historical average models for single-step forecasting. Smith et al. (2002) also confirmed that the seasonal ARIMA model outperforms the nonparametric regression models based on their experiments using 15-minute traffic volume data.

Conejo et al. (2005) combined the Discrete Wavelet Transform (DWT) and ARIMA models to forecast the day-ahead electricity prices, and they reported that the new model consistently outperformed the direct use of ARIMA models. Given that the application of ARIMA models in traffic volume forecasting is very popular and has attracted so many attentions as described before, in this study we will further investigate the performance of Wavelet Autoregressive Integrated Moving Average (WARIMA) models in short-term traffic volume forecasting to see if WARIMA models are more robust and accurate than direct ARIMA models. In the following section, a brief introduction of WARIMA models is presented first. After that is the description of data collection and test design. Based on these data the WARIMA models and the ARIMA models are then tested and compared. Finally discussions and conclusions are given based on the test results.

Model Formulation

The standard form of an ARIMA(p,d,q) model is shown in Eq. (1), where p, d, and q are the number of autoregressive terms, nonseasonal differences, and lagged forecast errors respectively (Box, 1976).

$$\phi_p(B)(1-B)^d X_t = \theta_0 + \theta_q(B)e_t \tag{1}$$

Where $\phi_p(B) = (1-\phi_1 B - ... - \phi_p B^p)$;

$\theta_q(B) = (1-\theta_1 B - ... - \theta_q B^q)$;

$\theta_0 = \mu(1-\phi_1 B - ... - \phi_p)$;

μ, ϕ, θ = coefficients to be estimated;

B = backshift operator; and

X_t, e_t = original data series and white noise respectively.

Assume traffic volume data are recorded every 5 minutes, and the historical traffic volume data series is $\{X_{t-N}, ..., X_{t-2}, X_{t-1}, X_t\}$. If X_t is just observed and an ARIMA(1,1,1) model is fitted based on the available N+1 observed data, then Eq. (2) can be used to forecast the traffic volume in the next 5-minute interval, X_{t+1},

$$\hat{X}_{t+1} = \theta_0 + X_t + \phi_1(X_t - X_{t-1}) + e_{t+1} - \theta_1 e_t \tag{2}$$

This prediction process is repeated every time a newly observed value is added and is called a single-step forecasting. Sometimes it might be useful to update the model coefficients in Eq. (1) or even recalibrate the ARIMA models when new observed data are available at each time interval. But in this study the single-step ARIMA models without real-time model updating are used.

In WARIMA models, instead of applying the ARIMA models directly, a DWT preprocessor is applied to the original traffic volume data. In this way, the original data are decomposed into several approximate and detailed constitutive series. The resulted approximate data are more stable and have much less outliers compared with the original one, thus the WARIMA models fitted based on the approximate data may have better forecasting performance than the direct ARIMA models. The basic idea of DWT is to project a signal or data series $f(t) \in L^2(R)$ onto a set of wavelet and scaling functions as follow (Burrus, 1998).

$$f(t) = \sum_{k \in Z} c_{j0,k} 2^{-j0/2} \varphi(2^{-j0} t - k) + \sum_{k \in Z} \sum_{j=0}^{j0} d_{j,k} 2^{-j/2} \psi(2^{-j} t - k) \tag{3}$$

Where $j0$ is the coarsest level of decomposition; $\varphi(\cdot)$ is the scaling function; and $\psi(\cdot)$ is the wavelet function. $j0$ can be chosen arbitrarily within the range defined by the length of the discrete data series. The goal of the DWT analysis is to obtain the coefficients $c_{j0,k}$ and $d_{j,k}$ in Eq. (3), which in practice are often calculated using the filter-bank algorithm (Burrus, 1998). With these DWT coefficients, the reconstruction of the data at each resolution level can be accomplished based on Eqs. (4-6). For example, the detailed information at level i is

$$D(i) = \sum_{k \in Z} d_{i,k} 2^{-i/2} \psi(2^{-i} t - k), \; i = 1,2,..., j_0 \tag{4}$$

The approximate information at level i is

$$A(j_0) = \sum_{k \in Z} c_{j_0,k} 2^{-j_0/2} \varphi(2^{-j_0} t - k) \tag{5}$$

$$A(i) = A(j_0) + \sum_{m=i+1}^{j_0} D(m), \; i = 1,2,..., j_0 - 1 \tag{6}$$

And $A(0)$ represents the original data. Figure 1 shows a two-level DWT decomposition and reconstruction, where $A(1)=A(2)+D(2)$ and all the constitutive series, $A(1)$, $A(2)$, $D(1)$, and $D(2)$, have the same length as the original series.

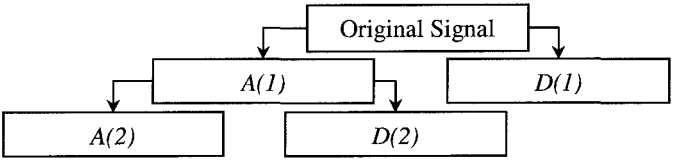


Figure 1. An example of a 2-level wavelet decomposition and reconstruction.

$A(1)$ is the finest approximate series, and $A(2)$ is the coarsest approximate series. $D(1)$ and $D(2)$ are the detailed data series. In this study, the approximate series will be used to fit the WARIMA models that are to be compared with the ARIMA models fitted based on the original date series.

Data Collection and Test Design

Data used in this study are provided by the Next Generation Simulation (NGSIM) program. These data were collected between Powell Street and Gilman Avenue on Interstate 80 in Emeryville, California. Twenty-four hour traffic volume data at loop detector stations 1, 3, 4, and 6 for December 1st~5th and December 8th~12th, 2003 are used. The original data are grouped to obtain the five-minute traffic volume data, which is appropriate for short-term traffic volume forecasting analysis.

The data collected from December 1st~4th and December 8th~11th at the four detector stations are used for fitting the WARIMA and ARIMA models, which are then applied to forecast the traffic volumes on December 5th and 12th respectively. Therefore, totally there are 8 data sets. The calibration and prediction processes are illustrated in Figure 2. The calibration process is to use SAS to find the best ARIMA(p,d,q) model. All feasible combinations of p, d, and q are tried and the one produces the lowest Mean Absolute Percentage Error (MAPE) is chosen. The solid lines represent the calibration and the dashed lines represent the prediction process. Three types of wavelets, Daubechies 4, 8, and 16, are used to test the ARIMA and WARIMA models. For each type of wavelet, four-level of DWT decomposition and reconstruction analysis is carried out to investigate the effects of decomposition levels on the forecasting accuracy.

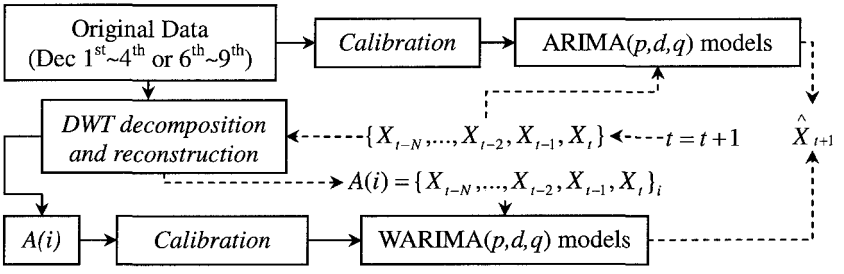


Figure 2. Calibration and prediction of ARIMA and WARIMA models.

Two commonly used indexes, MAPE and Variance of Absolute Percentage Error (VAPE), are used to measure the performance of the two models. The two indexes are defined in Eqs. (7) and (8). Where $\hat{vol}(k)$ is the predicted volume and $vol(k)$ is the measured volume at time interval k

$$MAPE = \frac{1}{N}\sum_{k=1}^{N}\left|\frac{\hat{vol}(k)-vol(k)}{vol(k)}\right| \tag{7}$$

$$VAPE = Variance\left(\left(\hat{vol}-vol\right)./vol\right) \tag{8}$$

Test Results Analysis

For all the 8 data sets, the average MAPE, VAPE, and standard deviation of prediction errors of the ARIMA models are 13.9%, 0.085, and 43 vehicles respectively. The standard deviation of prediction errors is an output from the SAS software and shows how reliable the prediction is. The corresponding values of the proposed WARIMA models, with 3 types of wavelets (DB4, DB8 and DB16) are listed in Table 1. In this table, MAPE and VAPE compare the WARIMA prediction results with the original (observed) data while MAPE1 and VAPE1 compare the WARIMA prediction results with the denoised data. It can be seen that all the average standard deviations of the WARIMA models are significantly smaller than that of the ARIMA models, which suggests that the WARIMA models are more reliable. However, all the MAPE and most VAPE values of the WARIMA models are greater than those of the ARIMA models, even though most of the differences are insignificant. This suggests that in contrast to the reported successful application of the WARIMA models in the prediction of electricity prices (Conejo, 2005); using the DWT analysis does not seem to improve the ARIMA models in terms of the MAPE.

Table 1. Prediction results of WARIMA models.

Wavelet Type	Decomp. Level	Std (Vehicle)	MAPE (%)	VAPE	MAPE1 (%)	VAPE1
DB4	*A(1)*	42	14.2	0.082	13.5	0.071
	A(2)	23	14.3	0.091	6.7	0.007
	A(3)	7	14.4	0.102	4.0	0.001
	A(4)	13	16.0	0.073	2.0	0.003
DB8	*A(1)*	32	14.7	0.078	9.9	0.025
	A(2)	24	14.4	0.085	6.9	0.008
	A(3)	18	14.0	0.079	5.3	0.004
	A(4)	14	14.5	0.098	4.2	0.003
DB16	*A(1)*	35	14.5	0.078	11.0	0.042
	A(2)	22	14.5	0.085	6.5	0.007
	A(3)	13	14.0	0.091	3.9	0.002
	A(4)	9	15.1	0.115	2.6	0.001

A close look at the values in Table 1 suggests that the WARIMA models can predict the denoised data with much smaller MAPE1 and VAPE1 values when higher decomposition and reconstruction level data are used for the model calibration. This seems to indicate that the original data has a lot of noise, and the wavelet decomposition is an effective way of denoising such that WARIMA model can accurately predict the denoised data, though the WARIMA model does not seem to outperform the ARIMA model when predicting the original data as discussed in the previous paragraph. Since wavelet decomposition is good at removing noise and retaining underlying trend of data, the WARIMA model may produce better results for 2-step or other multiple-step forecasting. In addition, different wavelet types can be tried to see if they may improve the prediction accuracies of WARIMA models.

Conclusions

In this study, the WARIMA models are tested and compared with the conventional ARIMA models for short-term traffic volume forecasting based on the real-world data. The results show that the WARIMA models have significantly smaller average standard deviations of prediction errors than the ARIMA models, but their average prediction accuracies are no better than that of the ARIMA models if the original data are compared with the prediction values. The performance of WARIMA models is much better if the predicted values are compared with the denoised data. Further studies can try different wavelet types or compare the multiple-step forecasting results of ARIMA and WARIMA models.

Acknowledgments

The authors would like to thank the FHWA and the NGSIM program for providing the data used in this study.

References

Ahmed, M.S. and Cook, A.R. (1979). "Analysis of freeway traffic time-series data by using Box-Jenkins techniques." *Transportation Research Record, No.722*, 1-9.

Box, G.E.P. and Jenkins, G.M. (1976). *Time series analysis: forecasting and control*, Holden-Day, Oakland, California.

Burrus, C.S., Gopinath, R.A., and Guo, H.T. (1998). *Introduction to wavelets and wavelet transforms: A primer*. Prentice Hall, Inc., New Jersey.

Conejo, A.J., Plazas, M.A., Espinola, R., and Molina, A.B. (2005). "Day-ahead electricity price forecasting using the wavelet transform and ARIMA models." *IEEE Transactions on Power Systems*, 20(2), 1-9.

Hamed, M.M., Al-Masaeid, H.R., and Bani Said, Z.M. (1995). "Short-term prediction of traffic volume in urban arterials." *Journal of Transportation Engineering*, 121(3), 249-254.

Jiang, X. and Adeli, H. (2005). "Dynamic wavelet neural network model for traffic flow forecasting." *Journal of Transportation Engineering*, 131(10), 771-779.

Nicholson, H. and Swann, C.D. (1974). "The prediction of traffic flow volumes based on spectral analysis." *Transportation Research*, 8, 533-538.

Nihan, N.L. and Holmesland, K.O. (1980). "Use of the Box and Jenkins time series technique in traffic forecasting." *Transportation*, 9(2), 125-143.

Smith, B.L., Williams, B.M., and Oswald, R.K. (2002). "Comparison of parametric and nonparametric models for traffic flow forecasting." *Transportation Research Part C*, 10(4), 303-321.

Stathopoulos, A. and Karlaftis, M.G. (2003). "A multivariate state space approach for urban traffic flow modeling and prediction." *Transportation Research Part C*, 11(2), 121-135.

Van Der Voort, M., Dougherty, M., and Watson, S. (1996). "Combining Kohonen maps with ARIMA time series models to forecast traffic flow." *Transportation Research Part C*, 4(5), 307-318.

Willams, B.M., Durvasula, P.K., and Brown, D.E. (1998). "Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models." *Transportation Research Record, No.1644*, 132-141.

Real-Time Traffic State Estimation on Urban Road Network: The Application of Unscented Kalman Filter

R. Pueboobpaphan[1] and T. Nakatsuji[2]

[1]Doctoral Student, Graduate School of Engineering, Hokkaido University, Kita-13, Nishi-8, Kita-ku, Sapporo 060-8628, Japan; Tel: +81-11-706-6822; Fax: +81-11-706-6216; email: ratta@eng.hokudai.ac.jp

[2]D.Eng., Assoc. Prof., Graduate School of Engineering, Hokkaido University, Kita-13, Nishi-8, Kita-ku, Sapporo 060-8628, Japan; Tel: +81-11-706-6215; Fax: +81-11-706-6216; email: naka@eng.hokudai.ac.jp

Abstract

Feedback approach for real-time traffic state estimation on urban road with signalized intersection is presented. The conventional discretization of kinematic wave model for freeway based on demand and supply concept is modified in order to be capable of describing traffic state on urban road with signalized intersection and to better estimate of traffic speed. The novel filter, namely Unscented Kalman filter (UKF), is used as a correction algorithm to make the estimate to be consistent with observed traffic data. Numerical tests show superiority of the modified model over the conventional model on speed estimation and the improvement in flow estimation when combining the model with UKF.

Introduction

Reliable traffic information is essential for development of efficient Advanced Traffic Management Systems (ATMS) and Advanced Traveler Information Systems (ATIS). Traffic information can be obtained by traffic detectors installed on the road network. However, it is very costly and unrealistic in practice to densely install detectors in order to have a complete view of traffic situation over the whole network. Alternatively, one may estimate traffic state from the limited sources of data using traffic flow model and feedback technique (Nanthawichit et al., 2003; Wang and Papageorgiou, 2005). Most of the literatures in this field focus on the application of feedback technique on freeway network but it is rarely seen in urban road with signalized intersection. In addition, most of them used Extended Kalman Filter (EKF) as filtering technique due to nonlinearity of system equations. However, there exist other filters for nonlinear system that have superior performance with the same order of complexity as of EKF.

In this paper, we present two new developments in traffic state estimation on urban road with signalized intersection. The first one is the modification of the well developed macroscopic model for freeway based on demand and supply concept (Daganzo, 1994; Lebacque, 1996). The modification takes the interruption from

traffic signal into account and considers the fact that drivers adjust their speed according to the immediate traffic density as well as the traffic situation ahead. The second one is the application of newly developed nonlinear filtering technique, Unscented Kalman Filter (UKF), to traffic state estimation. This filter has an ability to capture true mean and covariance accurately to second order Taylor series expansion and does not require explicit computation of matrix derivative.

Macroscopic Traffic Flow Model

We use the term link to denote the portion of road connecting between two nodes (or intersections). Each link is discretized into small segments and these segments are numbered in ascending order from 1, for the first segment upstream, to *N*, for the last segment downstream. Similar to other macroscopic models, the model used here includes equation of conservation as shown below:

$$\rho_j^l(k+1) = \rho_j^l(k) + \frac{\Delta t}{\Delta x_j^l}\left(q_{j-1}^l(k) - q_j^l(k)\right) \tag{1}$$

where $\rho_j^l(k)$ is density of segment j on link l at time step k, $q_j^l(k)$ is average flow crossing boundary point between segments j and j+1on link l during time step k, Δx_j^l is length of segment j on link l, Δt is time increment, k is time step. For stability reason, selection of Δx_j^l and Δt must meet the criteria $\left(v_{f,j}^l \cdot \Delta t\right)/\Delta x_j^l \le 1$ where $v_{f,j}^l$ is a free flow speed on segment j on link l. Outflow $q_j^l(k)$ from all segments j, *except the last segment of all links connecting to intersection*, is estimated as follows:

$$q_j^l(k) = \min\left\{D\left(\rho_j^l(k)\right), S\left(\rho_{j+1}^l(k)\right)\right\} \tag{2}$$

$$D\left(\rho_j^l(k)\right) = \begin{cases} Q_e\left(\rho_j^l(k)\right) & \text{if } \rho_j^l(k) < \rho_{c,j}^l \\ q_{\max,j}^l & \text{if } \rho_j^l(k) \ge \rho_{c,j}^l \end{cases} \tag{3}$$

$$S\left(\rho_{j+1}^l(k)\right) = \begin{cases} q_{\max,j+1}^l & \text{if } \rho_{j+1}^l(k) < \rho_{c,j+1}^l \\ Q_e\left(\rho_{j+1}^l(k)\right) & \text{if } \rho_{j+1}^l(k) \ge \rho_{c,j+1}^l \end{cases} \tag{4}$$

where $D\left(\rho_j^l(k)\right)$ is local demand (or maximum sending flow) of segment j on link l during time step k, $S\left(\rho_{j+1}^l(k)\right)$ is local supply (or maximum receiving flow) of segment j+1 on link l during time step k, $q_{\max,j}^l$ is capacity flow into segment j on link l, $\rho_{c,j}^l$ is critical density corresponding to the density at the capacity flow of segment j on link l, $Q_e(\cdot)$ is equilibrium flow-density relationship.

Special equation is required for the *outflow from last segment of all links connecting to intersection* in order to explicitly take signal effect into account as follows:

$$q^{mn}(k) = A^{mn}(k)G^{mn}(k) \cdot \mathrm{Min}\left\{\gamma^{mn}(k) \cdot D^m(k), S^{mn^*}(k)\right\} \tag{6}$$

where $q^{mn}(k)$ is the outflow from last segment of link m to the first segment of link n during time k, $D^m(k)$ is the local demand at the last segment of link m during time k, $S^{mn*}(k)$ is the *actual* supply of first segment of link n for the flow from link m during time k, $\gamma^{mn}(k)$ is the turning ratio from m to n during time k, $A^{mn}(k)=1$ if signal allows turning from m to n during time k and $A^{mn}(k)=0$ otherwise, $G^{mn}(k)=1$ if the gap on the opposing flow of movement mn during time k is accepted by the vehicles turning from m to n and $G^{mn}(k)=0$ otherwise. Since traffic signal is utilized in order to reduce the conflict at intersection, we may postulate the priority rule for merging from any of upstream link m to a downstream link n. For example, if link m has first priority and link i has second priority to merge into link n during time k, then $S^{mn*}(k)=S\left(\rho_1^n(k)\right)$ and $S^{in*}(k)=S\left(\rho_1^n(k)\right)-q^{mn}(k)$. Note that in case of urban road network with signalized intersection, the time step Δt is usually set to a small value (3 seconds in this study) and all time varying variables can be assumed constant during the interval.

Space mean speed $v_j^l(k)$ of segment j on link l during time k and average spot speed $w_j^l(k)$ at the boundary between segments j and j+1 on link l during time k are defined in this paper as:

$$v_j^l(k)=\begin{cases}\theta\cdot V_e\left(\rho_j^l(k)\right)+(1-\theta)\cdot V_e\left(\rho_{j+1}^l(k)\right), & \text{if } V_e\left(\rho_j^l(k)\right)>V_e\left(\rho_{j+1}^l(k)\right)\\ V_e\left(\rho_j^l(k)\right), & \text{otherwise}\end{cases} \tag{7}$$

$$w_j^l(k)=\phi\cdot v_j^l(k)+(1-\phi)\cdot v_{j+1}^l(k) \tag{8}$$

where $V_e(\cdot)$ is equilibrium speed-density relationship, θ and ϕ are constant parameters that need to be calibrated and their values range between zero and one.

Macroscopic Model and Unscented Kalman Filter

We note here that all the equations of macroscopic model given above are just the approximation of actual traffic dynamic and hence include some errors. Moreover, there are often some uncertainties and situations which cannot always be anticipated such as incident, unscheduled maintenance works, etc. Therefore, the application of feedback technique is required in order to keep the estimate to be consistent with the observation data.

In this paper, state vector $\mathbf{x}$ is defined as segment density $\mathbf{x}=\left[\cdots,\rho_1^l,\rho_2^l,\cdots\right]^T$, measurement vector $\mathbf{y}$ is defined as segment outflow and spot speed $\mathbf{y}=\left[\cdots,q_1^l,w_1^l,q_2^l,w_2^l\cdots\right]^T$. Eq. (1) is considered as state equation while eq. (2) or (6) and eq. (8) are considered as measurement equations. Because of nonlinear relationship between flow and density, these equations are also nonlinear. The general form of state and observation equations which include white noise errors are:

$$\mathbf{x}(k+1) = \mathbf{f}[\mathbf{x}(k)] + \xi(k) \tag{9}$$

$$\mathbf{y}(k) = \mathbf{g}[\mathbf{x}(k)] + \zeta(k) \tag{10}$$

where $\xi(k)$ and $\zeta(k)$ are random noises representing modeling errors and measurement errors with covariance $\mathbf{R}^{\mathbf{v}}$ and $\mathbf{R}^{\mathbf{n}}$, respectively. They are assumed to be independent, zero mean, and Gaussian. As actual distribution is not known and many random processes occurring in nature actually appear to be normally distributed, to some extent this assumption is therefore suitable. The algorithm of UKF presented here is based on Wan and Merwe (2001).

Let L denotes the dimension of state vector $\mathbf{x}(k-1)$ with a posteriori estimate as $\hat{\mathbf{x}}(k-1)$ and error covariance $\mathbf{P}_{\mathbf{x}}(k-1)$. A matrix $\mathbf{X}(k-1)$ of $2L+1$ sigma vectors $X_i(k-1)$ is drawn as follows:

$$\begin{aligned} X_0(k-1) &= \hat{\mathbf{x}}(k-1) \\ X_i(k-1) &= \hat{\mathbf{x}}(k-1) + \left(\sqrt{(L+\lambda)\mathbf{P}_{\mathbf{x}}(k-1)}\right)_i, \quad i = 1,....,L \\ X_i(k-1) &= \hat{\mathbf{x}}(k-1) - \left(\sqrt{(L+\lambda)\mathbf{P}_{\mathbf{x}}(k-1)}\right)_{i-L}, \quad i = L+1,....,2L \end{aligned} \tag{11}$$

where $\lambda = \alpha^2(L+\kappa) - L$ is a scaling parameter. Constant α determines the spread of sigma points around $\hat{\mathbf{x}}(k-1)$ and is usually set to $0.0001 \le \alpha \le 1$. Constant κ is a second scaling parameter (we set $\kappa = 0$). The weight W_i can be calculated as

$$\begin{aligned} W_0^m &= \frac{\lambda}{L+\lambda}, \\ W_0^c &= \frac{\lambda}{L+\lambda} + 1 - \alpha^2 + \beta \\ W_i^m &= W_i^c = \frac{1}{2(L+\lambda)}, \; i = 1,....,2L \end{aligned} \tag{12}$$

where β is used to incorporate prior knowledge of the distribution of $\mathbf{x}$ (in this study, we set $\beta = 0$). The time-update equations are

$$\mathbf{X}^*(k) = \mathbf{f}\left(\mathbf{X}(k-1)\right) \tag{13}$$

$$\hat{\mathbf{x}}^-(k) = \sum_{i=0}^{2L} W_i^m X_i^*(k) \tag{14}$$

$$\mathbf{P}^-(k) = \sum_{i=0}^{2L} W_i^c \left(X_i^*(k) - \hat{\mathbf{x}}^-(k)\right)\left(X_i^*(k) - \hat{\mathbf{x}}^-(k)\right)^T + \mathbf{R}^{\mathbf{v}} \tag{15}$$

Redraw sigma points from $\hat{\mathbf{x}}^-(k)$ and $\mathbf{P}^-(k)$ using eq. (11), we obtain $\mathbf{X}(k)$. The time-update of observation variables are

$$\mathbf{Y}(k) = \mathbf{g}\left(\mathbf{X}(k)\right) \tag{16}$$

$$\hat{\mathbf{y}}^{-}(k) = \sum_{i=0}^{2L} W_i^m Y_i(k) \tag{17}$$

The measurement-update equations are

$$\mathbf{P}_{yy}(k) = \sum_{i=0}^{2L} W_i^c \left(Y_i(k) - \hat{\mathbf{y}}^{-}(k)\right)\left(Y_i(k) - \hat{\mathbf{y}}^{-}(k)\right)^T + \mathbf{R}^n \tag{18}$$

$$\mathbf{P}_{xy}(k) = \sum_{i=0}^{2L} W_i^c \left(X_i(k) - \hat{\mathbf{x}}^{-}(k)\right)\left(Y_i(k) - \hat{\mathbf{y}}^{-}(k)\right)^T \tag{19}$$

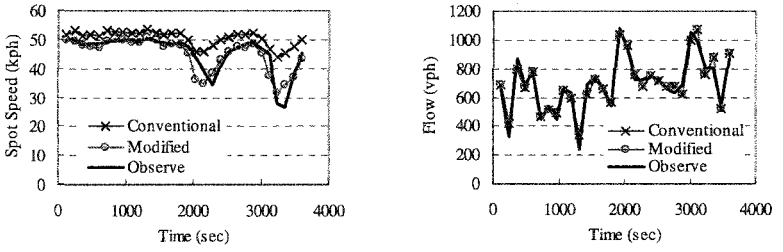
$$\mathbf{K}(k) = \mathbf{P}_{xy}(k) \cdot \mathbf{P}_{yy}^{-1}(k) \tag{20}$$

$$\hat{\mathbf{x}}(k) = \hat{\mathbf{x}}^{-}(k) + \mathbf{K}(k)\left[\mathbf{y}(k) - \hat{\mathbf{y}}^{-}(k)\right] \tag{21}$$

$$\mathbf{P}(k) = \mathbf{P}^{-}(k) - \mathbf{K}(k)\mathbf{P}_{yy}(k)\mathbf{K}^T(k) \tag{22}$$

Numerical Results

In this study, traffic data of an isolated intersection are obtained using simulation software, INTEGRATION. The total simulation time is 60 minutes and $\Delta t = 3$ s. All links have equal length of 300 meters and each link is divided into 4 segments of 75 meters. The parameters for all segments are the same as $v_f = 55$ km/hr, $q_{\max} = 1{,}650$ veh/hr/lane, $\rho_{\max} = 120$ veh/km/lane, and $v_c = 22.5$ km/hr. Inflow rates are assumed to increase at the mid of simulation period. We considered two set of inflow rate: set A and set B to reflect the case of moderate and heavy traffic respectively. The turning ratio of a major link is assumed to change at the mid of simulation and this change is then treated as known or unknown in the model to reflect the case of certainty and uncertainty. Assume that detectors are installed at the entrances and exits of the network. Figure 1 shows the results from conventional and modified model compared to the observed value on an entry link.



(a) Spot Speed (b) Flow

Figure 1 Estimation results; $\theta = 1$ and $\phi = 0.7$ for conventional model; $\theta = 0.5$ and $\phi = 0.7$ for modified model

Root Mean Square Error (RMSE) is computed to compare the estimation results and the observation of flow and spot speed at the mid of exit links as follows:

$$RMSEz = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(z_i - \hat{z}_i\right)^2} \tag{23}$$

Table 1 shows the RMSE of different cases. Case 1 – 3 used the same inflow set A while case 4 - 5 used inflow set B. The term "Mac" represents the case of conventional model, "Mac + Mod. Speed" represents the modified model, and "Mac + Mod. Speed + UKF" represents the modified model with UKF. "Perfect" and "Imperfect" represent respectively the case that turning ratios are perfectly or imperfectly known. The results showed the improvement of the modified model over the conventional model in speed estimation. When UKF was applied, substantial improvement in flow estimation can be found while speed estimation becomes slightly deteriorate. Since flow and spot speed are related to each other, improvement in flow estimation could sometimes result in deteriorate of spot speed estimation. Nevertheless, in overall, UKF provides the minimum variance estimate.

Table 1 Root Mean Square Error

case	Inflow	Name	RMSE flow		RMSE Spot Speed	
			Perfect	Imperfect	Perfect	Imperfect
1	Moderate	Mac	129.03	145.89	6.31	6.37
2	Moderate	Mac + Mod. Speed	129.03	145.89	6.02	6.10
3	Moderate	Mac + Mod. Speed + UKF	112.42	123.82	6.06	6.11
4	Heavy	Mac	144.83	186.25	6.75	6.92
5	Heavy	Mac + Mod. Speed	144.83	186.25	6.45	6.65
6	Heavy	Mac + Mod. Speed + UKF	128.91	161.78	6.54	6.69

Conclusion

This paper presented the modified macroscopic model in order to be capable of describing surface street traffic dynamic. Feedback technique using UKF was applied to the model to capture the errors of model and unexpected uncertainties. Numerical results showed the improvement of the modified model over the conventional one and the benefit gained from using UKF. Future work is directed to the simultaneous estimation of traffic state and turning ratio.

References

Daganzo, C.F. (1994) The Cell Transmission Model: A Dynamic Representation of Highway Traffic Consistent with the Hydrodynamic Theory. *Transportation Research, Vol. 28B, No. 4*, pp.269-287.

Lebacque, J.P. (1996) The Godunov Scheme and What It Means for First Order Traffic Flow Models. In Lesort, J.B. (ed), *Proceedings of the 13th International Symposium of Transportation and Traffic Theory*, Lyon, pp.647-677

Nanthawichit, C., Nakatsuji, T., and Suzuki, H. (2003) Application of Probe Vehicle Data for Real-Time Traffic State Estimation and Short-Term Travel Time Prediction on a Freeway, *Transportation Research Record 1855*, pp. 49-59.

Wang, Y., and Papageorgiou, M. (2005) Real-Time Freeway Traffic State Estimation based on Extended Kalman Filter: A General Approach. *Transportation Research, Vol. 39B, No. 2*, pp. 141-167.

Wan, E.A., and Merwe, R. van der. (2001) The Unscented Kalman Filter. In Haykin, S. (ed.), *Kalman Filtering and Neural Networks*, John Wiley and Sons, Inc.

Revised Approaches for Estimation of Time-Varying Origin-Destination Flows in Freeway Corridors

Jiao Pengpeng[1], Lu Huapu[2], Noboru Harata[3]

[1]Postdoctoral Researcher, Department of Urban Engineering, the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. Tel: +81-3-5841-6234, Fax: +81-3-5841-8527, Email: jiao_pengpeng@ut.t.u-tokyo.ac.jp (until Sep. 2005, Department of Civil Engineering, Tsinghua University, Beijing 100084, China)

[2]Professor, Department of Civil Engineering, Tsinghua University, Beijing, 100084 China. Tel: +86-10-62795339, Fax: +86-10-62795339, Email: luhp@mail.tsinghua.edu.cn

[3]Professor, Department of Urban Engineering, the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. Tel: +81-3-5841-6233, Fax: +81-3-5841-8527, Email: nhara@ut.t.u-tokyo.ac.jp

Abstract

This paper puts forward two revised approaches for estimating the real-time O-D flows in freeway corridors. The first approach is formulated in the form of parameter optimization (PO), with the objective function to minimize the absolute deviations between estimated and observed traffic counts. The second approach is formulated as a state-space model using Kalman filtering (KF), with O-D flow deviations as state vectors. These two models make full use of information from ramp traffic counts and mainline flow measurements, which are provided by traffic surveillance systems. The models also take into account the traffic delay by integrating the dynamic route travel time, which is estimated using a recursive method reported in this paper. A genetic algorithm is designed to solve the PO model, as well as an extended KF to solve the state-space model. The results of simulation examples indicate that the proposed approaches are rather promising for potential applications.

Introduction

Getting dynamic O-D flows directly from conducting surveys is very difficult and costly. Over the past two decades, the development of an accurate and efficient method for dynamic O-D flows estimation from link counts has received increasing attention among researchers. Cremer and Keller (1981), Cremer and Keller (1984), in a series of pioneering work for identifying intersection flow movements, have first employed relations constructed through time-series of traffic counts, and converted the underdetermined static model to an overdetermined dynamic formulation. Nihan and Davis (1987), Cremer and Keller (1987) have also respectively presented a set of dynamic O-D estimation models for intersections or small networks based on the prediction-error minimization methods. A rather general review of studies in this category is available in Bell (1991).

Another category of studies for estimating the complex dynamic O-D matrices focuses on formulating the model based on state-space. Okutani (1987) formulated a Kalman filtering model for dynamic O-D estimation and prediction, using an autoregressive model to capture the temporal interdependency among O-D flows and link counts. Along the same line, Ashok and Ben-Akiva (1993, 2002) have revised Okutani's work with the state variables representing the deviations of O-D flows from prior estimates based on corresponding historical data. However, all these methods are based on the assumption that an accurate dynamic traffic assignment (DTA) model exists. Unfortunately, the development of a reliable DTA model is a quite challenging task, and may not be solved in the near future.

Fortunately, there exist unique properties in freeway corridors that provide valuable information for their dynamic O-D estimation, and the dynamic relations between O-D pairs and link counts can thus be formulated. Chang and Wu (1994) proposed a recursive model along this line, and applied extended KF to solve it. Though KF is rather efficient, its results

are not very satisfying, and the traffic delay is not taken into account in the estimation of time-varying travel time, which might be rather big in a congested freeway. In this paper, a further research for freeway corridors will be conducted, including models and algorithms.

This paper consists of the following sections. The basic nonlinear dynamic interrelations between time-dependent O-D flows and time-series of traffic counts are illustrated in Section 2. A recursive approach to estimate dynamic travel time is proposed in Section 3. A revised parameter optimization model is represented in Section 4, with a genetic algorithm to solve it. A revised state-space model with the O-D deviations as state vectors is further formulated in Section 5. The evaluation results based on simulation experiments are reported in Section 6. Conclusions and further researches are summarized in the last section.

Basic Dynamic Interrelations and Travel Time Estimation

Consider a freeway corridor of N segments from 0 to (N-1) as shown in Figure 1.

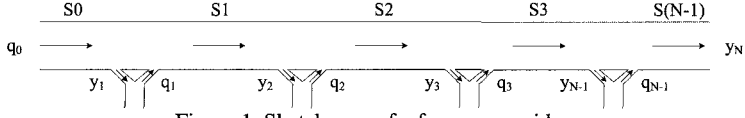


Figure 1. Sketch map of a freeway corridor

Variables used in this paper are defined below:

$q_0(k)$: the number of vehicles entering from upper section of S0 during interval k;

$q_i(k)$: the number of vehicles entering from on-ramp i during interval k, $i=1,2,\ldots,N-1$;

$y_j(k)$: the number of vehicles leaving from off-ramp j during interval k, $j=1,2,\ldots,N-1$;

$y_N(k)$: the number of vehicles leaving from S(N-1) during interval k;

$T_{ij}(k)$: the number of vehicles entering from on-ramp i during interval k with destination at off-ramp j, i.e. the time-varying O-D flows, where $0 \le i < j \le N$;

$b_{ij}(k)$: split parameter, it equals to the ratio of vehicles in $q_i(k)$ which leave the freeway corridor from off-ramp j;

$U_i(k)$: the number of vehicles passing upper section of segment i during interval k, $i=0,1,...,N-1$.

Here $q_0(k)$, $y_N(k)$, $q_i(k)$ and $y_j(k)$, as trip generation and attraction of each node, could be obtained by detectors, as well as the mainline traffic counts $U_i(k)$. Therefore, the problem to be solved here can be stated as: how to get the time-varying O-D flows $T_{ij}(k)$ or split parameters $b_{ij}(k)$ from the time-series of known traffic counts.

The variable $\rho_{ij}^m(k)$ is introduced to denote the proportion of O-D flows $q_i(k-m)b_{ij}(k-m)$ in interval $(k-m)$ that arrive at an off-ramp during interval k, namely, the contribution of $q_i(k-m)b_{ij}(k-m)$ to $y_j(k)$. $\rho_{ij}^m(k)$ is named influence factor (IF) of travel time in this paper. Therefore, the relation between entry traffic volume, split parameter and exit traffic volume can be stated as:

$$y_j(k) = \sum_{m=0}^{M} \sum_{i=1}^{j-1} \rho_{ij}^m(k) q_i(k-m) b_{ij}(k-m) \tag{1}$$

where M is the maximum number of intervals during which trip generation at upper node i has influence on trip attraction at lower node j, which is set to be the number of intervals

corresponding to the maximum travel time for vehicles to traverse the freeway corridor.

Obviously, there exist the following natural constraints of $\rho_{ij}^{m}(k)$:

$$0 \le \rho_{ij}^{m}(k) \le 1, \quad 0 \le i < j \le N, \quad m = 0,1,...,M \tag{2}$$

$$\sum_{m=0}^{M} \rho_{ij}^{m}(k+m) = 1, 0 \le i < j \le N \tag{3}$$

For mainline traffic counts, the dynamic relation can be represented as:

$$U_l(k) - q_l(k) = \sum_{m=0}^{M} \sum_{i=0}^{l-1} \sum_{j=l+1}^{N} q_i(k-m)\rho_{il}^{m}(k) b_{ij}(k-m), \qquad l = 1,2,...,N-1 \tag{4}$$

In summary, equations (1) and (4) formulate the dynamic interrelations between time-varying O-D flows and link traffic counts.

Recursive Estimation of Dynamic Travel Time

A key problem in above interrelations is $\rho_{ij}^{m}(k)$, which greatly increases the number of unknown variables. To get it, one should estimate the dynamic travel time of vehicles first.

Concerning freeway corridors, the average travel time $\tau_a(k)$ of vehicles on a sub-corridor a in interval k is available using existing method. Therefore, the route travel time during each interval can be obtained by recursive method. Denote $\eta^{ij}(k)$ as the average travel time for entering vehicles from on-ramp i in interval k to arrive at off-ramp j. Under the condition that $\eta^{i(j-1)}(k)$ is an integer times of Δk, one can get:

$$\eta^{ij}(k) = \eta^{i(j-1)}(k) + \tau_a[k + \eta^{i(j-1)}(k)/\Delta k], \qquad 0 \le i < j \le N \tag{5}$$

In case that $\eta^{i(j-1)}(k)$ is not an integer times of Δk, its adjacent two integers can be used for approximation. Define INT() as an integer conversion function, and $\xi^{i(j-1)}(k) = \mathrm{INT}[\eta^{i(j-1)}(k)/\Delta k]$, (5) can be transformed into the following formulation:

$$\eta^{ij}(k) = \eta^{i(j-1)}(k) + \frac{\tau_a[k + \xi^{i(j-1)}(k)] + \tau_a[k + \xi^{i(j-1)}(k) + 1]}{2}, \quad 0 \le i < j \le N \tag{6}$$

From (6), we can get the time-varying route travel time between each pair of nodes.

Parameter Optimization Model

From system equation (1) and (4), one can directly formulate an optimization model using least-squares (LSQ) method. However, the parameter optimization method to minimize the absolute deviations (LAD) between observed and estimated traffic counts is more attractive in the presence of outliers in the input data. In such case, an LSQ model might cause outliers to greatly influence the solution, whereas an LAD model has a less effect on the solution. Moreover, the LAD model is readily able to accommodate additional side-constraints which further control the determined split parameter estimates. Thus we get the following model:

$$\begin{aligned} J = \min & \sum_{k=1}^{M} \sum_{j=1}^{N} \left| y_j(k) - \sum_{m=0}^{M} \sum_{i=0}^{j-1} \rho_{ij}^{m}(k) q_i(k-m) b_{ij}(k) \right| \\ & + \min \sum_{k=1}^{M} \sum_{s=1}^{N-1} \left| U_s(k) - q_s(k) - \sum_{m=0}^{M} \sum_{i=0}^{s-1} \sum_{j=s+1}^{N} q_i(k-m)\rho_{is}^{m}(k) b_{ij}(k) \right| \end{aligned} \tag{7a}$$

$$\text{s.t.} \quad \sum_{j=i+1}^{N} b_{ij}(k) = 1, \quad b_{ij}(k) \ge 0, \quad 0 \le i < j \le N \tag{7b}$$

Due to the form of absolute deviation, there is no simple optimization method to solve the model. Here we use a genetic algorithm (GA) to solve it. The implementation

process is as following:

1) Set up encoding and decoding formulae, and individual fitness function; 2) Generate the initial population; 3) Evaluate the fitness value for each individual in the population; 4) Reproduction procedure; 5) Crossover procedure; 6) Mutation; and 7) Select new population for next generation with a specific selection method.

State-Space Model

To improve the efficiency, we further propose a state-space model.

Ashok and Ben-Akiva (1993) pointed out that using deviations of O-D flows from historical data as state vectors can overcome the inadequacy of autoregressive process. Furthermore, it can exploit the trip-making patterns in the O-D flows and their spatial and temporal variations, and also to be amenable to the approximation by a normal distribution. Denote variables with superscript H as corresponding historical data, we set up the transition equation as following autoregressive process in matrix formulation:

$$\mathbf{T}(k)-\mathbf{T}^H(k)=\sum_{d=1}^{p}\mathbf{C}_k^d[\mathbf{T}(k-d)-\mathbf{T}^H(k-d)]+\mathbf{W}(k) \tag{8}$$

where $\mathbf{C}_k^d$ is an $[N(N+1)/2]*[N(N+1)/2]$ matrix of effects of $[\mathbf{T}(k-d)-\mathbf{T}^H(k-d)]$ on $[\mathbf{T}(k)-\mathbf{T}^H(k)]$, $\mathbf{T}(k)$ is the vector form of $T_{ij}(k)$, p is the number of lagged O-D flow deviations assumed to affect the O-D flow deviation in interval k, i.e., the degree of the autoregressive process, and $\mathbf{W}(k)$ is an $N(N+1)/2$ vector of random errors. It is a column Gaussian white noise vector with zero mean and covariance $\mathbf{D}\delta_{kl}$, where $\mathbf{D}$ is a constant semi-positive matrix, and δ_{kl} is the Kronecker delta.

For measurement equations, essential system equations (1) and (4) can be transformed to following state-space matrix forms, with O-D deviations as state vectors:

$$\mathbf{Y}(k)-\mathbf{Y}^H(k)-\sum_{m=1}^{M}\mathbf{A}_k^m[\mathbf{T}(k-m)-\mathbf{T}^H(k-m)]=\mathbf{A}_k^0[\mathbf{T}(k)-\mathbf{T}^H(k)]+\mathbf{U}(k) \tag{9}$$

$$\mathbf{X}(k)-\mathbf{X}^H(k)-\sum_{m=1}^{M}\mathbf{G}_k^m[\mathbf{T}(k-m)-\mathbf{T}^H(k-m)]=\mathbf{G}_k^0[\mathbf{T}(k)-\mathbf{T}^H(k)]+\mathbf{V}(k) \tag{10}$$

where $\mathbf{Y}(k)$ is the vector form of $y_j(k)$, $\mathbf{X}(k)$ is the vector form of $U_l(k)-q_l(k)$, $\mathbf{T}(k)$ is the vector form of $T_{ij}(k)$, $\mathbf{A}_k^m$ is $[N\times N(N+1)/2]$ mapping matrix of contributions of $\mathbf{T}(k-m)$ to $\mathbf{Y}(k)$, $\mathbf{G}_k^m$ is $[(N-1)\times N(N+1)/2]$ mapping matrix of contributions of $\mathbf{T}(k-m)$ to $\mathbf{X}(k)$, $\mathbf{U}(k)$ and $\mathbf{V}(k)$ are Gaussian white noise vectors of measurement errors. Since all $\mathbf{T}(k-m)$ have already been estimated during foregoing intervals, we move them to left side of the equations to get standard measurement equations.

Eq. 8 along with (9) and (10) formulate the revised state-space model. It can be solved using extended Kalman filtering (Chui and Chen, 1991). The mapping matrix $\mathbf{A}_k^m$ and $\mathbf{G}_k^m$ can be obtained from (1) and (4), i.e., from estimated $\rho_{ij}^m(k)$.

Remark 1. Computation of the state transition matrix $\mathbf{C}_k^d$ involves estimating linear regression models for each O-D pair. The error covariance matrix $\mathbf{D}$ could be approximated from the residuals of these regressions.

Remark 2. The historical data used in the model can be obtained from traffic surveillance system, or from results of estimations conducted in previous days. In the latter way, the historical data should be classified by day of week, type of weather, special events, etc. Furthermore, they should be updated everyday.

Simulation Examples

A simple freeway corridor shown in Figure 2 was used for numerical analysis. This corridor contains one upper boundary, one on-ramp, two off-ramps, and one lower boundary.

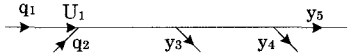


Figure 2. Example freeway corridor for simulation example

Here the time-varying $q_i(k)$, $y_j(k)$ and $U_l(k)$ are obtained from traffic simulation, as well as the actual values of dynamic O-D flows.

For comparison, three methods are performed to estimate the time-varying O-D flows, including existing Kalman filtering method (KF), revised PO model and revised KF (RKF) model in this paper. The following initial values are adopted for KF and RKF process, and the estimation results of them are also transformed to split parameters.

$$b_{13}=0.33,\ b_{14}=0.33,\ b_{15}=0.34,\ b_{23}=0.33,\ b_{24}=0.33,\ b_{25}=0.34$$

Using root mean square error (RMS) and RMS normalized (RMSN) as evaluation criteria, the statistical results are shown in Table 1. A graphical illustration of b_{13} is shown in Figure 3. For CPU time, KF is 0.57 seconds, PO is 3.4 seconds, and RKF is 0.76 seconds.

Table 1. Statistical Results of RMS and RMSN

Split Parameters	RMS			RMSN		
	KF	PO	RKF	KF	PO	RKF
b_{13}	0.021	0.012	0.022	10.72%	6.50%	11.38%
b_{14}	0.027	0.014	0.020	8.87%	4.62%	6.73%
b_{15}	0.025	0.015	0.022	4.93%	3.06%	4.25%
b_{23}	0.025	0.023	0.022	12.11%	11.10%	10.74%
b_{24}	0.026	0.015	0.022	8.53%	5.17%	7.48%
b_{25}	0.030	0.024	0.021	6.05%	4.85%	4.30%
Average	0.025	0.017	0.022	8.54%	5.88%	7.48%

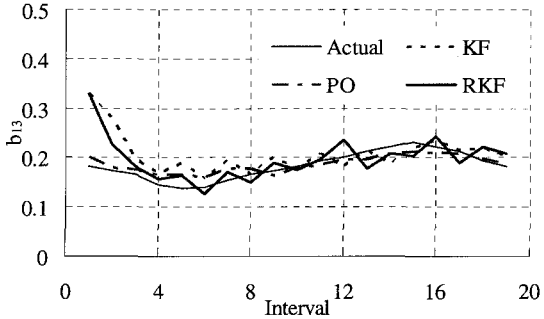


Figure 3. Estimation results of b_{13}

From above table and figure, one can find out the following results:

1) For accuracy, the revised PO model achieves the best estimation results, and the revised KF model is also more accurate than existing KF method.

2) For the capability to track dynamic O-D flows, the revised PO model is better than RKF and KF, because it is iterative in every interval, which can be found out from the estimation results of first four intervals.

3) For efficiency, the revised PO model needs most CPU time for estimation,

because it is iterative in nature. However, it is efficient enough for practical use. The revised KF model is more efficient than PO model.

Conclusions

In this paper, two revised models have been formulated for estimating dynamic O-D flows in freeway corridors under the assumption that time-series of link counts are available from detectors. A recursive estimation method of time-varying travel time is first proposed, and nonlinear dynamic interrelations between time-dependent O-D flows and traffic flow measurements are presented. Based on this, the first approach is formulated as a PO model, and the second one is in a state-space form. A genetic algorithm is developed to solve the PO model, as well as the extended KF to solve the state-space model. The reported simulation examples have demonstrated the accuracy and efficiency of revised approaches.

Further research should be directed towards four aspects. The first is to optimize the number and location of detectors under a budget constraint to provide as enough data as possible. The second is to extend the models to a real-size network. The third is to take into account traveler information provided by ATIS, and to integrate the dynamic estimation methods with ATMS, such as ramp metering and route guidance systems. And the fourth, to analyze the complexity of the estimation and test its performance under large network case.

Acknowledgement

This research was supported by Beijing Municipal Science and Technology Commission project (H030630530120) and the JSPS (Japan Society for the Promotion of Science) postdoctoral fellowship partially. The authors would like to thank the anonymous referees for the careful review of this paper.

References

Ashok, K. and Ben-Akiva, M.E. (1993) Dynamic origin-destination matrix estimation and prediction for real-time traffic management systems, *Proceedings 12th International Symposium of Transportation and Traffic Theory*, Berkeley, CA.

Ashok, K. and Ben-Akiva, M.E. (2002) Estimation and prediction of time-dependent origin-destination flows with a stochastic mapping to path flows and link flows, *Transportation Science*, 36(2), 184-198.

Bell, M.G.H. (1991) The real time estimation of origin-destination flows in the presence of platoon dispersion, *Transportation Research B*, 25, 115-125.

Chang, G.L. and Wu, J.F. (1994) Recursive estimation of time-varying origin-destination flows from traffic counts in freeway corridors, *Transportation Research B*, 28(2), 141-160.

Chui, C.K. and Chen, G. (1991) *Kalman filtering with real-rime applications.* Springer-Verlag, New York.

Cremer, M. and Keller, H. (1981) Dynamic identification of O-D flows from traffic counts at complex intersections, *Proceedings 8th International Symposium on Transportation and Traffic Theory*, University of Toronto Press, Toronto, Canada.

Cremer, M. and Keller, H. (1984) A Systems dynamics approach to the estimation of entry and exit O-D flows, *Proceedings 9th International Symposium on Transportation and Traffic Theory*, VUN Science Press, Utrecht, the Netherlands.

Cremer, M. and Keller, H. (1987) A new class of dynamic methods for the identification of origin-destination flows, *Transportation Research B*, 21, 117-132.

Nihan, N.L. and Davis, G.A. (1987) Recursive estimation of origin-destination matrices from input/output counts, *Transportation Research B*, 21, 149-163.

Okutani, I. (1987) The Kalman filtering approaches in some transportation and traffic problems, *Proceedings 10th International Symposium on Transportation and Traffic Theory*, Elsevier Science Publishing Co., Inc., New York.

A Robust Algorithm of Estimating Dynamic O-D Matrix for Large Freeway Networks with Measurement Errors

Pei-Wei Lin and Gang-Len Chang
Dep. of Civil and Environmental Engineering, Univ. of Maryland, College Park, MD 20742,
pwlin@wam.um.edu

Abstract

This study presents a robust algorithm that can deal with the incomplete volume information so as to significantly improve the estimation accuracy. To tackle the inevitable measurement data error or only partially available information, the proposed robust algorithm converts each piece of model input data into one interval with its upper and lower bounds best approximated from historical data or/and prior knowledge. A simulated system, the I-95 freeway corridor between I-495 and I-695, has been created to generate example data and to perform the numerical evaluation of the developed robust algorithm.

Introduction

Due to the increasing needs of using the O-D information for a variety of planning and operation studies, transportation researchers over the past decades have devoted considerable efforts in developing effective approaches for reliably estimating O-D demands either in a static or a time-varying format. Most existing advanced approaches (Cascetta, 1984; Cremer and Keller, 1984, 1987; Keller and Ploss, 1987; Nihan and Davis, 1987; Bell, 1991a, 1991b; Ashok and Ben-Akiva, 1993; Cascetta et al., 1993; Chang and Wu, 1994; Wu and Chang, 1996; Chang and Tao, 1999; Tavana and Mahmassani, 2000; Hazelton, 2000; Sherali and Park, 2001; Lin and Chang, 2005; Zhou and Mahmassani, 2005) for such a need require the use of a set of initial/prior O-D and ramp/mainline volumes as model inputs. Depending on the available information and the network structure, the O-D patterns estimated with those approaches may result in a large variance, and insufficient reliability to use in practice. Besides, many of those essential data may not be available in a real world network. Hence, it is imperative that any developed system for such applications be sufficiently robust to accommodate to real world operational constraints.

In general, link volume data are assumed to be available and do not contain measurement errors. The quality of such information, however, may impact the estimation accuracy. The assumption of having accurate link volume data is often subject to challenge, as a large amount of traffic volumes from detectors for dynamic O-D estimation are constantly suffering from the hardware quality deficiency. The issue of having unreliable link volume data is especially critical for a large network as its formulations for O-D estimation generally contain a significant number of unknown system parameters. Depending on the information availability and the network structure, the O-D patterns estimated with those approaches may result in a large variance, and insufficient reliability for use in practice (Lin and Chang, 2005). To contend with such deficiencies, this study develops a dynamic O-D model that employs an interval-based estimation algorithm to account for measurement variances in network link volumes, which may vary significantly from their mean volumes used in the existing O-D estimation models.

The remaining of the paper is organized as follows. A basic dynamic O-D estimation model for a freeway corridor is presented in the next section. Section 3 illustrates an interval-based solution algorithm to contend with input volume variance. Extensive numerical analyses for evaluating the effectiveness of the proposed solution algorithms are presented in Section 4. Conclusions and further enhancements are summarized in the last section.

A Basic Model Formulation for Dynamic O-D Estimation

Consider a typical freeway corridor with link count information as shown in Figure 1, where detectors are deployed at on-ramps, off-ramps and mainline links. The information that is readily available for estimation of its time-dependent O-D flow proportion or dynamic O-D distribution is the time series of entering flow, $q_i(k)$, exiting flow, $y_j(k)$, and mainline flow, $U_l(k)$.

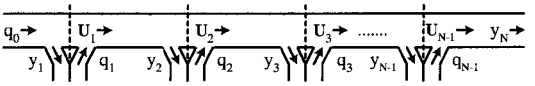


Figure 1. A Typical Freeway Corridor

Let $b_{ij}(k)$ denote the proportion of vehicles entering from origin i to destination j during time interval k. By definition, it is subject to the following two natural constraints:

$$0 \le b_{ij}(k) \le 1, \qquad 0 \le i < j \le N \tag{1}$$

$$\sum_{j=i+1}^{N} b_{ij}(k) = 1, \qquad i = 0, 1, ..., N-1 \tag{2}$$

As shown in Figure 1, the freeway corridor consists of N segments from 0 to N-1 and the set of other variables used in modeling the dynamic traffic flow and O-D relationships is defined as follows:

$q_0(k)$: The number of vehicles entering the upstream boundary of the freeway section during time interval k.
$q_i(k)$: The number of vehicles entering freeway from on-ramp i during time interval k, i=1,2,…,N-1.
$y_j(k)$: The number of vehicles leaving freeway from off-ramp j during time interval k, j=1,2,…,N-1.
$y_n(k)$: The mainline volume at the downstream end of the freeway section during time interval k.
$U_i(k)$: The number of vehicles crossing the upstream boundary of segment i during time interval k, i=1,2,…, N-1.
$T_{ij}(k)$: The number of vehicles entering the freeway from on-ramp i during time interval k that are destined to off-ramp j (i.e., the time-dependent O-D flow), where $0 \le i < j \le N$.
t_0 : The length of one unit time interval.
$t_{ij}(k)$: The average travel time from on-ramp i to off-ramp j departing during time interval k.
$\sigma_{ij}(k)$: The standard deviation of the travel time for vehicles traveling from on-ramp i to off-ramp j departing during time interval k.
$\rho^m_{ij}(k)$: The fraction of $T_{ij}(k-m)$ vehicles departing from entry node i during time interval k that takes m time intervals to exiting node j.
$\rho^m_{ilj}(k)$: The fraction of $T_{ij}(k-m)$ trips from entry node i during time interval k that takes m time intervals to mainline node *l*.

Lin and Chang (2005) assumed that the travel time of drivers departing from node i during time interval k to node j follows a normal distribution, i.e., $N[\mu_{ij}(k), \sigma^2_{ij}(k)]$. Since the travel time for an O-D pair departing during the same time interval follows a normal distribution, $\rho^m_{ij}(k)$ can be replaced with the following cumulative density function of a time interval, m.

$$\rho^m_{ij}(k) = \int_{m \cdot t_0}^{(m+1) \cdot t_0} f_{ij}(x)dx, \quad 0 \le \rho^m_{ij}(k) \le 1, \; m = 0, 1, ..., M \tag{3}$$

$$\sum_{m=0}^{M} \rho^m_{ij}(k+m) = 1, \quad 0 \le i < j \le N, \; m = 0, 1, ..., M \tag{4}$$

where $f_{ij}(x)$ is the density function of the travel time distribution with mean $t_{ij}(k)$ and standard deviation $\sigma_{ij}(k)$. By applying the above travel time distribution concept, the relationships between ramp volumes and O-D proportions, and the relationships between mainline volumes and O-D proportions can be written as:

$$y_j(k) = \sum_{m=0}^{M} \sum_{i=0}^{j-1} q_i(k-m) \rho^m_{ij}(k) b_{ij}(k-m) \tag{5}$$

$$U_\ell(k) - q_\ell(k) = \sum_{m=0}^{M} \sum_{i=0}^{\ell-1} \sum_{j=\ell+1}^{N} q_i(k-m) \rho^m_{ilj}(k) b_{ij}(k-m) \tag{6}$$

$$t_{i\ell j}(k) \sim N[\gamma_{i\ell j} \cdot \bar{t}_{ij}(k), \gamma^2_{i\ell j} \cdot \sigma^2_{ij}(k)] \quad (7)$$

$$\gamma_{i\ell j} = d_{i\ell} / d_{ij} \quad (8)$$

where $\gamma_{i\ell j}$ is the ratio of the distance $d_{i\ell}$ to the distance d_{ij}. In addition to the use of a normal distribution to capture the variation of travel times among drivers of the same departure time, one can also estimate the average $\bar{b}_{ij}(k)$ for consecutive intervals, instead of solving the O-D flow distribution matrix for each small interval (Chang and Wu, 1994). This is proposed to reduce the number of unknown parameters. With this enhancement, one can estimate the average travel time for each O-D pair with data provided by a surveillance system, and let the O-D proportions, $\bar{b}_{ij}(k)$, and standard deviations,$\sigma_{ij}(k)$ be the unknown set of parameters.

An Interval-Based Algorithm for Dynamic O-D Estimation

This section presents an enhanced solution algorithm that can estimate an O-D matrix under the circumstance that its traffic volumes are subjected to some level of measurement errors. In most existing approaches, the unknown parameters, $b_{ij}(k)$ and $\sigma_{ij}(k)$, are assumed to follow the random walk process: $b_{ij}(k+1) = b_{ij}(k) + w_{ij}(k)$ and $\sigma_{ij}(k+1) = \sigma_{ij}(k) + v_{ij}(k)$, where the random terms, $w_{ij}(k)$ and $v_{ij}(k)$ are independent Gaussian white noise sequences with zero mean and covariance.

To facilitate the formulations, the variables are defined as: $\mathbf{b(k)} = [b_{ij}(k)]$, $\boldsymbol{\sigma}\mathbf{(k)} = [\sigma_{ij}(k)]$, $\mathbf{W(k)} = [w_{ij}(k)]$, and $\mathbf{V(k)} = [v_{ij}(k)]$. Thus, the matrix forms are as follows:

$$\mathbf{b(k+1)=b(k)+W(k)} \text{ and } \boldsymbol{\sigma}\mathbf{(k+1)=}\boldsymbol{\sigma}\mathbf{(k)+V(k)} \quad (9)$$

With the above refinements for $\mathbf{b(k)}$ and $\boldsymbol{\sigma}\mathbf{(k)}$, Equations (5) and (6) can be restructured into the following matrix form:

$$\mathbf{Z(k)=H[}\boldsymbol{\sigma}_{\mathbf{ij}}\mathbf{(k)]\cdot b(k)+e(k)} \quad (10)$$

where $\mathbf{Z(k)} = [y_1(k), y_2(k), ..., y_N(k); U_1(k) - q_1(k), ..., U_{N-1}(k) - q_{N-1}(k)]^T$, $\mathbf{H[}\boldsymbol{\sigma}_{\mathbf{ij}}\mathbf{(k)]} = [h^k_{rs}]$ is a matrix with its entries given by the corresponding coefficients in Equations (5) and (6), and $\mathbf{e(k)}$ is an observation noise vector. Through Equations (9)-(10), a canonical state-space system model has been established. Such a system equation can be solved with any solution algorithm, such as the Kalman filtering or the minimum least squares estimation algorithms.

The matrices $\mathbf{Z(k)}$ and $\mathbf{H(k)}$ in Equation (10) consist of the time-series measurement of traffic volumes, $q_i(k)$, $y_i(k)$ and $U_i(k)$. If the traffic volume information is not precisely known due to measurement errors and progressively changes over time, most existing solution algorithms will not be applicable. The interval-based solution algorithm presented hereafter is proposed for contending with such a critical issue. Suppose that these two uncertain matrices of $\mathbf{Z(k)}$ and $\mathbf{H(k)}$ are only known to be bounded, one then can rewrite their relations as follows:

$$\mathbf{Z^I(k)=[Z(k)-|\Delta Z(k)|, Z(k)+|\Delta Z(k)|]} \quad (11)$$

$$\mathbf{H^I(k)=[H(k)-|\Delta H(k)|, H(k)+|\Delta H(k)|]} \quad (12)$$

where $|\mathbf{\Delta Z(k)}|$ and $|\mathbf{\Delta H(k)}|$ are the positive constant bounds for those unknowns. Hence, Equation (10) can be written as an interval representation:

$$\mathbf{Z^I(k)=H^I(k)\cdot b^I(k)+e^I(k)} \quad (13)$$

Equation (13) along with Equation (9) forms a generalized state space system, in which the volumes are treated as intervals rather than constants. To apply the interval-based solution algorithm for the dynamic O-D estimation, the most challenging part is to ensure that the natural constraints of the system model as shown in Equations (1) and (2) can be held during the recursive computing process. As reported in the literature (Nihan and Davis, 1987), most studies employ the methods of truncation and normalization to tackle this difficult issue. To incorporate the above concepts in developing the algorithm for the proposed formulations, one needs to modify the truncation and normalization process as follows:

- Modified Truncation Process

Note that the purpose of the truncation is to find the largest step of improvement for the unknown parameters b and σ for the next time interval so that the O-D proportions, b_{ij}, can still satisfy Equation (1).

To incorporate this concept in developing the solution algorithm, it is essential that the truncation process be restated as follows:

$$\begin{bmatrix} b_n^I \\ \sigma_n^I \end{bmatrix} = \begin{bmatrix} b_{n-1}^I \\ \sigma_{n-1}^I \end{bmatrix} + \alpha' \cdot \delta_n^I \cdot g_n^I, \text{ where } \alpha' = \underset{0 \le \alpha \le 1}{MAX}[\alpha \mid 0 \le [b_{n-1}^I] + \alpha \cdot \delta_n^I \cdot g_n^I \le 1] \quad (14)$$

where α is a value, which represents the largest step of improvement so that all the possible values of b_{ij} can satisfy Equation (1).

- Modified Normalization Process

The normalization step is employed to satisfy Equation (2), which means that the sum of the O-D proportions from the same origin is equal to one. This constraint is not applicable to the interval formulations. Since b_{ij}^I is an interval, many possible combinations could satisfy this constraint. It is not necessary that either the sum of the lower bound or the sum of the upper bound be equal to one as there exists several possible combinations. In such a case, the estimated O-D proportions are valid as long as there exist combinations that can satisfy the constraint. Hence, one can formulate such relationships with the following equations:

$$0 \le \sum_{j=m+1}^{N} \underline{b}_{mj}^{n-1} \le 1, \qquad i = 0, 1, \ldots, N\text{-}1 \quad (15)$$

$$\sum_{j=m+1}^{N} \bar{b}_{mj}^{n-1} \ge 1, \qquad i = 0, 1, \ldots, N\text{-}1 \quad (16)$$

The Equation (15) is to ensure that the sum of the lower bound, $\underline{b}_{mj}^{n-1}$, for O-D proportions with the same origin is equal to or less than one, and Equation (16) is to let the sum of the upper bound, $\bar{b}_{mj}^{n-1}$, for O-D proportions with the same origin be equal or larger than one. If all the O-D proportions satisfy these two constraints, it is guaranteed that there exists at least one combination that can satisfy Equation (2).

With Equations (9) and (13), one can construct a dynamic model for O-D estimation that takes into account potential measurement errors. This formulation system treats the traffic volumes as intervals so it consists of the upper and lower bound in the system equations. To solve such formulations, one may simply employ the standard solution algorithm, such as the standard Kalman filtering algorithm, for these two system equations. However, the results with those existing estimation algorithms do not encompass all possible optimal solutions for the interval system formulations. Hence, this study employs the *interval Kalman filtering scheme* (Chen, et al., 1997) to develop an interval-based solution algorithm.

Due to the nonlinear nature of the formulations and the concern of computing efficiency, this study has employed sequential extended Kalman filtering algorithm (Chui and Chen, 1999) and Gumbel distribution (an approximation of normal distribution) to develop a solution algorithm. A major part of the interval-based solution algorithm for the formulations of Equations (9) and (13) with the Interval Kalman Filtering method is presented below:

For j=1, 2, …, 2N-1

- $g_j^I = P_{j-1}^I (f_j^I)^T [f_j P_{j-1} (f_{j-1})^T + r_j]^{-1}$
- $P_n^I = P_{n-1}^I - g_n^I f_n^I P_{n-1}^I$
- $\delta_{n-1}^I = y_n^I(k) - f_n^I b^I(k-1)$
- $\alpha' = \underset{0 \le \alpha \le 1}{MAX}[\alpha \mid (14) \sim (16)]$, set $\begin{bmatrix} b_n^I \\ \sigma_n^I \end{bmatrix} = \begin{bmatrix} b_{n-1}^I \\ \sigma_{n-1}^I \end{bmatrix} + \alpha' \cdot \delta_n^I \cdot g_n^I$

Numerical Examples

This section presents the numerical evaluation results of the proposed algorithm using the I-95 northbound freeway corridor between two major beltways, I-495 and I-695. For computing efficiency,

each interchange is represented with a pair of on-ramp and off-ramp, and the network is thus reduced to 7 pairs of on-ramps and off-ramps, and 36 O-D sets as shown in Figure 2.

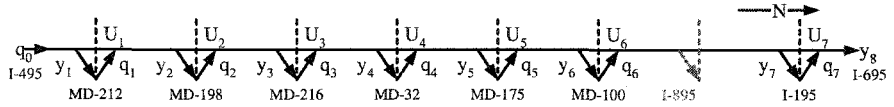


Figure 2. A Graphical Illustration of the Main Interchanges for the I-95 Corridor

To generate a meaningful data set for numerical analysis, the example freeway system under the assigned time-series O-D percentages was simulated with AIMSUN 4.0 (TSS-Transportation Simulation Systems, 2001), to produce time-dependent link traffic volumes. The simulation was executed for one hour using the dynamic O-Ds at an interval of 2 minutes. Following the simulation, one can obtain volume information for each link and the average travel time information for each O-D pair. In this example, the ramp/mainline volumes are added with measurement errors for up to 10%. To conduct an evaluation for the proposed algorithms, a series of time-varying O-D sets estimated with the extended Kalman filtering algorithm (i.e., Algorithm-1) presented by Lin and Chang (Lin and Chang, 2005) is set as a basis to compare with the results from the proposed interval-based algorithm (i.e., Algorithm-2). Figure 3 illustrates the graphical results from Algorithm-1 and Algorithm-2 along with the true O-D set. As shown in Figure 3, the interval-based algorithm (Algorithm-2) outperforms Algorithm-1.

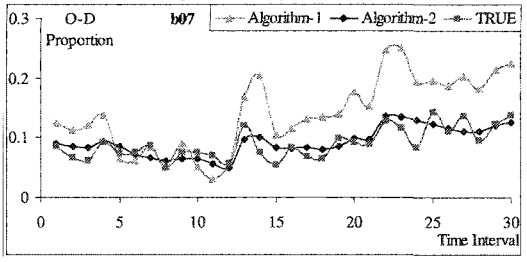


Figure 3. The Estimation Results for Scenario-2 (b07)

Table 1 shows the statistical results, of which the absolute error is computed for each O-D pair in every time interval. In addition to the time-average absolute error for every O-D pair, the maximum and minimum are chosen as the evaluation criteria. As shown in Table 1, the proposed algorithm (Algorithm-2) outperforms the commonly used algorithm (Algorithm-1) with the improvements of more than 30%.

Table 1. The Statistical Results and Improvements

Algorithm-1			Algorithm-2			Improvement		
Average	Max.	Min.	Average	Max.	Min.	Average	Max.	Min.
0.0679	0.1433	0.0225	0.0432	0.0968	0.0121	**36.38%**	**32.45%**	**46.22%**

Conclusions and Recommendations

This study has developed a robust algorithm to contend with the inevitable measurement errors embedded in most traffic sensors. This study has also performed the evaluation of the proposed algorithm using a large freeway network, the I-95 freeway corridor between I-495 and I-695 in Maryland. The results have revealed that the estimated time-varying O-D proportions for this example freeway corridor are more reliable due to the use of the interval-based algorithm. One of the critical issues remains to be investigated in the area of dynamic O-D estimation is to compute the reliability level a model can provide with the available information.

References

Ashok, K., and Ben-Akiva, M.E., "Dynamic Origin-Destination Matrix Estimation for Real-Time Traffic Management Systems," *Trans. and Traffic Theory*, pp. 465-484, 1993.

Bell, M.G.H., "The Estimation of Origin-Destination Matrices by Constrained Generalized Least Squares," *Trans. Research B*, Vol. 25, No. 1, pp. 13-22, 1991a.

Bell, M.G.H., "The Real-Time Estimation of Origin-Destination Flows in the Presence of Platoon Dispersion," *Trans. Research B*, Vol. 25, No.2/3, pp. 115-125, 1991b.

Cascetta, E. "Estimation of Trip Matrices from Traffic Counts and Survey Data: A Generalized Least Squares Estimator," *Trans. Research B*, Vol. 18 No. 4/5, pp. 289-299, 1984.

Cascetta, E. Inaudi, D., and Marguis, G., "Dynamic Estimators of Origin-Destination Matrices using Traffic Counts, " *Trans. Science*, Vol. 27, pp. 363-373, 1993.

Chang, G-L, and Tao, X., "An Integrated Model for Estimating Time-Varying Network Origin-Destination Distributions," *Trans. Research A*, Vol. 33, No. 5, pp. 381-399, 1999.

Chang, G-L, and Wu, J., "Recursive Estimations of Time-Varying Origin-Destination Flows from Traffic Counts in Freeway Corridors," *Trans. Research B*, Vol. 28, No. 2, pp. 141-160, 1994.

Chen, G., Wang, J., and Ghieh, L. S., "Interval Kalman Filtering," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 33, No. 1, pp. 250-258, 1997.

Chui, C. K, and Chen, G., *Kalman Filtering with Real-time Applications*, Springer-Verlag, 1999.

Cremer, M., and Keller, H., "A New Class of Dynamic Methods for the Identification of Origin-Destination Flows," *Trans. Research B*, Vol. 21, No. 2, pp. 117-132, 1987.

Cremer, M., and Keller, H., "A Systems Dynamics Approach to the Estimation of Entry and Exit O-D Flows," *the Pro. 9th Inter. Symposium on Trans. and Traffic Theory*, pp. 431-450, 1984.

Hazelton, M., "Estimation of Origin-Destination Matrices from Link Flows on Uncongested Networks," *Trans. Research B*, Vol. 34, No. 7, pp. 549-566, 2000.

Keller, H., and Ploss, G. "Real-time Identification of O-D Network flow from Counts for Urban Traffic Control," *Trans. and Traffic Theory*, pp. 267-284, 1987.

Lin, P. W., and Chang, G-L, "A Decomposition Algorithm for Estimating the Initial O-D Matrix in a Large Freeway Network," Submitted to *the IEEE Transactions on Intelligent Transportation Systems*, September 2005.

Lin, P-W, and Chang, G-L., "A Robust Model for Estimating Freeway Dynamic Origin-Destination Matrix," *Transportation Research Record*, No. 1923, pp. 110-118, December 2005.

Nihan, N.L., and Davis, G. A., "Recursive Estimation of Origin-Destination Matrices from Input/Output Counts," *Trans. Research B*, Vol. 21, No. 2, pp. 149-163, 1987.

Sherali, H. D., and Park, T., "Estimation of Dynamic Origin-Destination Trip Tables for a General Network," *Trans. Research B*, Vol. 35, No. 3, pp. 217-235, 2001.

Tavana, H. and Mahmassani, H., "Estimation of Dynamic Origin-Destination Flows from Sensor Data Using Bi-level Optimization Method," *the Pro. 80th TRB Annual Meeting*, 2000.

TSS-Transportation Simulation Systems, *GETRAM Operation Traffic Simulation Evrironment* – TEDI Version 4.0 User Manual, 2001.

Wu, J., and Chang, G-L, "Estimation of Time-Varying Origin-Destination Distributions with Dynamic Screenline Flows," *Trans. Research B*, Vol. 30, No. 4, pp. 277-290, 1996.

Zhou, X., and Mahmassani, H. S., "Dynamic Origin-Destination Demand Estimation Using Automatic Vehicle Identification Data," *the Pro. 84th TRB Annual Meeting*, 2005.

Estimation of Freeway Travel Time Based on Sparsely Distributed Detectors

Ying Liu and Gang-Len Chang

Department of Civil and Environmental Engineering, University of Maryland at College Park, College Park, MD, 20742; Tel: (301) 405-6959; Email: lyng@wam.umd.edu

Abstract

This study presents a new approach for estimating travel time information along freeway corridors, which experience recurrent congestions but have only a limited number of available detectors due to budget constraints. The proposed iterative estimation procedure, based on a set of empirically calibrated regression models, intends to rebuild the relations between travel times and accumulated flows within each segment of the target freeway corridor. To evaluate the effectiveness of the proposed methodology, this study has conducted extensive numerical experiments with simulated data from a CORSIM simulator. Experimental results under various traffic volume levels have revealed that the proposed method offers a promising property for use in travel time estimation based on sparsely distributed sensors.

Research Background

As a direct indicator of network congestion level, travel time information plays an important role in travelers' route and departure time choices. This research project aims to build an online travel time prediction system for I-70, one of the major commuting corridors in the Baltimore region. In reviewing travel time prediction methodologies (Rilett and Park, 1999; Chien and Chen, 2001; Chien and Kuchipudi, 2003), it is clear that most of those in the literature need extensive and reliable historical travel time data. The limited number of travel time samples available in this project can hardly meet such requirements, which has necessitated an effective travel time estimation approach.

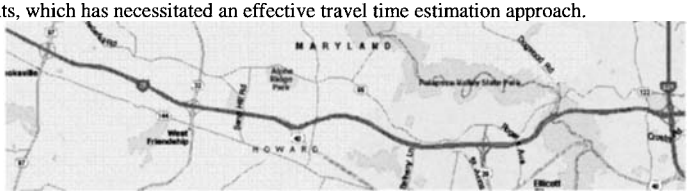


Figure 1. Illustration of the Study Area

Most existing travel time estimation methods in the literature can be classified into three categories. The first category provides the estimated travel time by comparing traffic measurements from up/downstream detectors, such as recognizing platoons, employing the flow conservation law, checking flow correlations, or building the regression relations between up/downstream flows (Kuhne and Immes, 1993; Nam and Drew, 1996; Dailey, 1993; Petty, etc., 1998). The second category uses point speed, either measured or estimated from detector data, to generate a section wide speed for computing the travel time (Smith, etc., 2004; coifman, 2002; Lindveld, etc., 2000; Van Lint and Van der Zijpp, 2002). The last category tries to reconstruct the relations between travel time and detected flow, speed, occupancy data with macroscopic flow equations or other models (Oh, etc., 2003; Sisiopiku, etc., 1994).

The I-70 segment under study experiences heavy congestions during morning/evening peak hours, which has excluded the first category of estimation approaches for this project. Due to budget constraints, the study area of more than 20 miles is only covered with 10 roadside detectors. Speed-based approaches cannot provide reliable travel time estimates under such long detector spacing (usually half-mile or 500m

in literature) (Liu, etc. 2005). Thus, this study will follow the direction of the third category and try to rebuild the relations between travel times and detector measurements with a set of regression models and an iterative estimation algorithm.

The rest of the paper is organized as follows. Next section analyzes Type I freeway segments with permanent downstream bottlenecks and tests two different regression models. Similarly, Section 3 discusses Type II freeway segments with temporary downstream bottlenecks such as queue spillbacks, and proposes the regression model as well as the iterative estimation procedures. Some operational issues such as effects of the sample size are discussed in Section 4. Concluding comments constitute the core of the last section.

Type I Segments with permanent downstream bottlenecks

Since travel time of a vehicle between two points is a segment-based parameter, the authors intend to select segment-related parameters as explanatory variables for rebuilding the relationship between travel time and detector measurements. Unfortunately, the flow/speed/occupancy data directly from detectors are only for a single point. Thus, an alternative is to use the number of vehicles within the segment as the regressor, since it is the simplest segment-related parameter obtainable from detector measurements.

On Type I segment whose congestion is due to a permanent downstream bottleneck (Figure 2), the relationship between travel time and the number of vehicles within the segment can be shown with Equation (1) if vehicles are assumed to follow the First-In-First-Out rule.

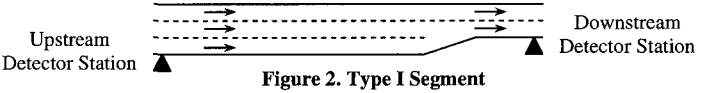


Figure 2. Type I Segment

$$tt(k) = \max\{ft, [n + af(k)]/C\} \tag{1}$$

Here, $tt(k)$ is the travel time of vehicles departed at time k; ft is the free flow travel time; n is the number of vehicles on the segment at time zero; C is the capacity of the downstream bottleneck; $af(k)$ is the accumulated number of vehicles at time k. Assuming that the flow measurements at upstream and downstream detectors at time k are $f_u(k)$ and $f_d(k)$, one has

$$af(k) = \sum_{i=1}^{k-1}[f_u(i) - f_d(i)] \tag{2}$$

Equation (1) can be rewritten as a two-segment expression as in Equation (3). Thus one has to fit a segmented curve and estimate the join point $C \times ft - n$.

$$tt(k) = \begin{cases} ft & : af(k) \le C \times ft - n \\ n/C + af(k)/C & : af(k) > C \times ft - n \end{cases} \tag{3}$$

Based on several related studies in the literature (Quandt, 1958; Hudson, 1966; Kim, etc., 2000), this paper employs Hudson's approach to find the Least Squares (LS) solution. For two submodels $f_1(x;\beta_1)$ and $f_2(x;\beta_2)$ joined together at $x = a$, the LS solution includes vectors $\beta_1 = \hat{\beta}_1$, $\beta_2 = \hat{\beta}_2$ and real values $a = \hat{a}$, $I = \hat{I}$, which minimize the Residual Sum of Squares (RSS):

$$R(\beta_1, \beta_2, a, I) = \sum_{i=1}^{I}[y_i - f_1(x_i, \beta_1)]^2 + \sum_{i=I+1}^{n}[y_i - f_2(x_i, \beta_2)]^2 \tag{4}$$
$$s.t. \quad f_1(\hat{a}, \hat{\beta}_1) = f_2(\hat{a}, \hat{\beta}_2), \qquad x_{\hat{I}} \le \hat{a} \le x_{\hat{I}+1}$$

The fitting procedures are stated below:

- Step 0: Reorganize the data points by the value of x
- Step 1: Assume $x_i < \hat{a} < x_{i+1}$ for some i. For all $I, 3 \le I \le n-3$ and $x_I \ne x_{I+1}$

- Perform regression on data points $\{x_1,...,x_I\}$ and $\{x_{I+1},...,x_n\}$ respectively
- Solve the join point by $f_1(a,\hat{\beta}_1)=f_2(a,\hat{\beta}_2)$
- If $x_{\hat{I}}<\hat{a}<x_{\hat{I}+1}$, compute $T_I=R(\beta_1,\beta_2,a,I)$. Otherwise $T_I=\infty$

- Step 2: Assume $\hat{a}=x_i$ for some i. For all $I, 3\le I\le n-3$, $x_I\ne x_{I+1}$ and $\{x_I,x_{I+1}\}$ does not contain a valid $\hat{a}$ in Step 1 estimation
  - Perform regression on data points $\{x_1,...,x_I\}$ and $\{x_{I+1},...,x_n\}$ subject to $f_1(x_I,\beta_1)=f_2(x_I,\beta_2)$. (Plackett, 1960)
  - Compute $S_I=R(\beta_1,\beta_2,a,I)$.
- Step 3: The overall solution is the one that yields the minimal RSS T or S

To test the reliability of the proposed segmented curve in relating travel time to the accumulated number of vehicles, the authors have constructed a highway segment of 9000ft with CORSIM, a microscopic simulation package. The three-lane segment is reduced to two lanes at the downstream end. During a simulation period of 6.5 hours, the output for analysis includes the upstream/downstream detector measurements and average travel time for every thirty seconds. The data from the first three hours are used to fit the model. Excluding the initialization period, there are a total of 352 data points. A program is written in Visual Basic 5.0 to perform the iterative fitting procedure. The minimal RSS from Step 1 is 293,001.4, whereas the minimal RSS from step 2 is 293,105.1. The best fitting parameters is as follows

$$\{\hat{\beta}_1,\hat{\beta}_2,\hat{a},\hat{I}\}=\{98.111,\ \begin{pmatrix}54.670\\0.899\end{pmatrix},\ 48.315,\ 63\}$$

Note that among all numerical experiments with modest congestions, the join point $x=a$ is always much smaller than the maximal accumulated number of vehicles. This implies that the first segment of the curve is short, which justifies the simplified one-segment model $tt(k)=a_1+a_2*af(k)$ under congestions. Using the same data set, the estimation result is shown in the following table.

Table 1. Fitting of the One-Segment Model: Type I Segment

	Coefficients	SE	t Stat	P-value	R Square	0.987
Intercept	73.021	3.085	23.667	0.000	Adjusted R Square	0.987
X	0.872	0.005	162.083	0.000	RSS	311,049.7

Data from the entire simulation period are used to test both calibrated models. The comparison of actual and estimated travel times is shown in Figure 3. The figure indicates that both models can well capture the change in travel time during the transition and congested periods.

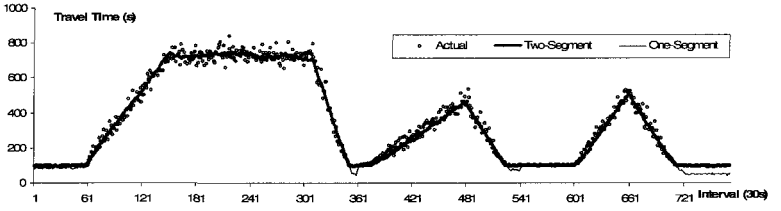


Figure 3. Comparison of Actual and Estimated Travel Time: Type I Segment

Type II Segments with temporary downstream bottlenecks

Type I segment with a permanent downstream bottleneck is mainly controlled by the fixed downstream capacity. This implies a monotonic relationship between travel time and the accumulated number of vehicles within the segment. However, for Type II segment with a temporary downstream

bottleneck such as queue spillbacks, the available downstream capacity will vary with time. Thus, a vehicle's travel time depends not only on the accumulated number of vehicles at its departure time, but also on the downstream capacity while it traverses the segment. For example, a vehicle entering the segment during free flow period may experience longer travel time if the downstream queue starts forming during its travel.

The authors propose to use the measurements of either speed or occupancy from the downstream detector as the indicator of downstream congestions. This paper selects speed measurements for convenience of calibration. The formulation of the proposed model is as follows:

$$tt(k) = a_1 + a_2 \times af(k) / av(k) \quad (5)$$

Here $a_i, i = 1,2$ are the model coefficients to be fitted. $av(k)$ is the average downstream speed during the travel of those vehicles departed at time interval k. Assuming the speed measurement at downstream detector at time k is $v_d(k)$, one has $av(k) = ave\{v_d(t) : t = k,\ldots,k + tt(k)\}$

To test the proposed model, the authors select a 3829ft segment from the I-70 simulator, which is built and empirically calibrated in the microscopic simulation environment CORSIM. The traffic pattern indicates that the queue forms and dissipates both from the downstream. During the simulation period of four hours, the upstream/downstream detector measurements are obtained from simulation output. The average travel time for every thirty seconds is computed by tracing sample vehicles. The first two hours' data, excluding those periods without travel time information, are used to fit the model. The estimation result is shown in the following table.

Table 2. Model Fitting Result: Type II Segment

	Coefficients	SE	t Stat	P-value	R Square	0.973
Intercept	43.941	2.329	18.868	0.000	Adjusted R Square	0.973
X	14.086	0.170	82.858	0.000	Observations	195

Data from the entire simulation period are used to test the calibrated model. Note that since the computation of $av(k)$ involves vehicles' travel time, it is not known until the travel time is determined. Thus, the following iterative estimation procedures are specially designed for such a need.

- Step 0: For departure time k, set $tt_0(k) = tt(k-1)$, if $tt(k-1)$ is available. Otherwise, set $tt_0(k)$ to the estimated free flow travel time.
- Step 1: Compute $av(k) = ave\{v_d(t) : t = k,\ldots,k + tt_0(k)\}$ and $tt(k) = \hat{a}_1 + \hat{a}_2 \times af(k) / av(k)$
- Step 2: If $tt(k)$ and $tt_0(k)$ fall in the same interval or the number of iterations has reached its upper bound, Stop. Otherwise, set $tt_0(k) = tt(k)$ and go to Step 1.

The comparison of actual and estimated travel times is shown in Figure 4. The data points falling on the x-axis imply that there are no sample vehicles departed during the corresponding interval and there is no actual travel time information available. The figure indicates that the proposed model can capture the change in travel time during the transition and congested periods.

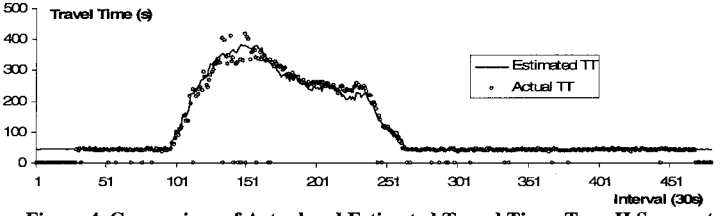


Figure 4. Comparison of Actual and Estimated Travel Time: Type II Segment

Some Operational Issues

Sample size for model calibration

Due to the limited budget, this research project will not have extensive travel time samples for model calibration. Thus, the authors have tested the robustness of the proposed model. With the same simulation experiment for Type II segment Table 3 presents the model calibration results with data points from intervals of 30 seconds, 300 seconds and 900 seconds. As shown in Figure 5, the fitting results are fairly stable even with a very small sample size.

Table 3. Model Fitting Results with Different Sample Sizes

30s	Coefficients	SE	t Stat	P-value	R Square	0.984
Intercept	41.596	0.852	48.799	0.000	Adjusted R Square	0.984
X	14.234	0.089	159.375	0.000	Observations	408
300s	Coefficients	SE	t Stat	P-value	R Square	0.988
Intercept	41.693	2.269	18.373	0.000	Adjusted R Square	0.988
X	14.246	0.245	58.176	0.000	Observations	42
900s	Coefficients	SE	t Stat	P-value	R Square	0.999
Intercept	38.699	1.075	35.999	0.000	Adjusted R Square	0.999
X	15.412	0.140	110.190	0.000	Observations	14

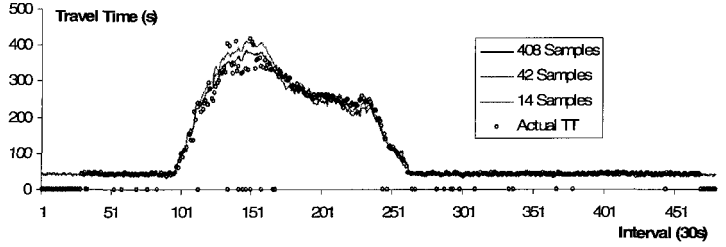


Figure 5. Comparison of Actual and Estimated Travel Time with Different Sample Sizes

Detector Location and Measurement Error

The proposed regression models are site specific. The regression coefficients will depend on various geometric features and traffic characteristics, such as the section length, road gradient, lane width and vehicle composition. For a given freeway segment, the proposed models use the accumulated number of vehicles within the road segment as one major regressor. This number is computed based on the flow measurements of the two end detectors. Thus, there should be no on/off ramps between these detectors.

To deal with the potential measurement errors, the accumulated number of vehicles can be recorded for several days. For an early morning time interval when the traffic is very low, the accumulated number of vehicles should be around a certain value over different days. If an apparent increasing/decreasing trend is observed, one can compute the average increase/decrease rate per interval and thus to adjust the accumulated number of vehicles.

Conclusions

This study has proposed two regression-based approaches to estimate travel time information, respectively, for freeway segments containing congestions due to permanent or temporary downstream bottlenecks. The regression models are designed to rebuild the relations between travel times and accumulated flows within each segment. The numerical experiments with simulated data from a CORSIM

simulator have demonstrated the potential of the proposed methodology, especially with respect to its following features:

- Reliable estimation results under various traffic volume levels (i.e., free flow and congestion and even the transition periods).
- Robustness in taking full advantages of available travel time samples for calibration.
- Reliable travel time estimates even under long detector spacing (e.g., 9000ft and 3829ft in the numerical study), which can reduce the required number of detectors for travel time estimation.

References

Chien, S., and Chen, M., "Dynamic freeway travel time prediction using probe vehicle data: Link-based vs. path-based." *Transportation Research Record 1768*, pp.157–161, 2001

Chien, S., Kuchipudi, C.M., "Dynamic Travel Time Prediction with Real-Time and Historic Data," *Journal of Transportation Engineering*, Vol. 129, No. 6, pp. 608-616, 2003

Coifman, B., "Estimating travel times and vehicle trajectories on freeways using dual loop detectors," *Transportation research A*, vol. 36, pp. 351-364, 2002

Dailey, D.J., "Travel time estimation using cross-correlation techniques," *Transportation research B*, vol. 27, pp. 97-107, 1993

Hudson, D.J., "Fitting Segmented Curves Whose Join Points Have To Be Estimated," *Journal of the American Statistical Association*, v.61, pp.1097-1129, 1966

Kim, H.J., Fay, M.P., Feuer, E.J., and Midthune, D.N., "Permutation tests for joinpoint regression with applications to cancer rates," *Statistics in Medicine*, v.19, pp.335-351, 2000

Kuhne, R.D., and Immes, S., "Freeway control systems for using section-related traffic variable detection," *Proceeding of the ASCE third AATT conference*, pp. 57-62, 1993

Lindveld, Ch.D.R., Thijs, R., Bovy, P.H.L., and Van der Zijpp, N.J., "Evaluation of on-line travel time estimators and predictors," *Transportation Research Record 1719*, pp. 45-53, 2000

Liu, Y., Lai, X., and Chang, G.L., "Detector Placement Strategies for Freeway Travel Time Estimation," working paper, University of Maryland, 2005

Nam, D.H., and Drew, D.R., "Traffic dynamics: method fro estimating freeway travel times in real time from flow measurement," *Journal of Transportation Engineering*, v.122, pp.185-191, 1996

Oh J.S., Jayakrishnan, R., and Recker, W. "Section travel time estimation from point detection data," *accepted at the 82nd Annual Meeting of the TRB*, 2003

Petty, L.F., Bickel, P., Ostland, M., Rice, J., Schoenberg, F., Jiang, J., and Ritov, Y., "Accurate estimation of travel times from single loop detectors," *Transportation research A*, vol. 32, pp. 1-17, 1998

Plackett, R.L., *Principles Of Regression Analysis*. Oxford: at the Clarendon Press, 1960

Quandt, R.E., "the estimation of the parameters of a linear regression system obeying two separate regimes," *Journal of the American Statistical Association*, v.53, pp.873-880, 1958

Rilett, L., and Park, D., "Direct forecasting of freeway corridor travel times using spectral basis neural networks." *Transportation Research Record 1617*, pp. 163–170, 1999

Sisiopiku, V.P., Rouphail, N.M., and Santiago, A. "Analysis of the Correlation between Arterial Travel Time and Detector Data from Simulation and Field Studies", *Transportation Research Record 1457*, Washington, D.C., pp. 166-173, 1994

Smith, B.L., Holt, R.B., and Park, B.B., "Travel time estimation for urban freeway performance measurement: understanding and improving upon the extrapolation method," *accepted at the 83rd Annual Meeting of the Transportation Research Board*, 2004

Van Lint, J.W.C., and Van der Zijpp, N.J., "An improved travel-time estimation algorithm using dual loop detectors," *accepted at the 81st Annual Meeting of the TRB*, 2002

Evaluation of Urban Traffic Management in China

Jiao Pengpeng[1], Lu Huapu[2], Noboru Harata[3]

[1]Postdoctoral Researcher, Department of Urban Engineering, the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. Tel: +81-3-5841-6234, Fax: +81-3-5841-8527, Email: jiao_pengpeng@ut.t.u-tokyo.ac.jp (until Sep. 2005, Department of Civil Engineering, Tsinghua University, Beijing 100084, China)
[2]Professor, Department of Civil Engineering, Tsinghua University, Beijing, 100084 China. Tel: +86-10-62795339, Fax: +86-10-62795339, Email: luhp@mail.tsinghua.edu.cn
[3]Professor, Department of Urban Engineering, the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. Tel: +81-3-5841-6233, Fax: +81-3-5841-8527, Email: nhara@ut.t.u-tokyo.ac.jp

Abstract
Urban traffic management is important for the growth of national economy. An integrated evaluation system is first proposed to evaluate the level of traffic management of each Chinese city, as well as to guide the developing direction of Chinese urban traffic management. An evaluation model is then formulated to evaluate the national economic benefit produced by the development of traffic management. The benefit consists of three aspects: time saving, energy saving and traffic accident decrease. All key parameters in the model are discussed thoroughly. Field data of 178 Chinese cities in 2000, 2001 and 2002 are collected, and the national economic benefit of these three years are evaluated. The results show that the integrated evaluation has greatly improved the level of Chinese urban traffic management, and has produced a large amount of national economic benefit.

Introduction
With the quick development of urbanization and motorization in China, the travel demand is increasing dramatically. Transportation system, especially urban traffic management system, is very important for the growth of Chinese national economy. In order to evaluate the level of traffic management of Chinese cities, also to guide the developing direction of traffic management, an integrated evaluation system of urban traffic management is necessary. In this paper, such an evaluation system will be proposed, including 9 aspects, such as policy, planning, infrastructure, etc. There are 63 indices in all. This evaluation system has been put in practice in China since 2000, and has produced great national economic benefit.

To evaluate the national economic benefit produced by advanced traffic management, an evaluation model is formulated. The total benefit is divided into three aspects: time saving, energy saving and traffic accident decrease. Each aspect is formulated as a model, and an integrated evaluation model is then formulated. All key parameters are discussed thoroughly.

Based on the evaluation system of urban traffic management, the indices of 178 Chinese cities in 2000, 2001 and 2002 are collected, and the national economic benefits of these three years are then evaluated. The results clearly indicate that the integrated evaluation system has greatly improved the level of urban traffic management in China, and has produced a large amount of national economic benefit.

This paper consists of following sections. The evaluation system is presented in Section 2. The evaluation model for national economic benefit is formulated in Section 3. A case study based on field data is reported in Section 4. Conclusions and further researches are summarized in the last section.

Evaluation System for Chinese Urban Traffic Management

To evaluate the level of traffic management of each city, also to guide the developing direction of Chinese urban traffic management, an integrated evaluation system is proposed, including nine aspects. The framework of the evaluation system is shown in Figure 1.

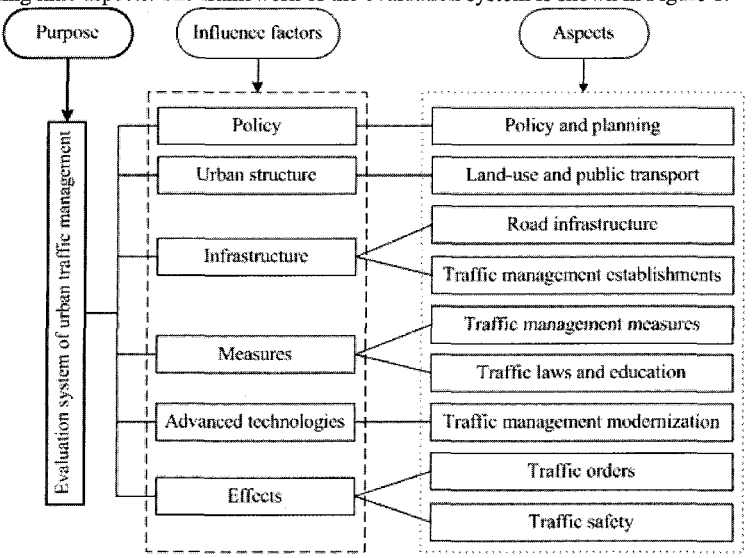


Figure 1. An integrated evaluation system for Chinese urban traffic management

Under each aspect, there are some detailed indices. For example, there are eight indices under the "Land-use and public transport", including traffic influence assessment, policy of public transport priority, modal split of public transport, number of buses per capita, safe operation distance, percentage of buses parking illegally, updating percentage of buses and percentage of taxis without passengers. Each index has its own definition, unit and calculation method, and can be obtained through survey respectively. Due to the page length limit, we have to omit them here.

Based on above evaluation system, the value of each index is first collected through survey, the weight of each index is determined by analytic hierarchy process (AHP), and then the evaluation result can be obtained using comprehensive fuzzy decision making method (FDM) (Hwang and Yoon, 1981). Based on this, we can evaluate the level of each city's traffic management.

Evaluation of National Economic Benefit of Urban Traffic Management

Under the guidance of above evaluation system, urban traffic management is developing very fast, and it has produced a large amount of social economic benefit for Chinese cities. Evaluation of such benefit will guide the developing direction of Chinese traffic management, and will produce even bigger benefit. Here, the evaluation is divided into three aspects: time saving, energy saving and traffic accident decrease.

1) Time saving

With the improvement of traffic management, the average speed of vehicles increases, and the personal travel time decreases. From the aspect of national economic evaluation, time is a kind of resources, the save of passengers' travel time increases the working time, so more GDP will be produced. Considering the economic benefit produced

by the save of travel time, the approximate evaluation model can be formulated as following:

$$B_T(t) = 0.5 b_t \sum_{i=1}^{n} \left\{ N_i(t) K_i(t) L_i(t) \left[\frac{1}{V_i(t_0)} - \frac{1}{V_i(t)} \right] \right\} \tag{1}$$

where $B_T(t)$ is the economic benefit produced by time saving in year t (RMB/year); b_t is the time value of passengers in year t (RMB/h*capita); $L_i(t)$ is the total travel distance of vehicle type i in year t (km); $V_i(t)$ and $V_i(t_0)$ is average travel speed of vehicle type i in year t and t_0 respectively (km/h); t_0 is the base year, in this paper it denotes year 1999, i.e., the last year before above evaluation system is put in practice in China; $N_i(t)$ is the number of ownerships of vehicle type i in year t; $K_i(t)$ is the average passenger number of vehicle type i in year t; n is the number of vehicle types.

According to Chinese standard (Chinese National Development and Reform Commission, 1998), passengers' time value is obtained from the GDP per capita, and half of the saved time will be used for work, so the coefficient 0.5 is taken into account in the evaluation model.

The time value b_t can be obtained through the following equation:

$$b_t = \frac{I_r(t)}{2000} \tag{2}$$

where $I_r(t)$ is the GDP per capita in year t (RMB/capita); the 2000 in the denominator is from $(52 \times 5 - 10) \times 8 = 2000$, where 52 means there are 52 weeks in one year, people work for 5 days per week, and 8 hours per day. 10 means 10 days of rest in one year.

The average passenger numbers $K_i(t)$ is usually obtained from sample survey, it can be formulated as following equation:

$$K_i(t) = \frac{P_i(t)}{M_i(t)} \tag{3}$$

where $P_i(t)$ is the total number of passengers counted in the survey, and $M_i(t)$ is the total number of vehicles counted in the survey.

From (1), (2) and (3), one can obtain the economic benefit produced by time saving.

2) Energy saving

The calculation of the benefit produced by energy saving mainly depends on the relation between vehicle speed and energy consumption. Based on the relation, the energy consumption per kilometer under different speed can be obtained, thus one can get the number of energy saving according to the change of vehicle speed. The evaluation model is formulated as following:

$$B_F(t) = C_F \sum_{i=1}^{n} \{ N_i(t) L_i(t) [Q_i(t_0) - Q_i(t)] \} \tag{4}$$

where $B_F(t)$ is the economic benefit produced by energy saving in year t; C_F is the

energy price (3.4 RMB/L in this paper); $Q_i(t)$ and $Q_i(t_0)$ are energy consumption per kilometer of vehicle type i in year t and t_0 respectively (L/km).

The energy consumption is different according to different vehicle types. In this paper, vehicles are classified into 4 types, small car, medium-size car, big car and truck, i.e., $i = 1,2,3,4$. Furthermore, for the calculation of energy consumption, we aggregate them into two types. The calculation method is as following (Wang et al., 2001):

$$Q_1 = 14.32\exp(-0.0003L_l) \times (1 + 4.75\exp(-0.18V_1)) \tag{5}$$

$$Q_j = 39.67L_l - 0.14(1 + 28.12V_j^{-1.79}) \quad j = 2,3,4 \tag{6}$$

where L_l is the average distance between two adjacent intersections (km).

From (4), (5) and (6), one can obtain the national economic benefit produced by energy saving.

3) Traffic accident decrease

For the national economic benefit produced by traffic accident decrease, since there is the statistical data of economic loss caused by traffic accident every year, we can get it directly by the following equation:

$$B_A = S_0 - S_t \tag{7}$$

where B_A is the economic benefit produced by traffic accident decrease; S_0 and S_t are economic loss in the base year and year t respectively (RMB).

Thus the evaluation model of national economic benefit produced by traffic management can be formulated as following:

$$B = B_T + B_F + B_A \tag{8}$$

Case Study

Based on the evaluation system in this paper, the indices of 178 Chinese cities are collected in 2000, 2001 and 2002 through the ministry of public security and the ministry of construction in China. These 178 cities are from 23 provinces in China, and the total population is 120.13 million, total number of vehicles is 12.33 million. The population distribution of these 178 cities is shown in Figure 2, from which one can find out that the sample cities are reasonable for China.

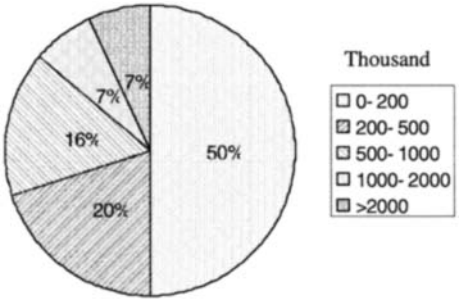


Figure 2. Distribution of sample cities

Using above evaluation method, the veritable national economic benefit is evaluated using actual data. The evaluated national economic benefit is shown in Table 1, and the proportions of three aspects are further shown in Figure 3. Figure 4 shows the distribution of economic per capita among different provinces in China, and the GDP per capita in 2002 (National Bureau of Statistics of China, 2003) is also shown for comparison.

Table 1. Evaluation Results of National Economic Benefit

	2000	2001	2002	Total
Time saving (billion RMB)	3.18	5.93	7.62	16.73
Energy saving (billion RMB)	1.40	3.09	4.20	8.69
Traffic accident decrease (billion RMB)	0.75	1.63	3.19	5.57
Total	5.33	10.65	15.01	30.99
Percentage of GDP (%)	0.22	0.39	0.52	0.39
Benefit per capita (RMB)	44.30	88.71	124.97	257.98
Benefit per vehicle (RMB)	567.97	987.20	1217.40	2772.57

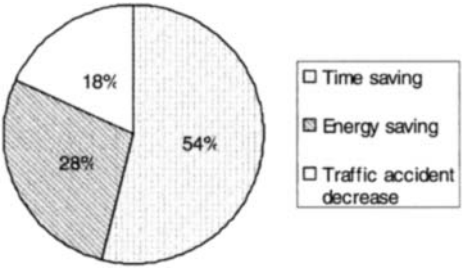


Figure 3. Distribution of economic benefit among three aspects

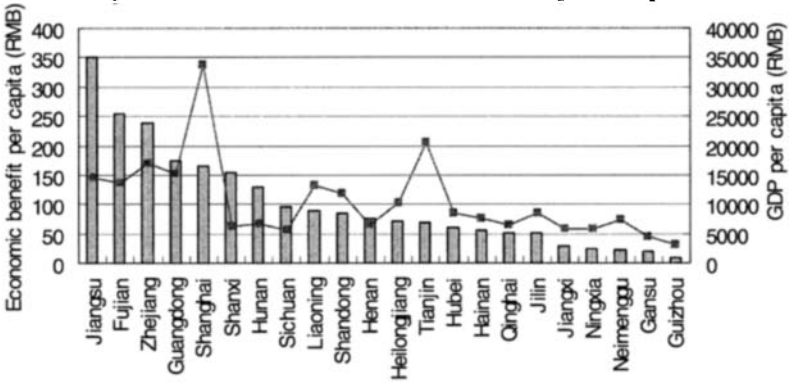


Figure 4. Distribution of economic benefit among different provinces

From above tables and figures, one can find out the following results:

1) Under the guidance of the evaluation system, the level of traffic management in China is developing very fast, and a large amount of national economic benefit has been produced.

2) With the development of traffic management, the national economic benefit turns bigger and bigger through these years.

3) The benefit produced by time saving occupies majority of three aspects. That is to say, the improvement of travel speed is the most effective way for national economy.

4) The economic benefit differs according to different areas in China. Generally, the

benefit is bigger in those developed provinces.

Conclusions

In this paper, an integrated evaluation system for Chinese urban traffic management has been proposed. This evaluation system was put in practice in China since year 2000, and has caused great improvement of the level of urban traffic management in China. To evaluate the national economic benefit produced by advanced traffic management, an evaluation model is formulated, including evaluation model for benefit produced by time saving, energy saving and traffic accident decrease. The key parameters are discussed thoroughly. The field data of 178 Chinese cities in 2000, 2001 and 2002 are collected, and the evaluation results based on them show that the integrated evaluation system has greatly improved the level of urban traffic management in China, and has produced a large number of national economic benefit.

Further research should be directed towards three aspects: The first is to adjust the evaluation system of urban traffic management according to Chinese situation. The second is to collect data from other Chinese provinces for more comprehensive economic benefit evaluation. And the third, to abstract the key factors that influence the level of traffic management, and to further guide the development of Chinese traffic management more efficiently.

Acknowledgement

This research was supported by Chinese Ministry of Public Security research fund and the JSPS (Japan Society for the Promotion of Science) postdoctoral fellowship partially. The authors would like to thank the anonymous referees for the careful review of this paper.

References

Chinese National Development and Reform Commission. (1998) *Economic evaluation method and parameters of construction project*, China Planning Press, Beijing. (in Chinese)

Hwang, C L. and Yoon, K. (1981) *Multiple attribute decision making*, Springer-Verlag, New York

National Bureau of Statistics of China. (2003) China statistical yearbook, China Statistical Press,, Beijing (in Chinese)

Wang, W. et al. (2001) *Analysis method of energy consumption and environment influence of urban transportation system*, Science Press, Beijing. (in Chinese)

Evaluation of Traffic Management Plans in CBD Area within Baghdad City

A. Al-Azzawi[1], N. Al-Saoudi[2], and M. Abdul-Ghani[3]

[1] Department of Highway and Transportation Engineering, Al-Mustansiriyah University; Bab Al-Muadham Post Office; P.O.Box 14099; Baghdad / Iraq; Mobile: 00964 7901 611 906; E-mail: afafazzawi@yahoo.com
[2 & 3] Building and Construction Engineering Dept., University of Technonlogy, Baghdad / Iraq.

Abstract
Generally, road users in most urban areas suffer from an increasing congestion and inconvenice. Mainly, because of the definicy in satisfying the eccentuated demand within these areas. To evalaute and enhance the traffic flow performace within Baghdad city, three arterial streets were selected. These are located within the Centeral Business district (CBD). Evaluating the existing traffic coditions has revealed that the area undergos a great deal of deterioration as represented by level of service (LOS) F. Resources constraints have obliged to investigate the traffic management meseaures first. Several operation and geometry alternatives were designed and tested. For the selected case study it appears that urgent and costy treatment would be the proper answer to the current problem.

Traffic Problems

Congestion is one of the most important problems in urban areas. Congestion is defined as traffic conditions that are not accepted by highway users, usually refered to as level of service F, [Highway capacity manual, 2000]. Congested flow has been conceived in two different, but related, ways. The first identifies congested flows with queuing upstream of bottlenecks, and regards it as a region of dense traffic that exists because flow into a roadway section exceeds the capacity downstream. This concept implies that events in congested flow are controlled by conditions downstream from the point of observation. Whereas, the second regards it as flow described by data on the right side of the flow-concentration diagram [Banks, 2000]. On a link with a signalized intersection, this situation is equivalent to drivers in queue who see a red signal at least twice [Kuorau and Tarko, 2000]. The Spillback phenomenon would be an expected condition, especailly at signalized intersection. It occurs when there is high volume of traffic for a particular movement, and it queues into the previous intersection having more demand than can be stored on an approach which creats this spillback problem [Khatib and Judd, 2000]. It's important to note

this problem, since this will creats a situation in which the upstream signal will not be able to allow traffic to enter the intersection. Spillback generates a long delay for the vehicles on all of the affected approaches. Possibly, one way operation system affords immediate and least expensive method of increasing traffic flow within an area. It costs little more than the expense of signs and markings necessary to indicate the direction of flow. Sometimes, it may have an adverse effects where it becomes more difficult to reach the business locations. Most likely, additional signs are required, since one-way signs must be erected at every intersection. Other special signs may be needed such as turn prohibition, do not enter, and lane control signs. Also, considerations must be given when different road geometry exists.

Background

Jrew et. al. [2001] carried out a study with the objectives of evaluation of the level of service at Al-Amal Al-Shabby street (arterial) in the suburb of Baghdad City. Moreover, development of alternative improvement strategies to overcome the traffic operation problems was also investigated. The designated arterial has four unsignalized intersections. At first, the arterial was analyzed as a street with unsignalized intersection (i.e. existing condition). The results showed that signal warrants were satisfied [FHWA, 1988]. Second, the arterial was analyzed as if it had signalized intersections. For this purpose, HCS-1994 program was used to evaluate the level of service. Upgrading of the level of service from (F) to (C) was then obtained. Hougeh and Kuwahara [2001] reported that oversaturated signalized intersections present a challenge for traffic engineers when determining the optimal signal timing. A so called inverse cause and effect procedure is adopted, where both arrival and departure flows at the signalized intersections are presented as smoothed functions as if there were no interruptions from signals. The objective function was to maximize the total vehicular output from the certain congested area subjected to prevailing capacity and operational constraints. Then, the optimized value of departure flow rate (effect) is converted to the value of split (cause). Also, the study deals with the simulation analysis. It is embeded the optimization procedure into the simulation model to be an integrated optimization and simulation to analyze the effect of physical queue. Messer et. al. [2001] stated that traffic congestion during peak periods is prevalent for most urban areas. Urban arterials systems have experienced increasing traffic congestion. Thus, there is a need for effectively managing traffic signal control systems during congested or oversaturated periods. Abu-Lebdeh and Benekhal [2001] developed modules to estimate capacity of oversaturated arterials. The input variables in these modules are capacities of individual intersections, offsets and vehicle queue lengths. The study also presents modules to quantify capacity loss due to blockage caused by downstream queues. The proposed modules showed that when determining arterial capacity in oversaturated conditions it's not sufficient to consider only the capacity of critical intersections, but also consideration must given be to the capacity of critical subsystems. The practical use of the modules was demonstrated for an oversaturated two-intersection system. The results showed that improper setting of offsets can lead to significant capacity loss. Al-Nuaimi [2002] conducted a study with the objective of evaluation of traffic performance within a selected CBD area in Baghdad city, the

capital of Iraq. Furthermore, several improvement strategies to overcome the traffic operation problems are designed and tested. The current and future conditions were considered. The selected study area includes a part of Al-Khulafa street, which is located in Al-Shorga CBD area and consists of Al-Ruhoon, Al-Wathba, and Al-Khullany intersections. The analysis and evaluation throughout the study are performed by using TRANSYT-7F program package. The objective was to identify the level of service for the studied arterial and intersections. The results revealed that the selected traffic facility suffer from serious degradation causing breakdown conditions. Thus, urgent considerations must be given regarding the upgrading of the the level of service. This might be done by implementation of improvement strategies depending on encouragement of TSM (Transportation System Management) measures implementation.

Study areas

The selected street network consists of three main streets in the CBD area of Baghdad City which are Al-Khulafa, Al-Rasheed and Al-Kefah streets. These streets are shown in Figure 1. Al-Khulafa street is a two-way, with three lanes per direction, over a distance of 2.64 km. It starts from Al-Khulafa square in Bab Al-Muadham district and ends at Al-Tahreer square. It is considered as one of the highly congested commercial street within Baghdad City. It has five main intersections. These are: Al-Khulafa square, Al-medan intersection, Al-Ameen intersection, Al-Wathba intersection and Al-Khullany square. The second street is Al-Rasheed street which starts from Bab Al-Muadham square and ends at Al-Tahreer square. Large proportion of this street is operated as two-way street, while the rest is operated as one-way street that extends from Al-Resafi to Hafed Al-Qadi intersections. This street is characterized by large number of commercial shops, private and public offices, high pedestrian movements, and on street parking permission. Five intersections are selected on Al-Rasheed street. These are: Bab Al-Muadham intersection, Al-Meedan intersection, Al-Resafi intersection, Hafed Al-Khadi intersection and Al-Sinak intersection. The third street is Al-Kefah street. This street starts at Al-Fadhel intersection and ends at Al-Gailany intersection. It is a one-way street with six lanes. In addition to the aforementioned street's characteristics, it has many industrial blocks. Four intersections on this street were selected to be evaluated by this study: Al-Fadhel intersection, Zubaida intersection, Al-Thawra intersection and Al-Gailany intersection.

Simulation of the existing conditions

TRANSYT-7F Package was applied to simulate the existing traffic flow for all intersections in the selected network at A.M and P.M peak periods. It is worth while to mention that for A.M peak period, twenty nodes out of twenty seven have LOS F. While the remaining seven nodes ranging between LOS C to E. These nodes are located on the boundaries of the CBD, so they were less congested in comparison with the others (inside the CBD). For the P.M peak period, the case was more or less the same. Twenty nodes out of twenty seven have LOS F, and the rest was between LOS D to E. At this point, it is recommended to consider the P.M data for further investigation. System performance measures in terms of v/c ratio and level of service

were evaluated. Actually, TRANSYT-7F presents several measures, these may include average delay, v/c ratio, total stops, fuel consumption, disutility index ... etc. But, in the current paper the authors limit the analysis results comparison to only two of these measures.

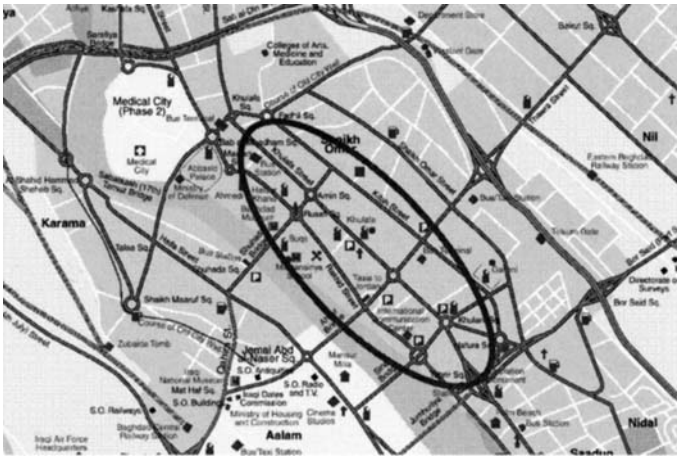


Figure (1) Location of the Selected street network for the Study.

Coordination alternatives of traffic flow

Coordination of traffic signals was applied in order to get use of a good TSM measure by which traffic congestion can be handled to an acceptable extent. All nodes within the network were coordinated. But, unfortunately, this solution to the existing traffic condition has offered no significant difference in the overall system performance. Therfore, another measures were have to be included and tested. For this purpose, two alternatives were designed to include geometric changes and traffic signals optimization for the base year 2003. Moreover, one alternative for the future years 2008 and 2013 was considered.

Base year: Alternative no. one

In this alternative the on street parking prohibition policy along some links is applied, in addition to the optimizing of signal timing by using CYCOPT program (provided in the main package of TRANSYT-7F). Moreover, restripping of the available road width to increase numbers of lanes was considered. Always, the criteria was to keep the minimum dimension of the lane not less than 2.4 m. Although the performance of some nodes was improved but, still the remaining has many difficulties in operating the traffic in more exceptable level. With these changes still nine nodes have LOS either E or F. Also, fifteen nodes have v/c ratio of 89 and more.

Base year: Alternative no. two

In this alternative additional measures was considered. Such as the one way system operation along Al-Rashid street (northbound). Currently, the one-way operation is applied in the section between Al-sink and Hafid Al-Qadi intersections, and between Al-Maidan and Bab Al Muadham squares. Optimizing the signal timing in Al-Sinak, Al-Rusafi, Al-Amin, Al-Thawra and Al-Fadhil intersections were also considered. In fact, the performance measures up to this point have showed that the network still operating at unexceptable LOS. So, it was necessary to investigate another possibility which will cost more than only implementing TSM measures. Therefore, traffic segregation in space was disscused and included at Al-Khulafa and Al-Khulani squares designs. These at-grade intersections were modified into overpasses in the east-west (Al-Khulafa) and north-south (Al-Khulani) directions. Generally, network performance was evaluated and the results of LOS ranged between A and C. Moreover, v/c ratios for all nodes were below 90%, except three nodes. Overall LOS for the whole network was C.

Future considerations

To take future growth into consideration, an expansion factor of 5% was applied. Also, two target years for short and medium term planning, 2008 and 2013, were regarded. Testing the network (alternative no. two) sufficency to accommodate the forecasted traffic movements has indicated that poor operation condition will be expected. As this was for year 2008, and surly the situation will be even worst for year 2013. Results showed that for year 2008 fourteen nodes will have v/c ratio of 90% and more. LOS ranged between A to F. Overall network LOS was F. While for year 2013 tewnty nodes will have v/c of 90% and more, and LOS ranged between A and F. Overall network LOS will be F.

Future year: Alternative no. three

To handle future traffic reasonbly, another TSM measure was added to alternative no. two. Then, rerouting of the private vehicles traffic into parallel alternative streets and limiting the use of the study network to buses only was introduced to form alternative no. 3. Due to the importance of the selected network location, it was suggested to apply the alternative during the peak periods. Generally, results have shown that the new planned alternative has offered a practical solution in absorbing the future traffic volumes within the area. As for year 2008, only one node will have v/c of more than 90% and overall network LOS will be C. While for year 2013, eight nodes will have v/c of more than 90% and overall network LOS will be E.

Conclusions

A thorough investigation to three streets network performance within Baghdad city was carried out. Simulation of the existing traffic condition has revealed that the network suffers a high level of congestion (LOS F). Therefore, different improvement strategies were evaluated using TRANSYT-7F package. Results have concluded that TSM measures in combination with costly geometric changes (constructing of two major overpasses) are required to cope with the increasing future traffic.

References

Abu-Lebdeh, G. and Benkhal, R.F. (2001), "Signal Coordination and Arterial Capacity in Oversaturated Conditions", Transportation Research Board, Washington D.C., TRB, January.

Al-Nuaimi, H.K., (2002), "Performance Evaluation and Improvement of Al-Khulafa street in Baghdad City", M.Sc. thesis, College of Engineering, Uni. of Al-Mustansiriya, July.

Banks, J.H., (2000), "An investigation of some characteristics of congested flow", Washington D.C., TRB, November.

Federal Highway Administration (FHWA), (1988)," Manual on Uniform Traffic Control Devices for Street and Highways", U.S. Department of Transportation, Washington, DC.

Hougeh, C.L. and Kuwahara, M., (2001), "An approach on Network Traffic Signal Control under the Real Time and Over-saturated Flow Condition", Transportation Research Board, Washington D.C., TRB, January.

Jrew, B.K., Al-Azzawi, A.A., and Shalal, Y.A., (2001), "Evaluation of Interrupted Traffic Flow along Arterial Street in Baghdad City", Engineering and Development Journal, June.

Khatib, Z.K. and Judd G.A. (2000), "Control Strategy for Oversaturated Signalized Intersection", Washington D.C., TRB, November.

Kuorau, Li. and Tarko, A. J. (2000), "A congested-oriented approach to detect incidents and to estimates capacities on signalized streets", Transportation Research Board, Washington D.C., TRB, January.

Messer, C., Urbanik II, T. and Park, B., (2001), "Traffic Signal Optimization Program for Oversaturated Conditions, A Genetic Algorithm Approach", Transportation Research Board, Washington D.C., January.

Transportation Research Board (2000), "Highway Capacity Manual", Special Report 209, Washington D.C.

Measuring the response of drivers to a yellow phase with a video based approach

Yue Liu[1], Gang-Len Chang[1], Ruihua Tao[2], Eric Tabacek[2], and Thomas Hicks[2]

[1]Department of Civil Engineering, University of Maryland at College Park, College Park, MD 20742; PH (301) 405-1953; FAX (301) 405-1235; email: troybest@umd.edu
[2]Office of Traffic and Safety, Maryland State Highway Administration, 7491 Connelley Drive, Hanover, MD 21076; PH (410) 787-5860; FAX (410) 582-9469; email: rtao@sha.state.md.us

Abstract: Understanding the response and acceleration/deceleration rate of driving populations to a yellow phase is essential for estimating the dynamic distribution of intersection dilemma zones. This paper presents a video-based method for measuring driver responses during a yellow phase, including their speed evolution profile, acceleration/deceleration rate, and the approximate reaction time. Such information is critical for understanding the spatial distribution of dynamic dilemma zones and the design of strategies to improve intersection safety. This paper details the key components of the proposed system and the systematic procedures for both field operations and data extraction. The results of a rigorous validation with an advanced experimental vehicle provided by Nissan are also reported in this paper.

Background

A reliable measurement of the speed evolution profile of driving populations from the start of a yellow phase to the red phase can offer invaluable information for understanding behavior patterns of drivers and the design of safety-improvement strategies. A well-captured speed profile can directly yield the following data for estimating the distribution of dynamic dilemma zones: 1. Measuring a driver's response time to the yellow phase; 2. Classifying drivers into different groups, based on their acceleration/deceleration patterns, and reaction times; 3. Developing a statistical model for estimating the spatial distribution of dynamic dilemma zones at signalized intersections; 4. Analyzing the car-following behavior of drivers during a yellow phase; 5. Investigating the interrelations between various driver responses, yellow phase design, and signal related crash rates.

A variety of sensors have been used in practice for direct measurement of vehicular speed, including microwave sensors, radio wave sensors, ultrasonic sensors, radar speed meters, and infrared speed meters. However, all these sensors are for direct measurement of a spot speed, which cannot reliably track an individual vehicle's movement over a target interval. Other direct speed measurement methods, such as distance-measuring instrument (DMI), global positioning systems (GPS), and cellular-phone location systems, have also been used to obtain the speed data along a roadway segment. But these methods can only capture the speed evolution of vehicles equipped with those measurement systems, which are likely to yield only biased and limited samples for analysis. One of the alternative methods is to measure a vehicle's speed indirectly from video images. A large body of methods for indirect speed measurements is available in the literature and has been used in practice. Such methods often involve quite costly and complex image processing work in spatial dimension, including identifying, extracting, and tracking vehicle for computing the speed. A relative cost-effective way to measure a vehicle's speed, as reported in the literature, is to employ video cameras or

camcorders. This approach often requires setting a speed trap at the video image with two reference lines separated by a known distance (Robertson, D. 2000; Dickinson et al. 1984; Ashworth et al. 1985). One can then compute the average speed between these two reference lines by dividing the distance with the travel time. This study intends to extend these video-based methods for spot speed measurements to a reliable and cost-effective system for measuring the spatial evolution of vehicle speeds. The key idea of the proposed approach is to superimpose reference lines over the video image and measure the vehicle's travel times between these lines sequentially to obtain the speed evolution profile. The distance between two adjacent reference lines is optimized to minimize the potential measurement errors under given operational conditions. The time when vehicles reach the reference line and the starting time of a yellow phase are also recorded by the program for extraction of speed evolution profile before-and-after a yellow phase.

System Components

The entire system for speed measurement includes the following components: 1. One DVD video camera, which can record at variable time-elapse rates up to 30 frames per second, along with several re-writable DVD video disks; 2. One adjustable tripod to allow a flexible camera orientation setup; 3. Orange cones placed at an identical distance along the roadway as reference points for video benchmarking and reference line generation; 4. A frame-by frame video editing computer program, which must be able to read the video file directly from the videodisk without any converting or capturing job, superimpose the reference lines onto the video image, slice the video footage into a small set of segments (up to a frame) to facilitate accurate analysis, and record the necessary timestamps when vehicles touch the line.

Implementation Issues

With the system developed for this study, several critical issues need to be addressed, including camera set-up, measurement accuracy, selection of speed trap length, video benchmarking, reference line generation, and data extraction.

Camera Set-up

The far-side camera should be set up based on the following criteria (see Figure 1): 1. The entire survey segment can be captured as long as necessary; 2. The signal phase changes can be captured; 3. The front wheel of vehicles can be identified as the detection point; 4. All the orange cones can be observed clearly in the video image.

Measurement Accuracy Analysis

According to the measuring approach, the average speed over each trap length is approximated as the spot speed at the reference line, there inevitably exists some difference from the actual spot speed. Therefore, if the trap length is sufficiently small and the vehicle keeps constant speed within the trap, the average speed will be equal to its spot speed, and there will be no error associated with the above conversion. So the length of the speed trap should be set as short as possible to reduce the approximation errors. On the other hand, the length of the trap should be maximized to reduce the time-elapse errors caused by a video camera. Hence, there exists a trade-off between conversion errors and time-elapse errors in setting the speed trap length. Note that vehicles traveling within the trap may execute different acceleration or deceleration rates. In this study, we use the worst scenario to assess the maximal possible measurement errors. For the speed conversion errors, the worst scenario occurs when a vehicle keeps accelerating or decelerating within a trap using the maximum acceleration rate ($16.0 ft/\sec^2$) or deceleration rate ($-11.2 ft/\sec^2$) (Gazis, D. 1960). For the time-elapse error, the worst scenario occurs if one frame of time is missing or over-counting

from the calculation of travel time between two reference lines. The maximal possible error estimation models are approximated with the following equations:

$$\varepsilon_{\max}^{c} = \frac{\left| v_{act} - \sqrt{v_{act}^2 - \frac{2aD}{(1.47)^2}} \right|}{2} \quad (1)$$

$$\varepsilon_{\max}^{t} = \frac{v_{act}^2}{K(v_{act} - \sqrt{v_{act}^2 - \frac{2aD}{(1.47)^2}}) + v_{act}} \quad (2)$$

Where: $\varepsilon_{\max}^{c}$ is the maximal speed conversion error (mph); $\varepsilon_{\max}^{t}$ is the maximal time-elapse error (mph); v_{act} is the actual speed of a vehicle at the reference line (mph); K is the number of frames per second; D is the length of the speed trap (ft); a is the maximal acceleration/deceleration rate (absolute values are used) within the speed trap ($ft/\sec^2$).

The above estimation models are deducted from basic vehicle moving dynamics and the acceleration/deceleration rates used are absolute values. The estimation model is somewhat conservative as vehicles don't often use the maximal acceleration/deceleration rate within a speed trap, and the missing or over-counting time from the calculation of travel time is always a fraction of one frame. The error estimation equations show that $\varepsilon_{\max}^{c}$ increases with length of a speed trap, and $\varepsilon_{\max}^{t}$ decreases with an increase in the speed trap length.

Selection of Speed Trap Length

Although it is difficult to compute a theoretically optimized value, this study has taken into account both types of errors and their trade-off in setting the speed trap length. It can be easily seen that an effective speed trap length shall lie at the point where $\varepsilon_{\max}^{c} = \varepsilon_{\max}^{t}$, so as to minimize and balance both types of errors. The speed trap lengths and the measurement errors under different speed levels can then be computed by the above equality. For each survey location, the average speed of the survey segment is used to decide which speed trap length should be used, and the selected speed trap length will then be applied in video benchmarking and speed data extraction, as summarized in Table 1.

Video Benchmarking and Reference Line Generation

The purpose of this task is to extract the spatial information of the target survey segment, and generate virtual reference lines at the video image where cones could not be used as reference points. The study has developed the following procedures (see Figure 2 and 3): 1. Take sample digital photos from different camera orientations during the field survey. 2. Mark grid corners at each video image to construct the coordinate systems. 3. Mark corresponding locators (the same object at different images, here the cone's vertex is used as the corresponding locator) for calibration. 4. Calibrate camera parameters and model the coordinate system. 5. Use the calibrated information for reference point extraction on the image. 6. Superimpose the extracted points over the video image to generate reference lines for speed measurements.

Data Extraction

Given the above procedures finished and survey videos available, a computer program was used to extract speed evolution data, as shown in Figure 4. During the extraction process, for each cycle, record the yellow phase starting and ending time separately, and identify all the

through vehicles trapped in the yellow phase. For each vehicle, record the time when it travels over each speed trap. Calculate each vehicle's speed evolution from the elapsed time and the distance traveled.

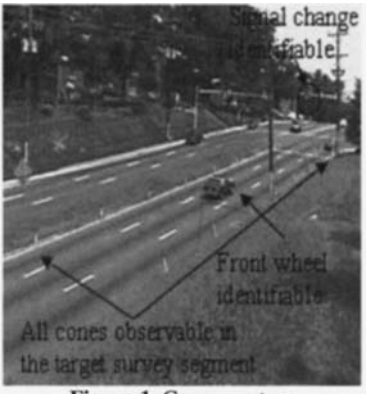


Figure 1. Camera set-up

Table 1. Selected speed trap lengths and maximal measurement errors

Speed[1] (mph)	Selected Speed Trap Length (ft)	Maximal Speed Conversion Error (mph)	Maximal Time-elapse Error (mph)
10	10[2]	2.89	0.76
15	10[2]	2.17	1.33
20	12	1.95	1.95
25	15	2.19	2.19
30	20	2.36	2.36
35	25	2.56	2.56
40	30	2.77	2.77
45	35	2.85	2.85
50	43	2.90	2.90
55	49	3.13	3.13
60	55	3.37	3.37

[1]The speed in the table is the average speed at the target survey segment.
[2]The length of 10 ft was set as minimum speed trap length for operational convenience.

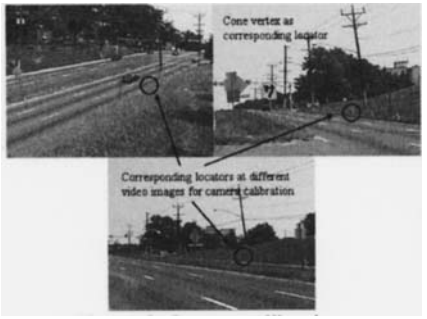


Figure 2. Camera calibration

Figure 3. Extracted reference points

Field Validation

To evaluate the accuracy of the proposed system for speed measurements, this study has conducted a field test at the intersection of MD 650 and Metzerott Rd (at northbound approach with an average speed about 40 mph). A Nissan Infinity Q45 instrumented with a CAN (Controller Area Network) message converter was employed in the test to provide the true speeds for comparison (the differences between video measured speeds and those from CAN were considered as errors). The CAN message converter is a measuring device which can convert the actual speed messages of the vehicle to decimal values. It was calibrated to the precision of ±0.0001 mph and connected to a laptop computer via a serial cable to display the speed of the experimental vehicle in a time frame of every 0.01 second. Two video cameras were used for validation. One was set at the far side to record the movements of the experimental vehicle in the surveyed segment, and the other was installed in the vehicle to record its actual speed displayed on the screen of the laptop. The synchronization of these two video cameras has yielded the consistency between the accurate speed by CAN and the

measured speed by video using the timestamp information (see Figure 5). Based on the above speed trap length selection design, the speed trap was set at 30 ft to minimize possibly maximal approximation and/or time-elapse errors. The field validation consists of 24 trials through the test site with entry speeds at six different levels (20-25, 25-30, 30-35, 35-40, 40-45, 45-50 mph), and each speed level has 4 trials (2 for pass, 2 for stop). There are a total of 180 speed records (each "pass" trial has 8 records and each "stop" trial has 7 records through the evolution process to the stop-line) for validation. The errors of speed measurements were calculated for each experiment and displayed in Table 2.

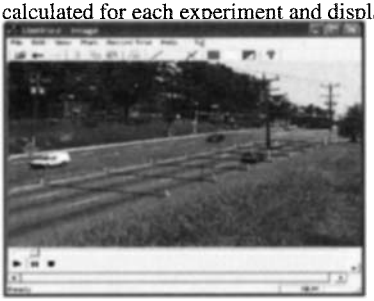


Figure 4. Data extraction

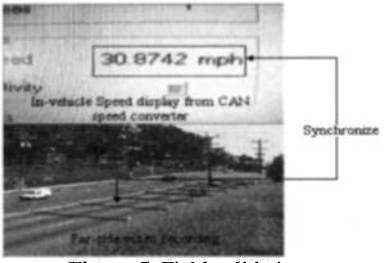


Figure 5. Field validation

Table 2. Errors of the Video-based Method under Different Entry Speeds

Entry speed level[1] (mph)		Speed Range[2] (mph)	Mean Error[3] (mph)	(Min, Max) Error[3] (mph)	Maximal Theoretical Error[4] (mph)
20-25	P	(20-32)	0.99	(0.01,3.59)	4.55 (>3.59)
	S	(0-26)	1.33	(0.04,3.68)	6.69 (>3.68)
25-30	P	(17-30)	1.22	(0.04,2.58)	5.45 (>2.58)
	S	(0-30)	1.57	(0.04,3.99)	6.69 (>3.99)
30-35	P	(32-39)	1.61	(0.17,3.56)	3.65 (>3.56)
	S	(0-34)	1.79	(0.00,3.77)	6.69 (>3.77)
35-40	P	(33-47)	0.75	(0.01,1.95)	4.08 (>1.95)
	S	(0-39)	1.87	(0.27,4.14)	6.69 (>4.14)
40-45	P	(41-50)	0.71	(0.01,3.18)	4.08 (>3.18)
	S	(0-43)	1.62	(0.11,3.86)	6.69 (>3.86)
45-50	P	(42-50)	1.26	(0.09,3.05)	4.08 (>3.05)
	S	(0-48)	1.61	(0.32,3.54)	6.69 (>3.54)
Sum	P	/	1.09	(0.01,3.59)	5.45 (>3.59)
	S	/	1.53	(0.00,4.14)	6.69 (>4.14)

[1]The entry speed is a spot speed when the test vehicle enters the survey segment.
[2]Speed range means speed evolution range.
[3]All the errors in the table are absolute errors.
[4]The maximal theoretical errors were the maximal of values computed by Eq. (1) and (2) given the speed trap length of 30 ft.

The maximum and minimum absolute values of the errors for the experiments and the maximum theoretical errors given by the Equations (1) and (2) were also listed. It is obvious that the errors of the speed measurements were less than the maximum theoretical errors, which suggests that the methodology developed in the study is sufficiently reliable for estimating the speed evolution. In Table 2, it is noticeable that across all the six levels of entry speeds, the experiments with "stop" maneuvers produced relatively larger measurement errors than with "pass" operations, which suggests that the accuracy level of speed measurements be sensitive to the acceleration/deceleration rate. The reason is that the length of the speed trap was set at 30ft on the basis of the speed level of 40-45 mph to minimize the potential measurement errors. However, when the vehicle's speed diverted away from that speed level, the measurement errors may increase and the preset speed trap length may not be the most effective selection. The way to improve accuracy of speed measurements for "stop" maneuvers is to use a best-fit-in length of speed traps, based on speed changes. However, it remains to be a challenge in practice.

Measuring the Response of Drivers

A driver's perception-reaction time in response to YELLOW is a critical factor that affects the dilemma zone distribution at signalized intersections (Xiang et al. 2005). Field measurements of a driver's perception-reaction time can offer invaluable information for understanding the interrelationship between driver behavior and surrounding factors. The proposed method offers a convenient way to approximate a driver's response time with his/her speed profile (approximately equal to a theoretical perception-reaction time). Figure 6 shows a speed evolution of a stop-maneuvered case in the field validation. A yellow phase started at the timestamp of 1164.01584 seconds. After that a significant speed reduction (10.43 mph) occurred between the timestamps of 1164.70886 and 1165.21386, as shown in Figure 6. Despite the average speed measurement error of ±1.53 mph for "stop" cases (see Table 2), the speed change in this case was still significant in such a short time period. Therefore, this speed reduction was identified as the driver's response to YELLOW, and the driver's response time was then estimated to lie between 0.69 and 1.20 seconds. One may use the average to represent the approximate response time of a driving population.

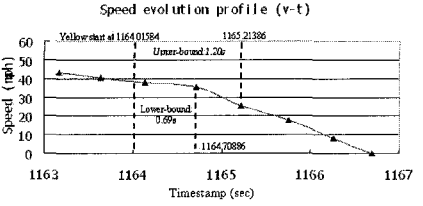


Figure 6. Measuring driver's response

Conclusion

It is found that the accuracy level of speed measurements by a video-based method is a function of several factors, such as length of the speed trap, the acceleration/deceleration rates and speed within the speed trap, time-elapse rate used and camera setup. Test results show that, if properly designed, the proposed video-based method is effective to measure the speed evolution, the acceleration/deceleration rate changes, and the response time of different driving populations to YELLOW, which provide all essential information for understanding the spatial distribution of dilemma zones.

References

Ashworth, R., D.G. Darkin, (1985) "Applications of Video Image Processing for Traffic Control Systems." Second International Conference on Road Traffic Control, London, UK, pp. 119-122.

Dickinson, K.W. and R.C. Waterfall. (1984). "Video Image Processing for Monitoring Road Traffic." IEEE International Conference on Road Traffic Data Collection, pp. 105-109.

Gazis, D., Herman, R., and Maradudin, A. (1960). "The problem of the amber signal light in traffic flow." *Operations Research*, Vol. 8, No. 1, pp. 112-132.

Robertson, D. (2000). "Spot Speed Studies." *Manual of Transportation Engineering Studies*, ITE, Washington, D.C., pp.33-51

Xiang, H., C.S. Chou, G.L. Chang, and R. Tao. (2005) "Observations and classification of Drivers Responses during Yellow-Light Signal Phase." Presented at 84th Annual Meeting of Transportation Research Board, Washington DC.

Tests of Traffic Sensors and Telemetry Services

Panos D. Prevedouros,[1] PhD and Goro Sulijoadikusumo,[2] MSCE, PE, PTOE, GISP

[1] CEE Department, University of Hawaii at Manoa, Holmes Hall 383, 2540 Dole Street, Honolulu, HI 96822, USA – Tel: (808) 956-9698 – Fax: (808) 956-5014 – E-mail: pdp@hawaii.edu.
[2] State of Hawaii, DOT, Highways Planning Branch, 869 Punchbowl Street, Ste. 301, Honolulu, HI 96813– Tel: (808) 587-1839– Fax: (808) 587-1787 – E-mail: Goro.Sulijoadikusumo@hawaii.gov.

ABSTRACT

This paper presents the results from field deployments and tests of eight portable and semi-portable detector systems. These tests are the culmination of four years of research conducted to determine suitable, primarily non-intrusive, traffic detector systems for use by the Hawaii DOT for its traffic monitoring program.

Three types of sensor installations were made, underground or sub-surface, on-ground and above ground, side-fired. All three above ground, side-fired, unintrusive sensors (SAS-1, SmartSensor and RTMS) provided reliable and reasonably accurate volume and speed data when properly installed and calibrated. Among the four on-ground sensors capable of providing vehicle classification data (in 13 classes), the Roadtrax BL piezoelectric sensor was the only one that was capable of providing adequate and reliable data.

The systems tests included evaluations of telemetry and other services for data retrieval and remote diagnostics. TrafInfo's Trafmate satellite modem and digital pager, and TrafficWerks' near-real time data retrieval via cellular sub-carrier and integrated archival system performed satisfactorily and were far superior to classical Hayes modems.

The research was extended to evaluate the potential for modifying signalized intersections for traffic data collection given the expanded functionality of the sensors, systems and services purchased. The remote collection of volume data without interfering operations at a busy 5-phase signalized intersection was possible from an intersection's 332 traffic cabinet with a loop actuated Type 170 controller using Canoga loop boards, and adding a TransHub and TrafInfo digital pager. Other tests showed the potential of remote collection of near-real time freeway data from legacy equipment by adding Canoga detector cards, CDMA modems and service integration with Internet data retrieval by TrafficWerks.

OBJECTIVES AND SENSORS TESTED

The purpose of this report was to investigate the suitability of a variety of mostly unintrusive detectors for collecting traffic data for Hawaii DOT's Traffic Monitoring System program. This study was conducted on the abundant foundation of similar sensor deployments and tests such as those reported by Klein (2001), Middleton and Parker (2004), and Hallenbeck and Weinblatt (2004).

The tests included assessments of **usability** which includes the elements that comprise a complete detector apparatus, detector, connection to counting box or other storage device, power requirement and power sources (grid, solar or battery) as well as telemetry options. Issues with installation, calibration and portability were examined. **Accuracy** comprises the traditional part of detector evaluation. It assesses their accuracy of axle or vehicle counts, speed and classification, as applicable to each device. Inductive loop data were used as the basis for most comparisons. Sample visual counts in both heavy and light traffic were

also conducted. A variety of outputs, accuracy, flexibility, ease-of-use and cost and labor considerations were combined to derive recommendations on the use of detectors.

This paper is a summary of the results and lessons learned from field deployments and tests of eight portable and semi-portable detector systems. The systems tested, grouped by how the sensors were installed, are listed below:

Underground or underbridge: (1) 3M M701 microloops and Canoga 800 Series detector cards, (2) Canoga detector card connected to pre-existing freeway loops, and (3) Canoga detector card connected to pre existing intersection loops.

On-ground: (1) Optical Sensor Systems fiber optic sensors and ITC TRS counter/classifier (2) Pneumatic tube sensors with JAMAR TRAX RD counter/classifier, (3) MSI Roadtrax BL piezoelectric sensors with PEEK ADR-2000 counter/classifier, and (4) Spectra Research ORADS (NTMS) portable laser sensor with IRD TCC 550 counter/classifier.

Above ground, side-fired (Figure 1): (1) RTMS microwave sensor and RTC data unit by EIS (2) SAS-1 acoustic sensor and SAS-CT board by Smartek, and (3) SmartSensor microwave sensor by Wavetronix.

Some on-ground sensor systems (as noted above) were incorporated in the tests because most above ground, non-intrusive sensors were determined to be not adequate for providing pavement design data (e.g., they cannot provide FHWA scheme F vehicle classification data). The idea was to use some portable on-road sensors installed for some brief time periods (typically 48-hours) and then use the axle correction factors along with the data from the non-intrusive sensors to calculate estimates of truck percentages in the traffic stream. However, problems with the on-ground sensors themselves or with the comparison equipment limited the testing and the adequacy of the results.

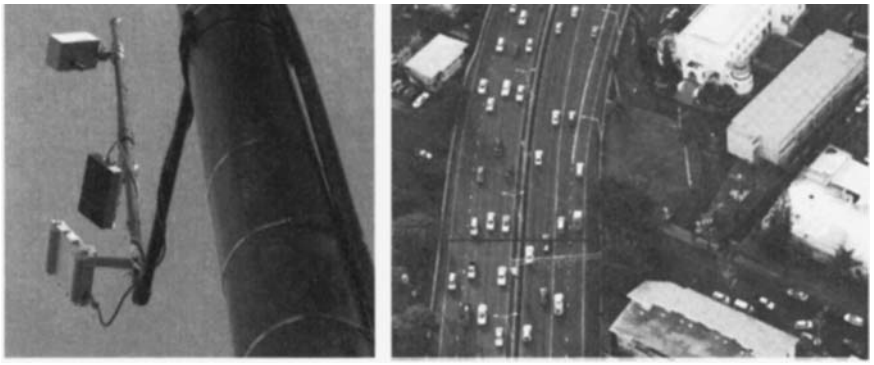


FIGURE 1. Deployment[1] of SmartSensor, SAS-1 and RTMS at an 8-lane freeway cross section.

TEST RESULTS

Locations were selected along principal and minor arterials which are the majority of roadways under the jurisdiction of HDOT. Initially, sensors were tested in the vicinity of a fully instrumented truck weighing station and progressively moved to more difficult sites to assess the suitability of sensors under various traffic congestion conditions, variable heights and offsets as well as the effects of barriers and cross-sectional width. Tests were conducted

[1] Microwave radar-based SmartSensor and RTMS were tested not only separately to avoid interference, but also together. No difference in measurements due to interference was observed.

on a 4-lane road with a significant amount of heavy vehicle traffic, 6, 8 and 10-lane freeway cross-sections, and a 4-lane collector road.

The ranges shown below (in terms of percent error) were used for the qualitative assessment of sensor accuracy.

Rating	Volume/Classification	Speed
Excellent	±1	±3
Very Good	±3	±6
Good	±5	±10
Possibly adequate	±10≤	±15≤
Inadequate	>±10	>±15

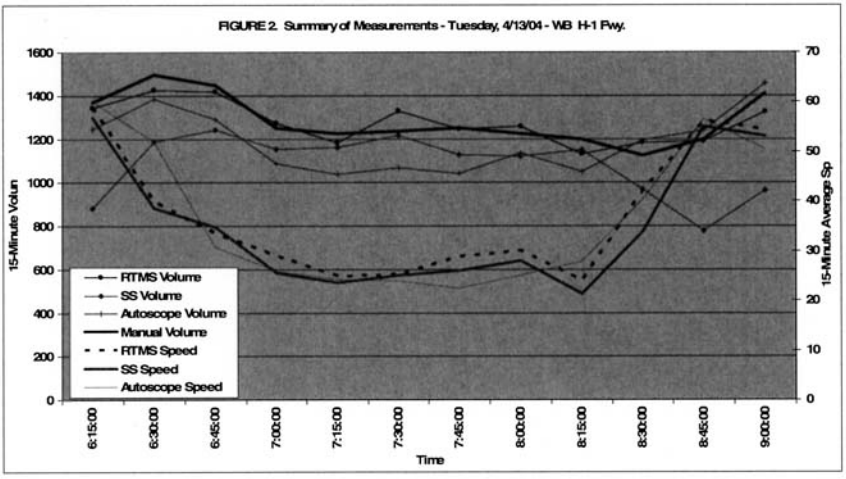
Acoustic (SAS-1) and microwave (RTMS and SmartSensor) sensors have a combination of positive attributes: reasonably accurate, fairly easy to deploy, relatively inexpensive to acquire. The SmartSensor's auto-ranging feature provides for a quicker setup and does not require the calibration of average vehicle length for accurate speed measurement in side-fired deployment – it is the most portable among these three sensors. All three sensors have real-time counting modes displayed on a notebook's screen which allows for the verification of count accuracy.

Weaknesses or limiting requirements of the three unintrusive sensors include the following (some are more relevant to portable deployments only):

- An offset of 20 ft. between the sensor's pole and the first lane of traffic being detected is required. This makes the deployment of these sensors impractical because poles are not commonly available at a 20 ft. or longer offset. The SAS-1 can detect the nearest lane adequately at a 15 ft. offset.
- A weather-proof cabinet is required for batteries and their termination (counter) units and the modem.
- A portable computer, some expertise in the setup software, patience as well as good timing. The latter means that setup should be conducted in the presence of light-to-moderate traffic, otherwise the "blips" (vehicle signatures) on the computer screen can be difficult to compare with actual traffic for the proper setting of lanes. This is harder for the RTMS because of its brief display time lag. The "blips" are essential for setting up the lanes, but they should not be used for testing count accuracy due to potential overlaps or double-counts. All sensor PC-based interface software have running-time data tabulations for this purpose.
- If highly accurate volume counts and speed measurements are required, sample data with a reliable on-pavement sensor such as the piezoelectric or the fiber optic sensors tested in this project should be collected for comparisons and fine-tuning of the sidefired sensor parameters. This can be avoided if a crew of two people is available to spend the better part of a workday for setting up a sensor to measure a 4-lane highway in the field by collecting multiple 5-minute data per lane (volume counts and speeds with a portable laser gun.)
- Deployment of these sensors in roadway cross-sections that include fixed barriers and similar objects should be avoided because of issues with detections on the lane immediately behind the barrier and other detection artifacts.
- Deployment of the SAS-1 in areas where loud music or loud stationary machinery are likely to be present should be avoided. The use of SAS-1 should be limited to noisy traffic environs.
- The RTMS has difficulty in detecting bicycles and mopeds whereas the SAS-1 has difficulty excluding these if they are not desired to be a part of the counts. This is an issue with SAS-1 in quiet environs where the sound of bicycle tires or chain can be detected. SmartSensor was not tested for this.

In general, the three side-fired sensors, SAS-1, SmartSensor and RTMS, can provide high quality data at a low cost, with minimal energy consumption and with modest setup and calibration time. All three can be used with a portable apparatus that includes an extensible

mast with a minimum height of 20 ft. Samples of volume and speed measurements taken at the site in Figure 1 are shown in Figure 2.



Various types of telemetry were investigated to examine system usability and practicality issues. Several telecommunications and data retrieval services were used: (1) Regular modem – Hayes @ 9600 baud), (2) Motorola 3 watt analog cellular phone with a ZYXEL low power cellular modem, (3) TrafInfo's Trafmate satellite modem, (4) TrafInfo's Trafmate digital pager, and (5) TrafficWerks' near-real time data retrieval via cellular sub-carrier (CDPD or CDMA modem) and integrated archival system.

On-site visits for reconfiguring sensors and data retrieval are sometimes expensive and can be hampered by weather and other adverse conditions. Telemetry provides a link to field devices so that data can be retrieved, and it offers limited remote diagnostic and control functions for the devices. Major lessons learned include the following:

(1) Long Interval (12 to 24 hr) data retrieval from field stations via satellite modems or digital pagers (TrafInfo service contract) is convenient, economical and reliable in most cases. Digital pagers are recommended for Hawaii due to the lack of nearby earth relay stations for satellites. TrafInfo offers some basic two-way communication with field devices including battery checks, clock checks and changes to data collection intervals.

(2) Near real-time data retrieval from field stations via CDMA modem (TrafficWerks service) is relatively expensive but refined, information-rich and easy to use. Its advantages include Internet Protocol (IP) addressability and compatibility with most sensors. TrafficWerks utilizes its IP addressability to integrate new or modern data stations with legacy data stations. Data from multiple stations can be accessed and various summaries can be requested through a password protected web interface. The field devices cannot be reconfigured remotely by the end-user; however the service provider can perform these basic functions.

A temporary underbridge installation of the 3M microloop sensors revealed a number of limitations in detection but the associated Canoga detector cards had extended features and functionality that warranted further testing. Specifically, the ability of communicating with each Canoga detector card independently to retrieve traffic data and without disrupting the operations of a signalized intersection with a loop actuated Model 170 controller was tested. These Type 170 controllers with standard loop detector cards, which can be replaced with the

Canoga cards, are common in signalized intersection or freeway management system cabinets and were investigated because this infrastructure is often underutilized for traffic monitoring programs.

The remote data collection from a freeway management system traffic cabinet was accomplished with several Canoga loop detector cards and one TrafficWerks digital CDPD modem (mobile telephony) at the Halawa Stream cross section of H-1 Freeway on Oahu and a similar setup but with a CDMA modem at section near the west end of Moanalua Freeway (H-201). Remote data collection from a traffic cabinet of an actuated controller at a signalized intersection was accomplished with several Canoga loop detector cards and one TrafInfo digital pager connected to a TransHub at an intersection on Route 11, in Hilo, Hawaii.

CONCLUSIONS

The conclusions may be summarized as follows. Both SmartSensor and RTMS are appropriate for quiet rural locations because sound artifacts within the detection area do not affect their measurements. The former was ranked higher based on its ease-of-setup, lower height requirement and exceptional feedback and assistance by the vendor. The SAS-1 is appropriate for applications where limited utility infrastructure is available as it can operate with marine (deep-cycle) batteries and solar panels due to its minimal energy consumption. Depending on solar panel size, a month's data can be collected with two 75 Ahr batteries. It is recommended for locations with constant traffic noise and no barriers or walls that generate echoes. For all side-fired deployments, separate detectors should be used with a median taller than about 3 ft. (or 1 meter), and for cross-sections along busy roads with six or more traffic lanes.

TrafInfo's data retrieval service is appropriate for rural, low volume and secondary counter stations due to its low cost and good reliability. It is also appropriate for data retrieval from signalized intersection loop counters. TrafficWerks' service is expensive and multifaceted. As such, it is appropriate for stations in high volume and high congestion highway cross-sections because of its double utility as a data collection-and-archiving tool, and as a congestion and operations monitoring service.

REFERENCES

Klein, L. A. (2001) *Sensor Technologies and Data Requirements for ITS Applications*, Artech House.

Middleton, D. and R. Parker (2000) *Initial Evaluation of Selected Detectors to Replace Inductive Loops on Freeways*, FHWA/TX-00/1439-7, Texas Transportation Institute, College Station, TX, April 2000.

Wald, M., *Microwave Vehicle Detection*, Final Report, CALTRANS, January 2004.

Hallenbeck, M. and H. Weinblatt (2004) *Equipment for Collecting Traffic Load Data*, NCHRP Report 509, TRB, Washington, D.C.

ACKNOWLEDGMENT: This paper is a product of research on INVESTIGATION OF TRAFFIC DETECTORS FOR USE IN HAWAII supported by the Hawaii DOT and the FHWA. The contents of this paper reflect the views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Hawaii DOT or the FHWA. This paper does not constitute a standard, specification or regulation.

Summary of Sensor Test Deployments*

Deployment	Location	Time Frame	Objectives	Results
1	EB H-1 Fwy. at Paiwa Interchange	1/2002	Demonstration of 3M Microloops and Canoga under a freeway viaduct.	Interference from rebar and suboptimal Microloop placement due to girders were noted difficulties.
2	Sand Island Access Rd., outbound	3/2002	Deployments of RTMS, Flexsense fiberoptic, BL piezoelectric, ORADS. Comparisons with loop data from truck weight station (PAT America).	Flexsense fiberoptic with ITC TRS recorder gave excellent results. RoadTrax BL piezoelectric sensor with PEED ADR 2000 recorder gave excellent results. RTMS X2 with RTC data unit gave very good results. Spectra Research ORADS with IRD TCC 500 recorder provided inadequate results.
3	Sand Island Access Rd., inbound	5/2002	Added SAS-1 and tested data retrieval with Hayes 9600 modem and cellular modem.	Both SAS-1 and RTMS provided very good volume and speed with most differences from loops being under +/-10%.
4	Sand Island Access Rd., outbound	6/2002	Tested volume, speed and classification measurements of BL piezoelectric and pneumatic tubes with JAMAR TRAX RD	Pneumatic tubes were rather inadequate largely due to the very high proportion of loaded trucks. They were good for speed and poor for classification. BL piezoelectric was deemed very good for classification (FHWA 13).
5	WB H-1 Fwy. at McCully, 30 ft. height	9/2002	Tests of SAS-1 and RTMS at various boom heights with and without remote data retrieval with TrafMate modems provided by TrafInfo. Field visit by TrafInfo engineer to troubleshoot TrafMates.	The 20 ft. height enabled the sidefired sensors to perform best at counting the near side traffic on the 3-lane freeway. Offset was at a fixed 20 ft. SAS-1 was oriented 30-degrees from vertical and for the last set was changed to 45 degrees (per Smartek advice.) Various satellite location and frequency issues were discovered from TrafMates. All volume errors were under +/-10% with overall being -2.1% for RTMS and -3.0% for SAS-1. Both appear to overcount in heavy traffic and undercount in moderate and light traffic.
6	WB H-1 Fwy. at McCully, 25 ft. height	9/2002		
7	WB H-1 Fwy. at McCully, 20 ft. height	10/2002		
8	WB H-1 Fwy. at McCully, TrafMate	10/2002		
9	Dole Street by East-West Road	12/2002	This location along a 4-lane collector enabled the identification of issues which are not present along busy, noisy arterials and freeways such as the effects of bicycles and mopeds, the effects of wind noise from multiple nearby utility cables and the effects of music and other human activities in a relatively quiet environment. First test of TrafMate digital pager.	SAS-1 needed considerable calibration to minimize effects of loud music (the beat of music registered as vehicular counts), the wind howl from nearby utility cables and grass trimming machinery. SAS-1 counted bicycles in the bike lane, RTMS did not. Overall count accuracy was +4.1% for RTMS and +3.9% for SAS-1. BL piezoelectric sensors were deployed for base counts. TrafMate digital pager worked better that satellite modem.
10	EB H-1 Fwy. prior to 6th Ave. off-ramp	1/2003	Tested volume, speed measurements of SAS-1 and RTMS. Loop station was over 500 ft. upstream and some motorists switched lanes. Light plant located on shoulder has a 12 ft. offset.	Due to short offset, RTMS could not count nearest lane adequately. The pattern of counts by SAS-1 and RTMS was nearly identical to the pattern from the loops. SAS-1 volume for the 3 freeway lanes was closest to the loop count. The TrafMate digital pager worked better than the satellite modem.
11	H-1 Fwy., Makiki St. cross-section	2/2003 5/2004	This site included SAS-1, SmartSensor (SS) and RTMS sensors, a solar panel and four deep-cycle batteries. Besides the tests for SmartSensor, this site was setup to provide continuous counts used in various on-going projects. Due to the proximity of a traffic surveillance camera, simultaneous video was recorded on VHS tapes and analyzed with Autoscope RackVision.	RTMS at a height of 30 ft. monitors the far side (WB H-1 freeway, 4 lanes), SAS-1 at a height of 25 ft. monitors the near side (EB H-1 freeway, 4 lanes), and SmartSensor at 20 ft. monitors all 8 lanes. Offset is 20 ft. RTMS and SS data are comparable for the three freeway lanes, but low for the furthest, auxiliary lane. Autoscope (freeway lanes only) is in agreement with them for speed, but undercounts by 8% due to occlusion (camera is on the far side). Despite manufacturers' advice, the simultaneous operation of both microwave sensors did not cause odd patterns or any changes in the counts.
12	Signalized intersection of arterial streets (Kanoelehua St. and Puaianako St., Hilo, Hawaii)	11/2003 - 9/2004	The objective was to address the feasibility of collecting traffic data from existing signalized intersection cabinets (about 900 in Hawaii), none of which are used to collect data. This test was intended to demonstrate that data collection does not interfere with regular controller functions. Canoga data cards, a TransHub, and a TrafInfo digital pager were used. Canoga setup by 3M engineer.	The deployment was entirely succesful in remotely receiving traffic counts from intersection loops used for a traffic-actuated signal, without any effects on signal controller operation. This deployment revealed the difficulties of not having a purpose-built data collection infrastructure, suboptimal performance of field devices (e.g., vehicle counts from long loops) and the impact on the data collected. This was the only location for which correction factors for collected volumes were derived.
13	H-1 Fwy., Halawa Stream cross-section (Interchange H1-13)	5/2003 - 12/2004	Update 332 cabinet with Type 170 Controller so that data can be received in near-real time over the Internet. Demonstrate incorporation of Autoscope RackVision with the available traffic surveillance camera	3M Canoga data cards were deployed with CDPD modem, later updated to CDMA. HDOT designed the interchange layout and TrafficWerks provided data retrieval and display over the Internet. RackVision installation and Canoga tuning by Econolite and 3M representatives.
14	Moanalua Fwy. (H201), Halawa Stream cross-section	11/2003 - 12/2004	Update 332 cabinet with Type 170 Controller so that data can be received in near-real time over the Internet.	As the above. Demonstrate the ability to retrieve data over the Internet from multiple stations and in various forms.

Note (*): Deployments 1, 12, 13 and 14 were technology demonstrators with minor testing of sensors. Deployments 2 to 11 were mostly focused on testing traffic sensors.

Capacity of Multi-Lane Roundabout: Methodology Based on Lane Utilization

Jing BIE[1], Hong K. LO[1] and S. C. WONG[2]

[1]Department of Civil Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong; email: cehklo@ust.hk
[2]Department of Civil Engineering, The University of Hong Kong, Pokfulam, Hong Kong; email: hhecwsc@hkucc.hku.hk

Abstract

Traffic roundabout is a priority junction, whose capacity is typically captured by the entry capacities of individual approaches while considering the effects of conflicting flows. For a single-lane roundabout, entry capacity is traditionally analyzed based on gap acceptance where entering vehicles seek an acceptable gap among the circulating flow. An example of this approach can be found in HCM (2000). This paper provides an extension to this approach for multi-lane roundabouts. Through an analysis of lane utilization, entry capacity is estimated for each entry lane. Reserve capacity is then used as a measure to assess the overall roundabout performance. This paper also studies the sensitivity of drivers' lane choices on the overall capacity of traffic roundabouts.

Introduction

Traffic roundabout is an unsignalized junction where priority rules take the role of traffic signals by assigning rights-of-way among the conflicting traffic streams. Vehicles entering the roundabout must give way to vehicles already in the roundabout, i.e. circulating flows. Roundabouts have been used successfully around the world and are adopted increasingly in the United States. By 2004, the number of modern roundabouts in the U.S. has leaped to around 800 (Baranowski, 2004). Predicting roundabout capacity is an important element in the planning and design of such facilities. Due to the interaction between flows of different origin-destination paths at roundabout, capacity is conveniently evaluated for each entry. Entry capacity is defined as the maximum inflow of an entry path, under a given amount of circulating flows which conflict the entry flow. Analytical models based on gap acceptance theory, such as Highway Capacity Manual (HCM, 2000), give entry capacity formulae only applicable to single-lane roundabouts. This paper proposes a method where the HCM formulae can be extended for application at multi-lane roundabouts. For simplicity, only double-lane roundabouts are illustrated while the same

methodology can be readily extended to roundabouts with three or more circulatory lanes. *Notes: The annotations and figures in this study are based on vehicles driving on the left side of a road (where roundabout circulation is clockwise).*

Flow Formulation

Circulating movements at a double-lane roundabout are distributed between an inner circulatory lane (i.e., the lane closer to the central traffic island) and an outer circulatory lane. Correspondingly, at entry, an approach usually offers two entry lanes (though other forms exist): a left entry lane and a right entry lane. We assume that, traffic entering outer lane always waits at left lane for acceptable gap and stays on the outer lane until exit; traffic entering inner lane always waits at right lane and stays on the inner lane until changes to the outer lane when exiting at the upcoming arm.

Demand matrix and movement modeling. For a roundabout with n arms (as arm $1,2,...,n$ in a clockwise order), there exist n^2 pairs of origin-destination arms. Denote $\mathbf{D}_{n\times n}$ as the demand matrix with $D_{i,j}$ representing the demand flow entering the roundabout from arm i and exiting to arm j. For demand $D_{i,j}$, a proportion of $\alpha_{ij} \in [0,1]$ enters inner lane from right entry lane while a proportion of $(1-\alpha_{ij}) \in [0,1]$ enters outer from left. From any approach i, there are n possible different movements. We designate the left-turn movement, which exits at the next arm, as the first movement. The k-th movement is exiting at the k-th arm after approach i. In this order, the n-th movement represents U-turn, which exits back at approach i. Taking the example of a four-arm roundabout as in Figure 1, the first movement is turning left (L), second going through (T), third turning right (R), and fourth making U-turn (U). We further develop the relative turning function $d(k,i) = \mathrm{mod}(i+k-1,n)+1$ where $\mathrm{mod}(p,q) = p - \mathrm{int}(p/q)$ is a remainder function. Physically, $d(k,i)$ represents the k-th approach from approach i in a clockwise order and therefore gives the exit arm of the k-th movement with entry approach i.

Flow separation and lane utilization. Figure 2 illustrates how the flow separation methodology is applied to analyze the channelization effect within the circle of a double-lane roundabout. Left and right entry lanes are measured separately. Entry flows on right lane enter inner lane and are affected by circulating flows on both inner lane and outer lane, whereas entry flows on the left lane enter into the outer lane and only give way to circulating flow on outer lane.

Arrival flow modeling. For any approach i, arrival flow on right entry lane consists of all traffic flows going to enter inner circular lane. Arrival flow on left entry lane consists of flows entering outer circular lane.

$$q_a^R(i) = \sum_{k=1}^{N} \alpha_{i,d(k,i)} D_{i,d(k,i)} = \sum_{j=1}^{N} \alpha_{ij} D_{i,j} \tag{1}$$

$$q_a^L(i) = \sum_{k=1}^{N} [1-\alpha_{i,d(k,i)}] D_{i,d(k,i)} = \sum_{j=1}^{N} (1-\alpha_{ij}) D_{i,j} \tag{2}$$

Conflicting flow modeling. Vehicles on right entry lane are entering inner circular lane and thus must give way to circulating vehicles on both inner and outer lanes. The conflicting flow for waiting vehicles on right lane of approach i is

$$q_c^R(i) = \sum_{k=1}^{n} \sum_{l=n-k+1}^{n} D_{d(k,i),d[l,d(k,i)]} \tag{3}$$

The conflicting flow for waiting vehicles on left entry lane of approach i, $q_c^L(i)$, only counts circulating flows on outer lane. It is not necessary to give way to circulating flows on inner lane. The conflicting flow can be written as

$$q_c^L(i) = \sum_{k=1}^{n} \sum_{l=n-k+2}^{n} (1-\alpha_{d(k,i),d[l,d(k,i)]}) D_{d(k,i),d[l,d(k,i)]} + \sum_{k=1}^{n} D_{d(k,i),d[n-k+1,d(k,i)]} \tag{4}$$

Notice that inner lane traffic exiting at arm $d(1,i)$ will change into outer lane before exiting and thus contribute to $q_c^L(i)$.

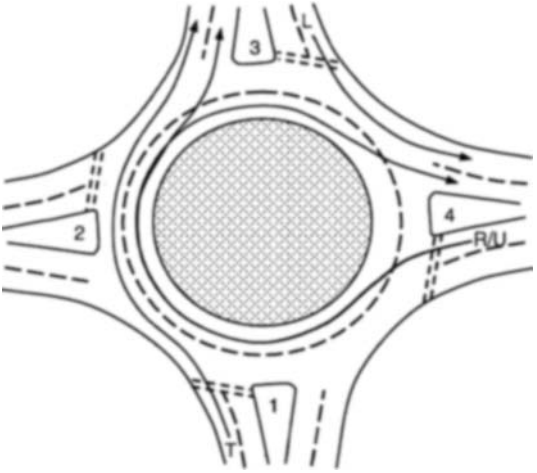
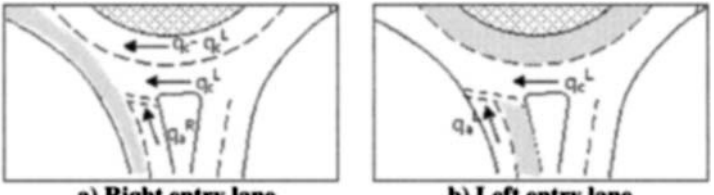


Figure 1. Double-lane roundabout and four turnings



a) Right entry lane **b) Left entry lane**

Figure 2. Flow separation

Capacity Analysis

Roundabout capacity formula in HCM (2000) is adopted for entry capacity study. We further utilize the concept of reserve capacity (Wong, 1996) to assess the overall roundabout performance.

Entry capacity analysis. Entry capacity Q_E is defined as a monotonic decreasing function of conflicting flow q_c. Entry capacity decreases as circulating flow increases, because there are then fewer chances for waiting vehicles to enter the circulation. When entry opportunities do present themselves, the number of vehicles possible to enter depends on the geometric characteristics of the roundabout layout. HCM (2000) formula gives

$$Q_E = q_c \exp(-q_c t_c / 3600) / [1 - \exp(-q_c t_f / 3600)] \tag{5}$$

where t_c is the critical gap in gap acceptance and t_f is the follow-up time. The critical gap and the follow-up time represent the geometric characteristics which affects entry flow movement. Entry capacity gives the maximum entry flow and bounds the arrival flow as

$$q_a \le \mu Q_E \tag{6}$$

where μ is the maximum acceptable degree of saturation (DoS).

Entry capacity is then calculated separately for each entry lane. Since q_c^L is a single stream, Eq. (5) can be directly adopted for left entry lane. For right entry lane, the conflicting flow consists of two streams. If independent Poisson arrival is assumed for flows both on outer lane and on inner lane, then the combined flow can also be modeled as Poisson arrival with a collective density. Thus, q_c^R fulfills the same role of q_c^L in gap acceptance theory and Eq. (5) can be used for right entry lane as well.

$$Q_E^R = q_c^R \exp(-q_c^R t_c^R / 3600) / [1 - \exp(-q_c^R t_f^R / 3600)] \tag{7}$$

$$Q_E^L = q_c^L \exp(-q_c^L t_c^L / 3600) / [1 - \exp(-q_c^L t_f^L / 3600)] \tag{8}$$

where the specific parameters are applied. Entry flows are thus bounded by

$$q_a^R(i) \le \mu Q_E^R(i) \text{ and } q_a^L(i) \le \mu Q_E^L(i) \tag{9}$$

As both q_a and Q_E are functions of $\{D_{i,j}\}$, the capacity constraint can be written in vector form as

$$\mathbf{q}_a(\mathbf{D}) \le \mu \mathbf{Q}_E(\mathbf{D}) \tag{10}$$

Reserve capacity analysis. A uniform growth factor is applied to compute the capacity reserve of the roundabout for a certain flow pattern $\{D_{i,j}\}$. A multiplier γ is imposed to the demand $D_{i,j}$ on each pair (i, j). The entry capacity constraint will require this growth factor not to exceed certain limit. The capacity constraint gives

$$\mathbf{q}_a(\gamma\mathbf{D}) \le \mu \mathbf{Q}_E(\gamma\mathbf{D}) \tag{11}$$

or equivalently,

$$\mu \mathbf{Q}_E(\gamma\mathbf{D}) - \mathbf{q}_a(\gamma\mathbf{D}) \ge 0 \tag{12}$$

It is easily verified that $\mathbf{F}(\gamma) = \mu \mathbf{Q}_E(\gamma\mathbf{D}) - \mathbf{q}_a(\gamma\mathbf{D})$ is monotonically decreasing with γ. Thus there is a bound $\gamma_{\max} = \min\{\gamma_{\max}^R(i), \gamma_{\max}^L(i)\}$ such that, for any $\gamma \le \gamma_{\max}$, (12) holds. Here, $\gamma_{\max}^R(i)$ and $\gamma_{\max}^L(i)$ are determined by

$$\gamma_{\max}^R(i) = \max\{\gamma : \mathbf{q}_a(\gamma\mathbf{D})_{(i)}^R \le \mu \mathbf{Q}_E(\gamma\mathbf{D})_{(i)}^R\} \tag{13}$$

$$\gamma_{\max}^L(i) = \max\{\gamma : \mathbf{q}_a(\gamma\mathbf{D})_{(i)}^L \le \mu \mathbf{Q}_E(\gamma\mathbf{D})_{(i)}^L\} \tag{14}$$

Numerical Study

Three case studies are provided to illustrate the methodology. All cases are based on the same four-arm double-lane roundabout where critical gap and follow-up time are calibrated to be, respectively, 4.2 and 3.0 seconds for all entry lanes on all entry arms. While the methodology can be readily extended for study of unbalanced arrivals, arrival flow patterns from different arms are assumed identical with a total arrival flow of 100 for each arm. Flow patterns of turning movements are listed in Table 1. Data in Case I are collected from field surveys. It represents the flow pattern of balanced movements. Cases II and III are manually created to represent unbalanced turnings: Case II with predominant left-turns and through traffic; Case III with many right-turns and U-turns.

Three lane use schemes are used in each case study. A fixed proportion of flow for each movement approaches right entry lane and enters inner circulatory lane. The percentages are listed in Table 2. Scheme a) has most road users compliant with the advised lane choices by a typical Drivers' Manual: left-turn uses the outer lane; right- or U-turns use the inner lane; and through traffic may use either. Scheme c) represents drivers' aversion against using the inner lane, as it is found in field survey that many right-turn drivers tend not to use the inner lane, possibly due to potential conflicts at exit. Scheme b) is a compromise between a) and c).

Table 1. Flow patterns in three studied cases

Case \ Movement	L	T	R	U	Sum
Case I - Field Data	22	36	29	13	100
Case II - L&T Predominant	30	50	20	0	100
Case III - R&U Predominant	20	20	40	20	100

Table 2. Percentage of right & inner lane usage

Scheme \ Movement	L	T	R	U
a	0	50%	80%	90%
b	0	50%	70%	70%
c	0	30%	50%	50%

The results are shown in Table 3. Comparing the three cases, it can be seen that flow patterns with high percentage of left-turn and drive-through movements usually provide higher capacity. Case II offers the highest capacity reserve, as 80% of traffic are left-turn or drive-through traffic, compared to 58% in Case I and 40% in Case III. This is reasonable, because the increased number of drivers making U-turn or right-turn will travel further around the roundabout, conflicting with more entry flows and thus reduce the capacity.

Among the three schemes of lane choices, Scheme b) is found to harbor a bigger capacity in most cases. This is anticipated due to a more balanced distribution of flows as shown in Table 3. Different lane usage choices generate different capacities. For a specific flow pattern, there exists the optimal lane usage scheme to maximize

the total capacity. It can be seen that balancing the usage on the two circular lanes would introduce higher capacity. Table 4 compares the dispersion of flows on the two circular lanes. It is estimated that the optimal ratio of lane usage would range between (0.45:0.55) and (0.55:0.45).

Table 3. Results under maximum DoS of 0.85

Case	Scheme	γ_{max}	q_a^R	q_c^R	Q_E^R	DoS	q_a^L	q_c^L	Q_E^L	DoS
I	a	*8.21*	434.4	1092.3	511.1	*0.85*	386.8	709.6	694.6	*0.56*
	b	*8.70*	412.2	1156.7	485.0	*0.85*	457.5	821.9	635.4	*0.72*
	c	*7.87*	250.2	1046.4	530.5	*0.47*	536.6	830.0	631.3	*0.85*
II	a	*9.85*	404.0	886.8	603.2	*0.67*	581.4	729.2	683.9	*0.85*
	b	*9.55*	372.6	859.8	616.4	*0.60*	582.8	726.1	685.6	*0.85*
	c	*8.21*	205.2	738.7	678.8	*0.30*	615.6	656.7	724.3	*0.85*
III	a	*7.04*	422.5	1126.6	497.0	*0.85*	281.7	647.8	729.3	*0.39*
	b	*7.58*	394.0	1212.2	463.5	*0.85*	363.7	787.9	652.8	*0.56*
	c	*7.74*	278.8	1239.0	453.4	*0.61*	495.6	929.2	583.1	*0.85*

Table 4. Ratio of lane usage by volume: inner lane to outer lane

Case\Scheme	a	b	c
I	*0.65:0.35*	*0.57:0.43*	*0.39:0.61*
II	*0.52:0.48*	*0.48:0.52*	*0.32:0.68*
III	*0.72:0.28*	*0.62:0.38*	*0.43:0.57*

Conclusions

This paper proposed a flow separation methodology in analyzing the entry capacity and reserve capacity of multi-lane roundabouts, which considers flow interactions and drivers' habitual lane choices. Drivers' lane usages can be considered as a design parameter so as to balance the flows on the two circulating lanes, thereby maximizing the junction capacity. This may be achieved by suitably refining the regulations and giving advices on lane usages.

Acknowledgements—This study is partially supported by the Competitive Earmarked Research Grants from the Research Grants Council of the Hong Kong Special Administrative Region (HKUST6283/04E and HKU7031/02E).

References

Baranowski, B. (2004). "History of modern roundabout." *RoundaboutsUSA*, http://www.roundaboutsusa.com/history.html (Jul. 15, 2004).

Transportation Research Board (TRB). (2000). *Highway Capacity Manual*, Washington, D.C.

Wong, S.C. (1996). "On the reserve capacity of priority junctions and roundabouts." *Transportation Research B* 30(6), 441-453.

Arizona's Experience with Context Sensitive Solutions to Implement NEPA

Deeb, Maria Angelica, P.E., M.E.P., [1] and Heidel, K. J. Ph.D.[2]

[1] City of Mesa, Transportation Division, 300 East Sixth Street, P.O. Box 1466, Mesa, AZ 85211-1466; PH (480)861-6504; email: mariaacafe2002@yahoo.com

[2]EcoPlan Associates, Inc., 701 W. Southern Avenue, Suite 203, Mesa, AZ 85210; PH (520) 299-2799; email: kheidel@azdot.gov

Abstract

Since its inception in 1969, the National Environmental Policy Act (NEPA) has provided a basis for preventing damage to the environment, stimulating the health and welfare of citizens, and enriching the understanding of ecological systems and the natural resources important to the nation. Processes and understanding for the implementation of NEPA continue to evolve and improve. In transportation facility planning, design and construction, an innovative and effective use of NEPA is referred to as Context Sensitive Solutions (CSS). CSS provides a framework for elevating the traditional NEPA – highway design process to a need based concept approach for planning and design. The CSS concept emphasizes the same concepts as the NEPA decision framework, promoting a multidisciplinary collaboration to develop transportation resources in the context of physical, scenic, aesthetic, historic, environmental, safety, and mobility constraints. CSS emphasizes in a more direct manner the need to have community insight as well as defining at the early planning stages the community values (defining the context where project is located) while striving for interdisciplinary collaboration as a mechanism for system consideration and coordination, to evaluate all aspects of purpose and need.

The Arizona Department of Transportation (ADOT) and the Federal Highway Administration (FHWA) in Arizona are using some elements of CSS for analyzing successes and identifying areas of potential improvement (lessons learned) in implementation of NEPA in project specific situations by:

- Using a documented, systematic and interdisciplinary approach;
- Giving the environment appropriate consideration with economic and technical considerations; and
- Including in proposals details on the environmental impacts, adverse impacts that can't be avoided, alternatives to proposed action and, consequences of taking the proposed action.

Components and approaches to a CSS program within Arizona are evaluated and discussed. Elements and challenges of a particular ADOT example are presented, along with lessons learned.

CSS and the Project Development Process

CSS is a collaborative, interdisciplinary approach that allows all stakeholders to develop a transportation facility that fits its physical setting and preserves scenic, aesthetic, historic and environmental resources, while maintaining safety and

mobility. Multiple resources document the ways that CSS considers the total context within which a transportation improvement project will exist. (Myerson, 1999 and Marriott, 1998)

Emerging technologies are transforming the way State Departments of Transportation (State DOTs) plan and implement new transportation infrastructure and improvements to existing networks. In recent decades, innovative Geographic Information Systems (GIS) have been combined with modeling tools, spatial analysis, data management, collaboration, and outreach to promote environmental streamlining and stewardship. CSS processes, together with GIS, clarify roles for proper communication and information dissemination within NEPA process, enhancing public involvement. CSS also fosters the coordination with interagency groups about protocols and standards pertaining to data, information management, modeling tools, and information security (Volpe, 2005). The CSS properly utilizes software, hardware, data, and personnel to effectively handle large volumes of data, analyze complex results, and present geographically referenced information. CSS provides tools so that multiple issues surrounding a particular project area can be displayed and analyzed, enhancing collaboration, recognizing benefits and costs, and supporting decision making based on a common –science- based database. Properly applied, CSS serves as the foundation for streamlined decision making, ensuring that all stakeholders base their analysis, judgments, and opinions on the same information. CSS then becomes an effective system for decision-making.

Using CSS, transportation professionals and communities go beyond the traditional transportation project development process to create new collaborative partnerships with carefully identified stakeholder groups. Transportation departments then design, construct, operate, and maintain transportation projects while also maintaining constant communication with stakeholders and project partners. Some of the new ways of working that have been identified as part of CSS (FHWA- European Roads, 2001) include efforts to: (i) identify stakeholders and project partners in the preliminary scoping phase; (ii) write a statement of purpose including not only transportation needs assessment, but also a statement of environmental and community values. Identify potential project problems, create a consensus vision, and establish project goals or criteria which can later serve as measures for evaluating the completed project; (iii) include non-traditional solutions, such as using alternative routes or modes as part of the evaluation of alternatives and alternative designs and allow stakeholders the ability to assess the advantages and disadvantages of a variety of approaches to addressing a project's "purpose and need"; (iv) keep track of comments made in early phases as an essential part of a quality design process. The final design is crucial in determining the project's ultimate impacts on the road's context; (v) perform consensus building during the entire process to accelerate later review and approval processes. Accelerate approvals from other state, local or federal agencies by involving them as stakeholders in the process from the very beginning. Use CSS to identify community concerns through early participation in the process to avoid conflict and community opposition at the approval stage of a project; (vi) maintain communications with concerned stakeholders by sharing information on

post-planning project changes, including schedule delays, modifications to mitigation or community-desired improvements, or changes to construction detours. Commitments made during the project development process should be honored to the maximum extent possible during the final phases of the project; (vii) CSS principles carry through under maintenance and operation agreements with communities. The maintenance of CSS roads involves more than cleaning and repair, and includes ongoing monitoring and modifications of road operations and design; and, (ix) use both qualitative and quantitative performance measures to provide feedback and to improve other CSS projects. These measures range from award programs to traffic counts and other "hard" data.

CSS and ADOT

ADOT has faced multiple challenges in developing its CSS processes. Not only is Arizona experiencing rapid growth and changing demographics, ADOT is also challenged by the Arizona's especially diverse climate zones, ranging from low desert plains in the south to high mountain alpine areas in the north. Federal or state governments and tribes, each of which has separate additional participation and partnering requirements for transportation decisions, hold vast tracts of Arizona.

Recently, ADOT contracted with the American Association of State Highway Transportation Officials (AASHTO) to complete a broad reaching, agency wide, internal assessment of its environmental performance. (Kober, 2004) The work considered ADOT's programs in a total context, including not only the regulatory framework, but also additional opportunities for voluntary improvements. Representatives from throughout the agency at all staff levels participated in a series of workshops to consider opportunities and strategies for improving ADOT's environmental performance and stewardship. Eleven Technical Memoranda were prepared and reviewed involving13 environmental performance areas. ADOT has initiated implementation of recommendations from this report.

ADOT CSS Example

In 1999, ADOT implemented CSS principles as part of the project known as *State Route (SR) 82: Rock-fall mitigation and safety project.* The project was initially designed as a straightforward roadway realignment and rock-fall ditch sufficient to prevent dangerous conditions resulting from rocks falling onto the highway during and after storm events. The local community was actively opposed to the original project design, and several interest groups and organizations came forward. These groups helped the project team understand environmental and community concerns, including the desire to protect a nearby stream and rare bird habitat, and to preserve the area's scenic character.

ADOT management re-evaluated the project scope and developed a Public Involvement and Environmental Mitigation Process. ADOT engaged a broad spectrum of stakeholders in a team: local residents, other interested citizens, ADOT staff, consultants, and contractors. ADOT engaged the services of a professional facilitator to assist the team in developing and evaluating seventeen alternatives using

criteria related to purpose and need, engineering feasibility, socioeconomic and environmental resources, and public safety. More than 50 measures were developed to mitigate potential social, economic, and environmental impacts that could result from implementation of the selected construction alternative. Mitigation measures were selected which would not only maintain access, but would also minimize the amount of disturbance and the loss of foliage; the potential for erosion and sediment release into the waterway, protect nesting rare birds, reduce the frequency and duration of traffic disruptions, and maintain the area's visual quality.

Before the construction phase of the project, partners attended a two-day workshop, focusing first on environmental sensitivities and mitigation measures, and then on partnering. Partnering goals included not only being responsive to the unique environmental sensitivities of the area, but also a number of elements related to professionalism, cooperation, respect, being proactive, listening, safety, and maintaining the schedule. The outcome of this successful workshop was a formal partnering agreement signed by all parties.

ADOT CSS Lessons Learned

Despite a contentious beginning, the planning and construction phases of the SR 82 Rockfall Project were marked by teamwork and cooperation, resulting from CSS-based Public Participation and Environmental Mitigation Process efforts. Innovative program elements such as weekly briefings for construction staff kept environmental concerns at the forefront of workers' awareness, and a specialized public information program kept residents and visitors apprised of restrictions and progress. In addition, the hiring of not only a facilitator, but also a private local environmental monitor increased trust. Additional creative elements included the assessment of liquidated damages for destruction of trees; the careful specification of terms (i.e. a definition of the word "tree") and a well-defined plant salvage plan. Final evaluations for the collaborative process were among the highest in ADOT's thirteen-year partnering history.

Case Study Conclusions

Using CSS in the project integrated highway development and enabled effective utilization of flexible design standards, policies, codes, and guidelines, and the consideration of design exceptions when feasible in order to make highway improvements more compatible with the surrounding communities and environment. A close out report (PM:A, 2004) detailed exemplary CSS activities during the planning, design, and construction process. Ultimately, the Patagonia Rock-fall Public Involvement and Environmental Mitigation Process received state and national recognition from the National Partnership for Highway Quality.

System Summary

Both the literature and the case study were used to summarize and systematize CSS efforts. The following paragraphs are broken out by the related transportation project phase or agency that corresponds to the recommended implementation timeframe. Several of recommended items listed need to be further developed; ADOT and other transportation professionals can expand their utility. As the

recommendations are discussed, and further refined, they can form a basis for CSS implementation not only in ADOT, but also in other state Department of Transportation agencies.

Recommendations related to the Project Assessment

Recommendation 1.

At the Project Assessment (PA) project phase there needs to be a review and revision of the current PA procedures to better identify community and environmental issues. The process must also identify and list specially detailed requirements by which projects in community-sensitive or environmentally sensitive areas (especially on designated scenic routes) are placed in a special category in which community sensitivity/activism, unique environmental conditions and sensitivities, and any other special designations for resources in the area are considered.

ADOT Response.

ADOT has implemented Policy 88-2, which addressed PA documents. The policy is now posted in ADOT website (http://www.azdot.gov/Highways/RdwyEng/RoadwayPredesign/PDF/memos.pdf). To enhance the information on the PA and support transportation design and engineer specialists, ADOT's Environmental & Enhancement Group (EEG), Transportation Enhancement Section has a website that lists all scenic routes (http://www.azdot.gov/Highways/EEG/enhancement_scenic_roads/scenic_roads/index.asp) and has a website that lists all transportation related environmental considerations. (http://www.azdot.gov/Highways/eeg/index.asp) EEG also provides monthly lunchtime training events open to all ADOT staff.

Recommendation 2.

Ensure that current and full data and information (e.g., topographic mapping) are made available at the beginning of the project analysis stage; if current and full data and information are not available at the beginning of the PA stage, they should be made available prior to initiating final design.

ADOT Response.

The Department's Process Manual (ADOT, 1995) provides an overview of the agency, and the process and requirements in each step of the process, through which individual projects are conceived, developed and constructed. Information access and management is listed as part of the development of the document, which is developed and reviewed to ensure validity. No updates are known of the manual. (http://www.azdot.gov/highways/ppms/ProjDevProcMan.pdf)

Recommendation 3.

Consider community and environmental sensitivities in assessing the overall project budget and schedule in addition to quantifiable measures such as project condition and size.

ADOT Response.

In 2001 ADOT's Program and Project Management Group (PPMS) developed a training course on *Implementation of Design Scheduling*, which is accompanied by a manual that allows for flexible transportation project scheduling. These, accompanied by ADOT's training course on *Successful Management of the Process Development process*, which stresses the importance of well developed schedules and cost budgets, should allow for project managers to generate schedules and budgets that ensure that this recommendation is addressed. (http://www.azdot.gov/Highways/PPMS/DesignSURFTrainingRevised.pdf).

Recommendation 4.
Ensure that all projects in community sensitive or environmentally sensitive areas emphasize early scoping efforts, including agency and public input.

ADOT response.
ADOT, in partnership with FHWA, has made available training regarding CSS for ADOT staff, developing concepts stated in *Flexibility in Highway Design* (FHWA, 1997). ADOT partners with consultants and other agencies to ensure that information about sensitive areas is identified early in the process (seeking input through dissemination of scoping letters) and findings are included in all project documents as part of the design and decision making process. ADOT EEG has guidance for the development of these letters as well as the agencies and community organizations that need to receive scoping letters.

Recommendations related to the Public Involvement

Recommendation 1.
Ensure early involvement of the affected public, including the local community, the interested community, as well as other stakeholders.

ADOT response.
ADOT created in 2004 a new office to improve public involvement. The new office is named: Communication and Community Partnership Office. Their mission addresses the recommendation listed above. (website: http://www.azdot.gov/ccpartnerships/index.asp)

Recommendation 2.
To encourage public input during the scoping process, provide adequate detail on the range of alternatives, scope items under consideration, and articulate the known range of potential issues.

ADOT response.
During the scoping, planning, and design, EEG oversees the preparation of environmental documents for all highway improvement and enhancement projects. These documents insure that all relevant environmental factors are appropriately addressed and mitigated. The primary documents include Environmental Determinations (Categorical Exclusions) for projects that do not have significant environmental impacts, Environmental Assessments (EAs) for projects that may have significant impacts, and Environmental Impact Statements (EISs) for projects having significant impacts. Supplemental documents include Section 4(f) Evaluations for projects that impact publicly owned recreation areas, wildlife refuges, or historic sites. The development of these documents jump starts the scoping process, and as part of the planning process, EEG implements an effective public involvement program to assist ADOT in making sound project decisions. The primary public involvement activities include Public Meetings and Public Hearings. They are conducted in an open-forum format with optional brief presentations. The Public Hearings also feature a court reporter who prepares a verbatim transcript of the hearing's proceedings.

Recommendation 3.
Ensure that ADOT follows through with all commitments made to the community.

ADOT response.
ADOT's commitments are all listed in the project documents as part of the Contract & Specification documents, mitigation section (ADOT, 1995). These commitments list the responsible party for implementation and follow-up, as well as the timeframe for completion. ADOT maintains high standards of professional ethics.

Recommendations related to the Bidding and Contract Operations

Recommendation 1.

Assure that the bid information and Plans and Special Provisions clearly identify and articulate the critical aspects of a project.

ADOT response.

ADOT 's Contract and Specifications process requires that a project bid only when an Environmental Clearance, Utility Clearance and a Right of Way Clearance have been obtained. All clearances list the requirements designated as mitigation measures to protect the sensitivity of the project area. ADOT construction personnel should have special knowledge of the issues addressed in these clearances in order to determine if the contractor is complying with all the Mitigation Measures, and be able to impose penalties for lack of compliance.
(http://www.azdot.gov/Highways/cns/index.asp)

Recommendation 2.

Require that all contractors who intend to bid on projects in community or environmentally sensitive areas attend a pre-bid meeting, which will underscore the sensitivity and needs for protection in the project area. Bids will thus include all relevant elements addressing these special requirements.

ADOT response.

ADOT has developed a Partnering Process that fosters communication between all stakeholders early in the design process, and that continues through the bidding award process. A pre-bid meeting is scheduled inviting all potential bidders, if the project requires it. The contact information of the ADOT Project manager is also listed in the bid documents in order to answer any questions, if needed. Before construction they facilitate a Pre-construction partnering meeting between ADOT, stakeholders and the contractor addressing all mitigations and concerns.
(http://www.azdot.gov/CCPartnerships/Partnering/
Policy/ITD_Policy_on_Partnering_signed.pdf)

Recommendation related to FHWA Project Involvement

Recommendation.

Federal projects encourage FHWA's involvement through the PA, planning, environmental evaluation, pre-design, design, and construction stages. Ensure that FHWA representatives are included in both Preconstruction, Partnering, and Construction Closeout workshops. This will allow FHWA to provide valuable information to the team and to make knowledgeable and timely decisions as the project progresses.

ADOT response.

ADOT continuously trains staff in the need for keeping FHWA involved at all stages of transportation project development process. ADOT and FHWA in Arizona have been using this philosophy for more than a decade. For example the SR-87 project (the last 4-lane section to Payson) has been open to traffic for many years. During the development of the EIS ADOT, FHWA along with the Tonto National Forest used the basic CSS concepts to develop a series of award winning construction projects. Other projects to consider would be the SR-179 route thru the Village of Oak Creek and Sedona, the Super Red-Tan system interchange, and more recently the on-going efforts dealing with elk crossings on SR-260. From a planning viewpoint, some ADOT and FHWA CSS examples are: the multi-agency efforts to deal with Habitat Connectivity on a state-wide basis, the development of an updated Highway Guidelines Manual on Public Lands and the research currently being done on US-93 (Lake Mead) dealing with Bighorn Sheep.

National Highway Institute (NHI) courses such as: *NEPA and Transportation Decision Making Process* and *Linking Planning and NEPA: Towards Streamlined Decision-making* emphasize this issue in their training.

Recommendation related to Contractor Involvement

Recommendation.

Assure that all underlying objectives are determined, defined, and discussed prior to construction. All key members of the construction team, including the contractor and key subcontractor personnel need to work with the members of the construction team. Assure that all questions regarding the critical aspects of a project, especially those measures which affect the sensitivity of the proposed project, have been resolved prior to initiation of construction.

ADOT response.

The CSS process allows for agency standards and criteria to be developed through the scoping and planning process, and to mold the outcome of the project's design process. ADOT's example project used performance measurement standards to define and evaluate mitigations measures and evaluate construction compliance; these standards were part of the information disclosed and discussed in the pre-construction meeting.

Recommendation related to Public Information.

Recommendation.

Conduct regular meetings with the public to advise them of what is planned to happen in their area. Assure that public interest and information flow contributes to a knowledgeable community as well as a knowledgeable project team.

ADOT response.

ADOT has conducted semi-annual meetings, and used a broad based team approach in some projects. The approach depends on the District or route involved. ADOT has links for most routes under their jurisdiction, as well as an ADOT website (www.azdot.gov).

Conclusions

CSS can present a powerful tool to enhance the way transportation departments do business. ADOT is working to incorporate useful CSS elements into its full spectrum of operations and procedures.

CSS fits well into the existing transportation project development process.

Transportation agencies can use CSS to:

- Develop project concepts through needs study evaluation and outside requests while addressing long range transportation planning.
- Use project concepts with data gathering and programming prioritization to define alternatives.
- Update agency standards using CSS criteria to perform preliminary engineering and develop preferred plans and alternatives.
- Continue CSS concepts through final design and construction.

CSS principles improve the transportation planning process by not only addressing transportation needs but also:

- Enhancing community assets.
- Protecting the Natural & Built environments.

CSS processes can be part of the way transportation departments do business.

- Assemble, utilize and manage diverse technical resources.
- Integrate environmental & engineering technical tasks.
- Incorporate meaningful public involvement.
- Consider multiple alternatives.
- Maintain open and honest communications and decision processes.
- Fully document all key decisions.

References

Arizona Department of Transportation (ADOT). (1995). "Project Development Process Manual".

Arizona Department of Transportation (ADOT). website. <(www.azdot.gov)>

Federal Highway Administration (FHWA). (1997). *Flexibility in Highway Design* (FHWA Pub. No. FHWA-PD-97-062).

Federal Highway Administration (FHWA). (2001). *Geometric Design Practices for European Roads- International Technology Exchange Program* (FHWA Pub. No. FHWA-PL-01-026 June 2001).

Kober, Wayne. (2004). "*Strategies to Integrate Environmental Stewardship Into ADOT's Business*", ADOT final report 543, FHWA-AZ-04-543.

Myerson, Deborah L. (1999). "Getting It Right in the Right-of-Way: Citizen Participation in Context-Sensitive Highway Design," *Scenic America.*

Marriott, Paul Daniel. (1998). *Saving Historic Road: Design and Policy Guidelines*, Preservation Press and John Wiley & Sons, Inc.

Nevada Department of Transportation (2002), *Pattern and Palette of Place: A Landscape and Aesthetic Master Plan for the Nevada State Highway System.*

PM:A/Planning & Management: Applied. (2004). *Results of Close-out Construction partnering workshop, STP 082-A(006)A/ H423901C Nogales – Tombstone highway (SR 82) / Rock-fall Mitigation and Safety.*

Volpe National Transportation System Center for the Federal Highway Administration Office for Project Development and Environmental Review (2005). "GIS for Environmental Stewarship and Streamlining: An Overview of State DOT Practices".

Architecture for On-Line Deployment of DynaMIT in Hampton Roads, VA

Byungkyu "Brian" Park[1], Devi Manohar Pampati[2], Ramachandran Balakrishna[3]

[1]*Assistant Professor, Department of Civil Engineering, University of Virginia, Charlottesville, VA 22903. Phone: (434) 924-6347. Fax: (434) 982-2591. E-mail: bpark@virginia.edu*
[2]*Graduate Research Assistant, Department of Civil Engineering, University of Virginia, Charlottesville, VA 22903. Phone: (434) 924-1420. Fax: (434) 982-2951. E-mail: dmp5e@virginia.edu*
[3]*Doctoral Candidate, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139. Phone: (617) 252-1116. Fax: (617) 252-1130. E-mail: rama@mit.edu*

Abstract

This paper presents the architecture and implementation for the online deployment of DynaMIT, one of two traffic estimation and prediction systems developed under the aegis of the Federal Highway Administration, in Hampton Roads, VA. Appropriate interfaces between real-time data (sensor counts and incident information) and DynaMIT were developed and tested. In addition, the run time of DynaMIT on a 64-bit computer was tested to ensure that online deployment could meet real-time requirements.

Introduction

Intelligent Transportation Systems (ITS) have been developed for the real-time management of urban traffic networks. Such systems harness advanced technologies and modeling techniques to generate route guidance and alleviate congestion due to perturbations such as incidents. It is widely believed that route guidance based on short-term predictions are more effective in reducing travel times than approaches based on instantaneous measurements (*Ben-Akiva et al. [2,3]*). Prediction-based guidance generation systems such as DynaMIT (*Ben-Akiva et al. [1]*) and DYNASMART (*Mahmassani [4]*) provide guidance by monitoring current traffic conditions and predicting traffic conditions in the near future. These advanced systems can improve network performance by allowing traffic management centers (TMC) to operate in a proactive rather than reactive manner.

The first step for realizing the benefits of Traffic Estimation and Prediction Systems (TrEPS) is to explore the feasibility of implementing such systems on-line in a TMC. Integration of a TrEPS with real-time surveillance and incident information systems is a key step in achieving the real-time implementation of TrEPS. This paper focuses on the interfaces developed for integrating DynaMIT, one of two TrEPS prototypes, with real time information such as sensor counts and incident information for achieving online implementation. In addition to this capability, DynaMIT should also be able to process this real-time information and perform the necessary computations within a certain time frame in order for a TMC to use its predictions for real-time operations. Therefore, this paper also discusses the computation performance of DynaMIT and its feasibility in online operations. Thus, The primary objective of this paper is to demonstrate the feasibility of online integration of DynaMIT in a TMC. The following requirements must be met towards this objective:

- Integration of Surveillance and Incident Management subsystems with DynaMIT
- Real-time running of DynaMIT with 5-minute estimation and 30-minute prediction intervals

Overview of DynaMIT

Dynamic network assignment for the Management of Information to Travelers (DynaMIT) is a DTA system designed to generate consistent, anticipatory route guidance. DynaMIT combines a microscopic demand simulator and a mesoscopic supply simulator to capture complex demand and supply processes and their interactions. Models of origin-destination flows, pre-trip and en-route driver decisions, traffic dynamics, queuing and spillback allow the system to estimate and predict network state in a realistic manner. DynaMIT is designed to prevent overreaction by ensuring that the generated guidance is consistent with the conditions that drivers are expected to experience. This is achieved through explicit modelling of drivers' reaction to information. The flexible simulation system can adapt to diverse Advanced Traveller Information System (ATIS) requirements, and is designed to handle a wide range of scenarios including incidents, special events, weather conditions, highway construction activities and fluctuations in demand.

On-Line Integration Methodology

This section outlines the methodology followed for performing the integration of DynaMIT in the Hampton Roads region. This includes developing a set of requirements for the integration of DynaMIT with surveillance, incident management and information dissemination subsystems. Although most of these requirements are fairly generic and may be applicable to many traffic management centers, some level of customization is unavoidable depending on the complexity of the present architecture in various traffic management centers.

Surveillance Integration Requirements

The surveillance system routinely collects traffic measurements from the network, and stores them for future use. The system may be comprised of a wide variety of sensing technologies, including standard loop detectors, acoustic sensors and closed-circuit television cameras. Traffic data obtained from the surveillance system typically consist of estimates of flows, speeds and occupancies. Such data must be communicated to DynaMIT immediately, in order to facilitate state estimation and timely guidance generation. Data communication requirements for the surveillance system interfaces include the ability to

- Process data simultaneously from various measurement sources accessible by the TMC.
- Map the identification numbers of sensors on the field with those coded in the network used by DynaMIT.
- Flexibly aggregate traffic data to the estimation time interval width used by DynaMIT.
- Automatically process and transmit a new set of relevant traffic data every estimation interval.
- Report measures of traffic data quality/reliability for each interval and for each sensor, to assist in the process of assigning lower weights to data associated with detector malfunctions.

- Convert the data obtained from different sources into DynaMIT input format, and only transmit the data that is relevant for running DynaMIT.

Incident Management Interface Requirements

Incident detection can be a non-uniform process, wherein the information about an incident can be obtained from various sources such as surveillance cameras, drivers, moving patrols, etc. The scope of this interface is limited to the provision of the necessary inputs needed for DynaMIT to simulate the incidents and their impact on traffic conditions. Incident data interfaces must satisfy the following needs:

- After the detection/confirmation of an incident, the interface must provide an interactive method for mapping the incident location to the DynaMIT network. The TMC operator must be able to instruct DynaMIT on the affected segments.
- After the selection of the segment, the interface should allow the operator to enter the parameters that are required by DynaMIT. These parameters are start time of the incident, end time of the incident and the available capacity during the incident (the fraction or percentage of the original capacity).
- Once the operator enters the required parameters, the interface must save the information to a file that is in DynaMIT input format.
- Once the information about a new incident is entered, the interface must append the entry to the previous incidents in the file. Also, it must provide the functionality to remove previous incidents.

Data Flow in STL

Sensor Data Flow

Figure 1 describes the flow of sensor data from HRSTC to the database in the Smart Travel Laboratory (STL) at the University of Virginia. The following points summarize the process of data flow in figure 1:

- Data is obtained from HRSTC every 1 minute through Java-based Database Connection (JDBC).
- This sensor data is screened under the Extraction, Transformation and Loading (ETL) process in order to differentiate between good and bad data.
- After the screening tests are performed, the data is stored in two ways: archiving process and real-time process. In the archiving process, this data is stored in a staging database and at the end of the day and then it is permanently archived in the Oracle database after processing. The real time process involves storing the obtained data directly in the Oracle database.

Incident Data Flow

As shown in Figure 2, the incident flow from Hampton Roads to the STL database is similar to the sensor data flow. Currently, the process of extraction of incident information is functioning such that as soon as the operators at HRSTC enter the information on a new incident, it is captured by the STL extraction process and stored in the Oracle database. Thus, even if the operator does not finish that particular record, the incident information is obtained. This is very important for testing the online implementation of DynaMIT at STL, as it enables the operator at STL to become aware of the incident as soon as it is detected, and take necessary steps in order to implement DynaMIT under incident conditions.

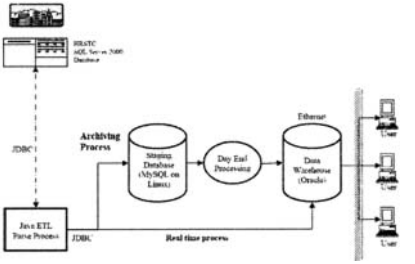


Figure 1. Sensor data flow from HRSTC to Smart Travel Lab (STL)

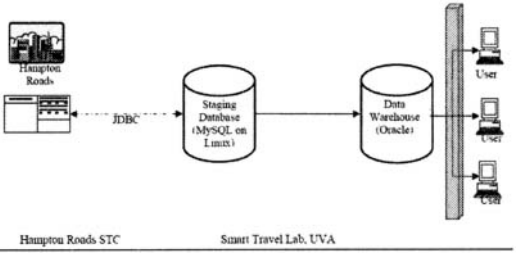


Figure 2. Incident data flow from HRSTC to Smart Travel Lab

Online Implementation

The integration approach mainly involved connecting the DynaMIT server to the Oracle database in STL via a Java Based Database Connection (JDBC). After the connection was established, a set of programs were developed for extracting the required sensor input and data quality information every 5 minutes and displaying incident information as soon as the database is updated with new incident information. The approaches followed for achieving the integration of sensor and incident information is described in detail in the following sections.

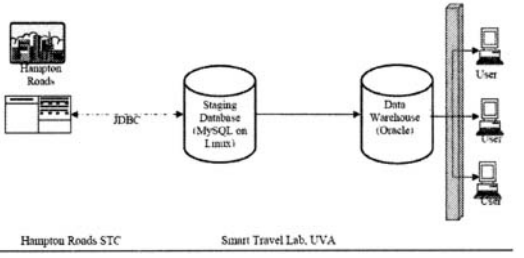


Figure 3. Integration of Surveillance data with DynaMIT-R

Sensor Data Integration

Figure 3 illustrates the approach used for performing the integration of sensor data. The main module in this process is a JAVA program that repeatedly checks the sensor flow database for the availability of data for the last 5-minute interval. If the sensor flows exist, the program extracts the sensor data and its corresponding data quality, which is then processed

by a C++ program for converting the sensor data file into DynaMIT input format. This program resides on the DynaMIT server and connects to the Oracle database.

Incident Data Integration

As shown in Figure 4, the approach followed for integrating incident information with DynaMIT is similar to the approach used for integrating sensor data: a JAVA program queries for new information every minute. Also, in order to facilitate ease of use for the operator, an alerting approach is followed: if the program detects new incident information, it alerts the operator by opening a web browser window displaying the new incident information. By parsing the information in the web browser, the operator decides the parameters that are to be entered into the incident input file of DynaMIT.

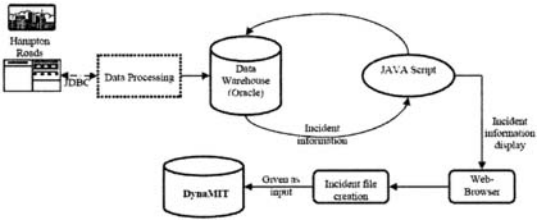


Figure 4. Integration of Incident information with DynaMIT-R

Once an incident input file is created, DynaMIT will rerun to estimate and predict traffic conditions under the incident. Since the incident duration and capacity reduction factor inserted in the incident input file may change over time, it is envisioned that the input file would be updated for each estimation interval.

Incident Input Interface

In addition to real-time access to incident information and obtaining incident alerts, an incident management sub-system also consists of an interface. This interface helps the operator in selecting the location of the incident and entering the information needed for simulating the incident in DynaMIT. After entering the necessary parameters, the interface converts the input information entered into DynaMIT input format. Figure 5 shows a snapshot of the interface developed.

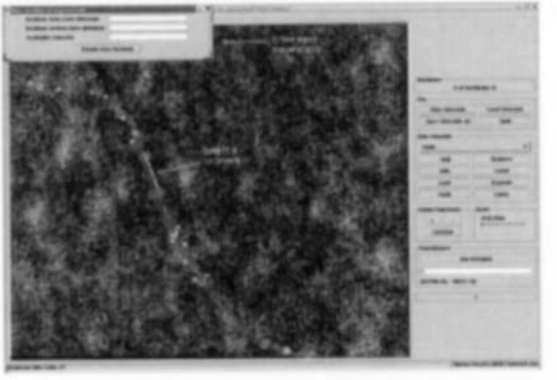


Figure 5. Incident Input Interface

DynaMIT Run-Time Analysis

As discussed earlier, a key objective for achieving the online implementation of DynaMIT is that the run time of DynaMIT for each horizon (which includes both estimation and prediction steps) should be less than the estimation interval width. In this case study, a 5-minute estimation interval and 30-minute prediction interval are desired. The run time for each horizon should therefore be less than 5-minutes. Tests were performed with a 64-bit AMD dual-processor with 2 GB of memory. It showed an average run time of 3 minutes and 10 seconds for the peak period for the Hampton roads region (4PM to 6PM). As expected, the run times for off-peak conditions were lower, indicating that real-time implementations of DynaMIT can be achieved.

Conclusions

Integration of real-time traffic sensor and incident data with DynaMIT is the first step towards the on-line implementation of DynaMIT at a TMC. This integration has been successfully completed such that sensor data is extracted every 5 minutes and converted into a DynaMIT input file, while incident data is checked every 1 minute. A user-friendly mechanism to alert the TMC operator about new incidents has been developed. An incident management interface is also developed which makes it easy for an operator to input incident information into DynaMIT. Also, with the use of a 64-bit processor, the objective of feasible run-times has been achieved, which enables DynaMIT to run in real-time at least for the HRSTC network.

Acknowledgements and Disclaimer

This material is based upon work supported by the Federal Highway Administration under Contract Number DTFH61-03-C-00107. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the Federal Highway Administration.

References

1. Ben-Akiva, M., Bierlaire, M., Bottom, J., Koutsopoulos, H.N., Mishalani, R., 1997. Development of a Route Guidance Generation System for Real-Time Application, Presented at IFAC Conference, Chania, Greece.
2. Ben-Akiva, M., de Palma, A., and Kaysi, I. Dynamic Network Models and Driver Information Systems, *Transportation Research* 25A(5), 1991, pp. 251-266.
3. Ben-Akiva, M.E., de Palma, and A., Kaysi, I. The Impact of Predictive Information on Guidance Efficiency: An Analytical Approach. In Bianco, L., and Toth, P. (Eds), *Advanced Methods in Transportation Analysis*, Springer-Verlag, Berlin, 1996, pp. 413-432.
4. Mahmassani, H.S. Dynamic Network Traffic Assignment and Simulation Methodology for Advanced System Management Applications. *Proc. 81st Annual Meeting of the Transportation Research Board.* CDROM. 2002.
5. MIT, *DynaMIT-R User's Guide version 2.0*, Intelligent Transportation Systems Program, Massachusetts Institute of Technology, February 2003.

WisTransPortal: A Wisconsin Traffic Operations Data Hub

Steven T. Parker, Ph.D.[1] and Yang Tao, Ph.D.[2]

[1]Traffic Operations and Safety Laboratory, Department of Civil and Environmental Engineering, University of Wisconsin-Madison, 1415 Engineering Drive, Madison, WI 53706; PH (608) 265-4921; FAX (608) 263-2512; email: sparker@engr.wisc.edu

[2]Traffic Operations and Safety Laboratory, Department of Civil and Environmental Engineering, University of Wisconsin-Madison, 1415 Engineering Drive, Madison, WI 53706; PH (608) 262-3542; FAX (608) 263-2512; email: yangtao@wisc.edu

Abstract

The Traffic Operations and Safety (TOPS) Laboratory at the University of Wisconsin-Madison, in partnership with the Wisconsin Department of Transportation (WisDOT), has started development of a data management system (WisTransPortal) to facilitate continuous collection and archiving of Intelligent Transportation Systems (ITS) data in Wisconsin. The basic objective of the WisTransPortal project is to develop capabilities for a statewide ITS data hub to support multiple applications in traffic operations and safety. Those capabilities include integration, management, analysis, and dissemination of real-time and historical ITS / traffic operations data through a centralized database and communications infrastructure.

The data archiving component of the WisTransPortal consists of a collection of automated services that connect to various WisDOT and other ITS data sources and prepare the data for archiving in a common relational database. The current development phase of WisTransPortal is focused primarily on replicating and integrating five existing WisDOT and regional data sources: traffic detector data, lane and ramp closure data, traffic incident data, historical crash reports, and road weather information.

The design of the WisTransPortal system is based on a distributed model that separates logical software components such as Web presentation, application logic, and data management components. The benefit of this design is in the ability to support the development of a wide range of user applications and to allow, at the same time, for gradual improvements to the underlying database model. In addition to the core archiving and data clearinghouse functions of the WisTransPortal, the system is designed to provide a platform for the development of high-level software modules to support existing and new WisDOT transportation operations applications.

In this paper, we describe the ITS architecture and technical organization of the WisTransPortal. The key components of the system include a direct link from the WisTransPortal system to the WisDOT fiber optic network, data management in Oracle Database 10g, and web-application development based on Java Enterprise (J2EE) technologies. An important technology direction of the project is the use of Open Source Java and J2EE frameworks such as Struts and Hibernate to facilitate robust code development, multi-layered design, and portability.

1. Introduction

The WisTransPortal Project was started in June 2003 at the UW-Madison Traffic Operations and Safety (TOPS) Laboratory to support the Wisconsin DOT (WisDOT) in the following areas: intelligent transportation systems (ITS) data archiving, real-time traffic information data sharing, transportation operations applications development, and transportation research. The Phase I Project, which was completed in December 2005, included developing a business plan and technical design for the WisTransPortal system, deploying the essential hardware and software infrastructure, and developing initial data archiving capabilities. This paper describes the ITS project architecture and the hardware and software design of the Phase I system. Future directions for the Phase II project, which starts in 2006, are also described. Additional information about the WisTransPortal can be found at the project website: http://transportal.cee.wisc.edu/.

2. WisTransPortal ITS Project Architecture

The National ITS Architecture is a common framework developed by U.S. Department of Transportation for planning, defining, and integrating intelligent transportation systems. Subsystems in the Physical Architecture view of the National ITS Architecture represent the principle structural elements of an intelligent transportation system. Terminators represent the people, systems, and general environment that interface to ITS. The scope of the WisTransPortal project involves the following six subsystems and two terminators of the National ITS architecture:

Archived Data Management Subsystem: This subsystem refers to the WisTransPortal itself, which provides data archiving, data integration, and data clearinghouse capabilities for a broad range of ITS subsystems in the WisDOT transportation system.

Traffic Management Subsystem: WisDOT maintains a statewide Traffic Operations Center in Milwaukee and additional ITS operation centers in Madison and Wausau. These centers provide continuous traffic detector data, freeway video camera feeds, lane and ramp closure reports, and selected county CAD / 911 traffic incident data.

Emergency Management Subsystem: The WisDOT Division of State Patrol (DSP) is developing a continuous XML feed to provide real-time state patrol traffic incident data over the WisDOT network.

Maintenance and Construction Management Subsystem: Wisconsin is divided into five transportation regions. The regional offices are primarily end-users of the WisTransPortal system. By using the system, maintenance and construction information from these regions can be better managed.

Information Service Provider Subsystem: WisDOT maintains 59 road-weather (R/WIS) collection stations throughout Wisconsin. Continuous data collection is operated through an outside information service provider and made available to WisDOT over FTP and the Web.

Personal Information Access Subsystem: The WisTransPortal is being developed to provide data clearinghouse and traveler information tools to the general public, as well as to engineers and planners inside WisDOT. The WisTransPortal is also geared to provide data integration services for other automated systems, such as Wisconsin 511.

Map Update Provider Terminator: The WisDOT Bureau of Information Technology Services provides the state trunk numbered (STN) highway and Wisconsin local roads (WisLR) GIS base maps and physical roadway characteristics information.

DMV Terminator: The WisDOT Division of Motor Vehicles maintains a historical crash database for Wisconsin which is available annually as a database extract.

3. WisTransPortal Hardware Organization

The WisTransPortal is physically located at the TOPS Laboratory in the University of Wisconsin-Madison Department of Civil and Environmental Engineering. The main hardware components include a database server and external storage device, application servers, networking and fiber optic equipment, and additional hardware to support system and data availability. A diagram of the hardware organization of the system is shown in Figure 2. An overview of the main system components follows:

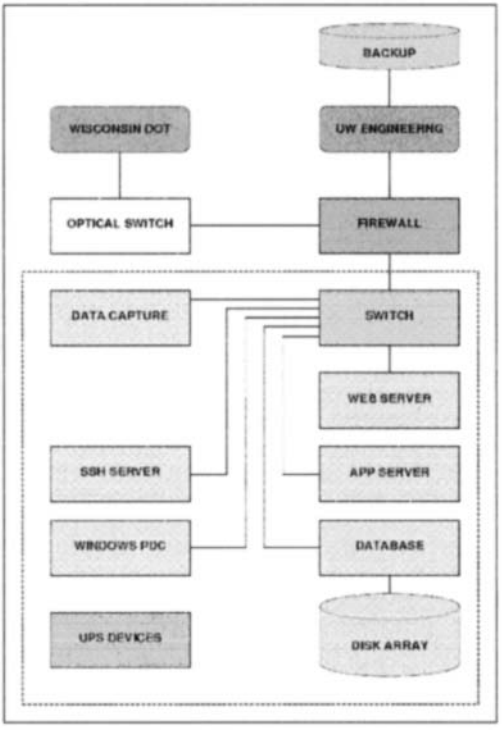


Figure 2. Hardware Diagram of the WisTransPortal System.

Database Server: The WisTransPortal database hardware consists of a dual processor Dell PowerEdge 2850 server with an external Dell PowerVault 220s SCSI disk array. The server and disk array are both configured with multiple hardware RAID volumes: one large RAID

1+0 for the database files and several smaller RAID 1 volumes for the OS, database binaries and redo logs. The database software consists of Oracle Database 10g Enterprise Edition, which includes enhanced support over previous editions for very large database objects. In particular, the WisTransPortal currently makes extensive use of Oracle partitioned tables and indexes, which allows one to transparently decompose data into smaller pieces to improve query performance and availability.

Application Servers: The WisTransPortal system consists of multiple Dell PowerEdge 1850 servers to distribute higher level application logic. All servers, including the database server, run the Windows Server 2003 operating system with Active Directory. The application servers are configured to decouple logical software components such as web services, data processing, and data acquisition. This software architecture is discussed in more detail in the next section. The WisTransPortal also includes separate application servers for remote access over SSH / SFTP and for the Windows domain controllers. The various servers are connected by a high-speed Cisco 2950 network switch.

Networking: The UW-Madison College of Engineering provides the basic network infrastructure for the WisTransPortal system. This includes DHCP and DNS services, as well as a college-wide firewall. In addition, the WisTransPortal has its own hardware firewall. An important component of the WisTransPortal system is the deployment of a direct fiber connection to the WisDOT fiber optic network. The fiber network configuration, which is still being deployed, is designed to establish a secure separation of the College of Engineering and WisDOT networks through the combined use of multi-port capabilities of the WisTransPortal firewall and IP routing techniques.

System Availability: An important goal of the WisTransPortal design is to maximize system and data availability through the use of redundancy. This includes RAID disk configurations, as noted above, dual power supplies and network interface cards in all servers, and dual UPS units. In addition, the system software design is geared to use redundant application servers to support process failover. Current development is focused on implementing the data archiving programs to operate simultaneously on separate (primary and secondary) application servers. Finally, system and database backups are handled through the UW-Madison backup system, which provides combined on-line recovery and off-site backup. As in any IT enterprise, there is a trade-off between reducing potential points of system failure by building redundancy into the architecture and overall deployment cost.

4. WisTransPortal Software Organization

The WisTransPortal software design incorporates two basic functions: ITS data archiving and ITS data dissemination. The data archiving category consists of a collection of automated Java programs that regularly update the WisTransPortal database with ITS data from WisDOT and other sources. The data dissemination category consists of a web-application platform based on Java Enterprise (J2EE) servlet and JSP technologies and Open Source Java web-application frameworks such as Apache Struts and Hibernate. A diagram of the software organization of the WisTransPortal is shown in Figure 3. An overview of the main software components follows:

Presentation Layer: The presentation layer encapsulates the high level web-application logic responsible for providing formatted output to users, such as HTML pages, map-based presentations, and XML or delimited data files. The system platform for the presentation layer consists of: Apache HTTP web server, Apache Tomcat servlet container, and ESRI

ArcIMS map server. Current development is geared towards web-based data clearinghouse tools using the open source Apache Struts web-application framework. Struts is a well-established framework for developing Java servlets and JSP pages based on the J2EE Model 2 / Model-View-Controller (MVC) design pattern. The presentation layer includes the View and Controller components of the Struts framework.

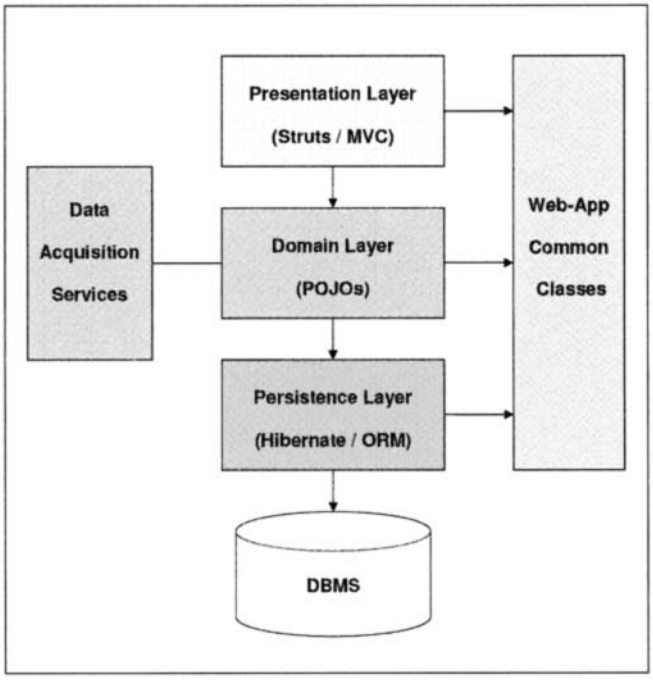


Figure 3. The WisTransPortal Software Organization.

Domain Layer: The domain layer encapsulates the data model and core application logic of the WisTransPortal system, including the Model components of the Struts MVC framework. A key role of the domain layer is to provide a representation of the WisTransPortal database to the presentation layer via standardized application programming interfaces (APIs). It also handles various internal data management activities. The application layer is being developed with ordinary Java classes, commonly referred to as Plain Old Java Objects (POJOs).

Persistence Layer: The persistence layer provides an interface between the application code and the low-level database management system. Its purpose is to manage the complexity that arises at the interface between the object oriented programming and relational database paradigms. It also provides a layer of independence between the application code and database model. The open source Hibernate object-relational mapping framework is currently being investigated for use in the WisTransPortal. As described above, Oracle Database 10g Enterprise Edition serves as the underlying database management system.

Web-Application Common Classes: Every web-application includes system-wide utility classes that do not strictly belong to any layer. The WisTransPortal system design follows an accepted practice of organizing these classes into one broad layer to provide consistent functionality across the remaining layers.

Data Acquisition Services: The data acquisition layer encapsulates the automated data acquisition logic and provides an interface to a temporary database / cache of real-time ITS data. As noted above, current development is based on a collection of automated Java programs, each running as a separate Windows service. An important design goal of the data archiving category is to separate the external data acquisition processing from the internal data model. In this way, the data archiving and web-application categories are organized around a common data model and development framework.

The WisTransPortal software design decomposes the data archiving and web-application logic into distinct modules and interfaces. Although this design does not place any specific requirements on the underlying hardware organization, the WisTransPortal software organization closely follows the hardware organization described in the previous section. Each software module is delegated to a separate server, with the exception of the archiving processes which are run on separate (primary and secondary) servers for the sake of redundancy. This provides a scalable design and has benefits for many system administration tasks. However, there is a trade-off in terms of an increased deployment cost and communication time between software components.

5. Future Work

The WisTransPortal Phase II Project, which starts in mid-2006, represents the next round of system development. The Phase II project goal is to complete the hardware build-out with respect to the WisDOT fiber connection and to further develop the database and software components of the system. Specifically, the Phase II project has the following objectives: enhanced support for data archiving capabilities, support for real-time data sharing capabilities, support for web-applications development (online query tools and map presentations), and support for ITS data standards initiatives.

6. References

The Apache Struts Project (2005). *Action Framework User Guide. Version 1.3.0-dev.* http://struts.apache.org/.

ASTM (American Society for Testing and Materials) International (2003). *ASTM Standard E2259-03a: Standard Guide for Archiving and Retrieving Intelligent Transportation Systems Generated Data.* West Conshohocken, PA.

The Hibernate Project (2005). *Hibernate Reference Documentation. Version: 3.1* http://www.hibernate.org/.

Singh, Stearns, Johnson, et al. (2002). *Designing Enterprise Applications with the J2EE Platform, Second Edition.* Sun Microsystems, Inc. http://java.sun.com/blueprints/guidelines/.

U.S. Department of Transportation (2005). *National ITS Architecture Version 5.1.* http://itsarch.iteris.com/itsarch/.

Improving Actuated Traffic Signal Control Operations Using Concept of Dynamic Gap-Out

Byungkyu "Brian" Park[1], Curtis Myzie[2], and Seli James Agbolosu-Amison[3]

[1]Assistant Professor, Department of Civil Engineering, University of Virginia, E-mail: bpark@virginia.edu, Phone: 434-924-6347, Fax: 434-982-2591
[2]Undergraduate Research Assistant, University of Virginia, E-mail: cdm5z@virginia.edu
[3]Graduate Research Assistant, University of Virginia, Email: sja3z@virginia.edu

Abstract

Traffic signal timing optimization and control is one of the most cost-effective ways of improving urban arterial network congestions. Actuated traffic signal control system is widely deployed around the world to provide green times where needed. The actuated control uses a pre-specified gap-out time to determine an early termination of current phase green time. However, the existing traffic signal control system relies on fixed point detectors that are installed on near stop bar and/or a few hundred feet upstream of the stop bar such that the system cannot capture entire vehicular movements along the approach link. This paper presents a dynamic gap-out concept and evaluates it using a microscopic simulation model with a COM interface. The results based on microscopic simulation runs under two one-way streets with 50 volume scenarios indicate that the dynamic gap-out can reduce vehicular delays by 20% when compared to existing regular gap-out.

Introduction

Traffic signal timing optimization is considered one of the most cost-effective ways of improving urban arterial network congestions. As such many researchers have focused on the development and enhancement of signal timing control and optimization practices. Several analytical computer-based programs were developed to generate better signal timing plans, including TRANSYT-7F (Wallace et al. 1991), SYNCHRO (Trafficware 2001), PASSER-II (Messer et al., 1974), etc.

Besides these programs, it was vehicle detection technology that drove the development of traffic signal control systems. In early days, when vehicle detectors were not available, pre-timed traffic signal control was implemented on the basis of manually collected traffic counts. Advances in the electronic technology enabled the use of loop-based automatic vehicle detection system. With the ability to detect vehicle existence on a certain approach, actuated traffic signal control systems were deployed. The idea of actuated control was to provide green where needed. However, the actuated control does not have capability of predicting system conditions or flexibility in traffic signal timing plan updates. It only works with predefined time-of-day timing plans. Naturally, the concept of adaptive signal control (ASC) emerged in 1980s, which can update signal timing plans in real-time based on the system conditions. The ASC systems, however, utilize rather extensive coverage of the urban network by placing more vehicle detection systems. Even though the concept of ASC has been very attractive, several field implementation results indicate that the benefits are still skeptical. Among these, actuated traffic signal control is most widely deployed in real world. One of the advanced features available in the actuated control over pre-timed control is an

ability to terminate current phase green time and to provide extra green time where it is needed. This is called gap-out and is intended to serve next phase where cars are waiting provided that no more cars are arriving on the current approach within the predetermined gap-out time.

With the emerging wireless location technology (WLT) including GPS-equipped vehicles, cellular phone tracking, etc., individual vehicle locations can be obtained in real time without placing fixed sensors on the roads. It is not surprising that vehicle infrastructure integration (VII), in which integrates WLT with infrastructure, is one of nine-tiers that the Federal Highway Administration (FHWA) decided to focus on for improving transportation systems. VII will enable communications among individual vehicles as well as with infrastructure management systems. This would require a new paradigm of traffic signal control algorithms as existing algorithms could not effectively take advantage of these abundant amounts of traffic data. One issue to be considered is market penetration of VII. Even though VII is becoming widely available with GPS-enabled navigation systems or OnStar system, it is not likely to be deployed to 100% of vehicles in a near future. However, it would be very important to provide potential benefits of utilizing VII technology in the area of traffic signal control systems.

This paper presents an estimate of benefits of utilizing VII technology in the area of actuated traffic signal control system, especially in the use of gap-out feature. For actuated signal control system, gap-out provides a flexibility in the allocation of green times where needed. However, each gap-out requires a certain amount of time (i.e., gap-out time or extension time) to make sure no more cars are arriving on that approach. One could ask what if individual vehicular information becomes available. Thus, a dynamic gap-out concept was proposed and evaluated using a microscopic simulation model, VISSIM. In order to implement the proposed dynamic gap-out, it was assumed individual vehicular status (e.g., location, speed, and acceleration rate) can be sent to traffic signal controller via either VII implementation or area wide video image detection system.

Proposed Dynamic Gap-out Logic

The operation of a dynamic gap-out signal controller is quite similar to that of an existing regular gap-out under actuated controller. Both systems check to see if there is demand on an approach and use a gap-out time to determine when a signal should change its phase. The difference is that the dynamic gap-out system is able to evaluate vehicles at a greater distance from the stop bar - the controller knows if a vehicle will reach the stop bar within the gap-out time. With this additional information, the traffic signal controller is able to make better signal state decisions. For example, take an intersection of two one-way streets, A and B. The signal on A is green and the signal on B is red. A vehicle passes through the traffic signal on street A while a vehicle is waiting for the traffic signal to change green on B. Rather than wait for the duration of the gap-out time before changing the traffic signal green for B, the dynamic gap-out controller changes immediately if there will be no cars arriving on A within the gap-out time. Consequently, this dynamic gap-out controller saves the car on B a delay of the gap-out time. With a regular gap-out controller, the vehicle would experience this delay.

Evaluations of Dynamic Gap-out Using Simulation

In order to demonstrate the benefits of the proposed dynamic gap-out logic under actuated signal control system, this study used a simple intersection with two one-way streets.

For this network, all approaches to the intersection were single lanes and no turning movements were considered. The network was designed to allow only vehicular movements in the NB and EB directions. This setup was chosen to avoid issues related to lane changing behaviors. Figure 1 shows a hypothetical layout of the network setup in SYHNCRO.

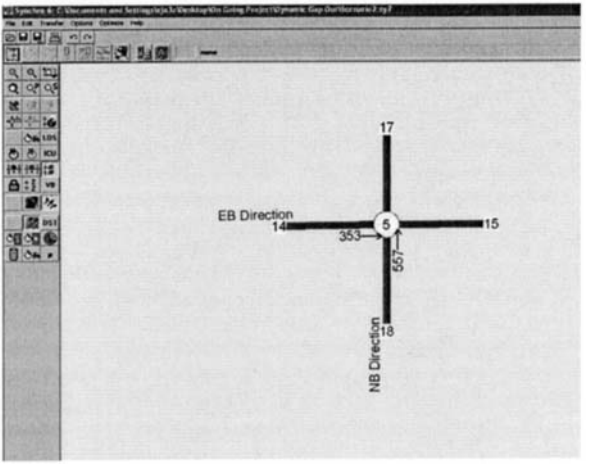


Figure 1. Two one-way Streets Modeled in SYNCHRO

Development of Traffic Volume Scenarios

To evaluate the concept of dynamic gap-out, various traffic volume scenarios covering different traffic conditions were developed. The process involved, designing lower and upper traffic demand volume conditions for NB and EB directions. The lower and upper demand volumes were designed with a criteria that their volume levels cover volume to capacity (v/c) ratios between 0.1 and 1.0. A total of 50 volume scenarios were developed for both NB and EB directions with the aid of statistical experimental technique, Latin Hypercube Design (McKay et al., 1979), and these volume scenarios were then used to analyze potential benefits of dynamic gap-out.

Traffic Signal Timing Optimization

SYNCHRO, a macroscopic simulation and optimization model for signalized intersections, was chosen to optimize 50 volume scenarios. This is to ensure that both dynamic gap-out and regular gap-out are evaluated under optimal timing plans. The optimization process in SYNCHRO involved optimizing intersection cycle length, and then followed by optimizing intersection splits. It is noted that during the optimization, all timing plans were optimized under regular gap-out control. The reason is that currently no tools are available in traffic simulation and optimization models for optimizing timing plans under dynamic gap-out conditions. These timing plans developed were then transferred to VISSIM, a microscopic simulation model for the evaluation of dynamic gap-out.

Implementation in VISSIM

As it appears that none of existing microscopic simulation programs provides dynamic gap-out logic, an application programming interface (API) for dynamic gap-out logic was developed in this study. The proposed gap-out simulation was done using a commercially available microscopic simulation program, VISSIM. It features interfaces using Microsoft's COM and plug-ins, allowing programmers to add functionality, automate simulations, and access VISSIM's internal data. Some of the functionality needed to implement a dynamic gap-out signal controller was not available through the COM interface, unfortunately. Consequently, the simulation setup became more complex and was divided into two parts: a Java program that runs a series of simulations and collects resulting performance measures (i.e., delay times), the other part a .DLL file plug-in that contains all the signal control logic.

The Java program was developed with assistance from Jawin (http://jawinproject.sourceforge.net/jawin.html), a software tool that generates a bridge between Java applications and COM applications, such as VISSIM. The program connects to an instance of VISSIM through COM and then runs a series of VISSIM simulations using varying parameters. Some of the parameters include the flow rates for each of the vehicle entry points, the gap-out time, and vehicle detector range. After each simulation run with a specific set of parameters, VISSIM outputs the total delay for all vehicles in the network during the length of the simulation. The program outputs all these delays after all the simulations have finished.

The signal control plug-in was written in C++ and contains the logic for both the dynamic gap-out and regular gap-out signal controls. The logic was kept simple because the test network used in this study has only a single intersection with two one-way streets. With only two signal phases, all the possible states can be enumerated and actions can be taken based on the current state and vehicle demands. One possible state that is enumerated is the case in which one of the signal phases has been green less than the minimum green time, in which case no change to the signal states is ever made.

The regular gap-out uses VISSIM's vehicle detectors to affect the signals in the standard actuated signal control manner. It queries the vehicle detector on each approach to determine the demands. If a signal is green and no vehicle has been seen in the last gap-out time period and there is demand on the red signal, then switch the signals. The dynamic gap-out operates in a similar fashion. The difference is that demands are calculated differently and the current signal phase will gap-out immediately if a vehicle is not going to arrive within the gap-out time. There is a demand on a signal if a vehicle will reach that signal within the gap-out time or if a vehicle is within a certain distance of the stop bar.

The important state for both regular gap-out and dynamic gap-out is when both phases do not *need* to change, but *can* change if there is a demand on the phase that is currently red and not on the phase that is green. In regular gap-out, this case is handled as follows: if the time since the last car on the green approach is greater than the gap-out time, switch the signals. Under the proposed dynamic gap-out, if another vehicle is not going to arrive on the green approach within a length of time as long as the gap-out time, then go ahead and switch the phases.

Results and Discussion

In order to consider variability in microscopic simulation models, four replications were made for each of 50 volume scenarios. Total vehicle delays under regular gap-out and dynamic gap-out were plotted and shown in Figure 2. It can clearly be seen that the total vehicle delays are mostly concentrated below the 45-degree line indicating dynamic gap-out produces less delays than that of regular gap-out.

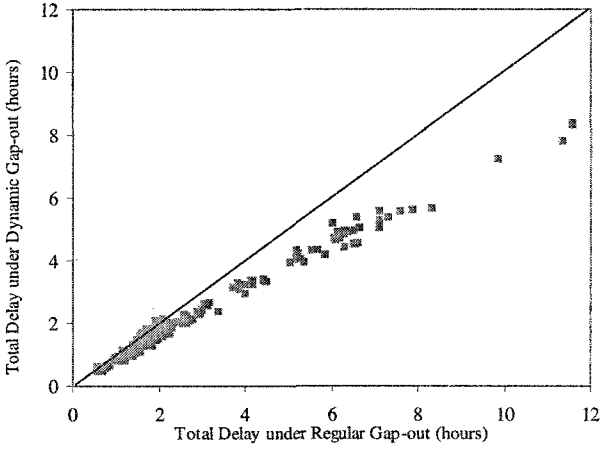


Figure 2. X-Y Plot of Total Vehicle Delays under Dynamic Gap-out and Regular Gap-out

Further, a summary of total vehicle delays for 50 volume scenarios is presented in Table 1. The total vehicle delay measures (i.e., average and standard deviation) were computed for all 50 volume scenarios. As shown in Table 1, implementation of dynamic gap-out could reduce about 20% of total vehicle delay over regular gap-out. Obviously, high delays were seen under regular gap-out because the traffic signal controller has to wait for predetermined gap-out time to elapse before changing signal phases.

Table 1. Summary of Total Vehicle Delays for 50 Volume Scenarios

	Total Vehicle Delays (hours)	
	Regular Gap-out (RG)	Dynamic Gap-out (RG)
Average total delays	2.79	2.21
Standard deviation	0.018	0.027
Percent reduction (RG – DG)/RG	20.64%	

Conclusions and Recommendations

This paper proposed a dynamic gap-out logic under actuated signal control system based on the assumption that individual vehicular information (i.e., location, speed, acceleration) can be obtained from either VII or advanced area wide video detection system.

In order to evaluate the benefits of implementing dynamic gap-out under microscopic simulation model, an API program using Java and C++ was developed. Using an isolated intersection with two one-way streets and 50 volume scenarios, a multiple of VISSIM simulation runs under both dynamic gap-out and regular gap-out were made. The results indicated significant benefits can be achieved by utilizing dynamic gap-out over existing regular gap-out under actuated control. It is noted that the actual benefits could be higher as the timing plans used in dynamic gap-out scenarios were optimized under regular gap-out logic.

In order to deploy the proposed dynamic gap-out logic in real world, the following is recommended for future research.

- Validation of area-wide video detection system: This is necessary until VII market penetration reaches 100%. Even with 100% market penetration, this would be useful for complementing any potential errors with VII information.
- Optimization under dynamic gap-out: Since existing optimization programs do not support traffic signal timing optimization under dynamic gap-out.

Acknowledgements

This research is sponsored in part by the National Science Foundation Grant (Award No. CMS-0510404) to the University of Virginia.

References

McKay, M. D., R. J. Beckman, and W. J. Conover (1979), A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, Vol. 21, No. 2, pp. 239-245.

Messer, C. J., H. E. Haenel, and E. A. Koeppe (1974), *A Report on the User's Manual for Progression Analysis and Signal System Evaluation Routine – PASSER II*, TTI Research Report 165-14, College Station, Texas.

Trafficware (2001), SYNCHRO User's Manual.

VISSIM Version Manual (2004), PTV Planug Transport Verkehr AG. Innovative Transportation Concepts, Inc.

Wallace, C. E., K. G. Courage, M. A. Hadi, and A. C. Gan (1998), *TRANSYT-7F User's Guide*. Transportation Research Center, University of Florida, Gainesville, Florida.

A Single Loop Detector Diagnostic: Mode On-Time Test

Benjamin Coifman[1] and Ho Lee[2]

[1]Associate Professor, Dept. of Civil and Environmental Engineering and Geodetic Science and Dept. of Electrical and Computer Engineering, Ohio State Univ., 470 Hitchcock Hall, 2070 Neil Ave., Columbus, OH 43210; E-mail: Coifman.1@osu.edu

[2]Graduate Research Assistant, Dept. of Civil and Environmental Engineering and Geodetic Science, Ohio State Univ., Columbus, OH 43210; E-mail: lee.2406@osu.edu

Abstract

Loop detectors are the most common vehicle detector for freeway traffic surveillance. Little is done in conventional practice to ensure that the detectors are functioning properly. When validation tests are applied the most common are to check that the detector seems to be counting the correct number of vehicles or that average speeds are reasonable. But such tests rely on aggregate measures across many vehicles and cannot catch many errors in occupancy, a common input to ramp metering. Our group has already shown the benefits of examining individual vehicles at dual loop detectors, this paper extends that strategy to single loops. In particular, we examine the mode on-time, where the on-time is the time one vehicle is over the detector. The strategy exploits the fact that at most locations, most vehicles will be passenger vehicles of known length range, and similarly, that most vehicles should be traveling at free flow speeds. Since on-time is the ratio of a vehicle's length over speed, the mode value at the end of a day should fall within a small range.

Introduction

Loop detectors are the most common vehicle detector for freeway traffic surveillance. It is important to identify non-performing and inaccurate detectors, as any errors made by the detectors will propagate to all control decisions based on the detector's data. The most common test is to simply check that the detector seems to be counting the correct number of vehicles or that average speeds are reasonable. But such tests rely on aggregate measures across many vehicles and cannot catch many errors in occupancy, which is often used as an input to ramp metering.

Many practitioners and researchers have worked to formalize heuristic tests to check average measurements within statistical tolerance (Jacobson et al. 1990; Cleghorn et al. 1991; Nihan 1997). These systems often go undocumented in the literature because they are either designed in house by an operating agency or consulting firm, e.g., see Chen and May(1987) for examples. Because these automated systems only use aggregated data, they must accept a large sample variance and potentially miss problems altogether.

Chen and May (1987) deviated from the aggregated measurement approach and used individual vehicle actuations to verify detector data. Their methodology examines the distribution of vehicles' on-time, that is, the time the detector is occupied by a vehicle. Unlike conventional aggregate measured, their approach is sensitive to errors such as "pulse breakups", where a single vehicle is detected multiple times. Coifman (1999) compared the measured on-times for each loop in a dual loop detector to identify errors. Because the two loops are closely spaced within a single lane in a dual loop, the on-times from the loops during free flow condition should be virtually identical regardless of vehicle length. At lower speeds vehicle acceleration can result in a difference of on-times and congested periods were excluded from the earlier analysis. Coifman and Dhoorjaty (2004) extended these efforts, developing eight detector tests to identify various errors both at single loop and dual loop detectors.

This paper shifts the focus exclusively to single loop detectors (though the tests could also be applied to loops in a dual loop or even other detector technologies). Detectors exhibit a wide range in sensitivity, as manifest in all on-times, either tending to be longer or shorter. If this bias is not measured, it will impact the accuracy and benefits of subsequent decisions for ramp metering and other freeway controls. Since on-time is the ratio of a vehicle's length over speed, the mode value at the end of a day should fall within a small range. Any exceptions would either indicate a detector bias or unusual traffic patterns, the latter can usually be quickly identified or eliminated by looking at results from multiple days. The results are compared to manual validation. Several supplementary tests are also developed to augment the mode on-time, namely; under minimum on-time, and over maximum on-time.

Data

This work uses individual vehicle actuations, i.e., event data, sampled at 60 Hz from a dual loop detector station in the Berkeley Highway Laboratory along I-80, north of Oakland, California (Coifman et al, 2000). Four different sensor units were used to collect data for this study, i.e., Peek, Reno, 3M and IST (for details, see Coifman, 2006). Each loop sensor was deployed for at least 24 hr across the same dual loop detectors in each lane at the station, covering five lanes in each direction. The redundancy of the dual loop detectors can be used for additional validation of single loop detector performance.

A loop detector records two transitions for each passing vehicle, *turning-on* as the vehicle enters the detection zone and *turning-off* as it leaves. The duration which a vehicle occupies a loop detector is the on-time. The difference of turn-off time from the preceding vehicle and current vehicle's turn-on time is the off-time.

Ultimately, any operational error at the detector can be classified into one of seven broad symptoms, as follows. Two errors will extend on-time measurement: premature turn-on and delayed turn-off. Three errors will shorten on-time measurement: delayed turn-on, premature turn-off, and flicker (turning off and back on in the middle of a vehicle). A missed vehicle will result in no on-time measurement when one should have been recorded, and a detection in the absence of a vehicle will lead to the opposite problem. Errors in on-time measurement will impact any subsequent occupancy measurement, as well as speed or length estimation from single loops. Note that two lanes do not have data for the 3M sensor due to the sensors locking up in these lanes and not providing any data.

Detector tests

Mode of on-times

The mode of on-times is simply the most commonly observed value of on-time. As mentioned in previously, the on-time measurement depends both on vehicle length and on speed, both of which vary from vehicle to vehicle and often as a function of time of day. However, during most of the day at most locations, most vehicles are expected to be passenger vehicle of know length range. Using the dual loop detectors to measure effective vehicle length, Fig 1A shows the cumulative distribution over 24 hrs, with approximately 90% of the measurements falling within the 18 to 22 ft range, i.e., passenger vehicles. Similarly, over any given 24 hr period, the majority of the time will usually be characterized by free flow conditions. As a direct result, the mode on-time should usually fall within a small range that corresponds to a passenger vehicle at free flow speed. The on-time from a 20 ft vehicle traveling at 65 mph is 13/60 sec. Fig 1B shows frequency of on-time over 24hrs at one loop detector on lane 2. Although this lane saw many hours of congestion during the day (and thus, larger on-times during these periods) the mode on-time corresponds to 13/60 sec.

Since on-time is inversely proportional to speed for a given vehicle length and traffic in the outer lanes moves at slower speed than the inner lanes, the mode on-time is expected to increase slightly as one moves to the outer lanes. Assuming that most vehicles indeed fall in

the 18 to 22 ft range and allowing a large range of acceptable free flow speeds (40 mph to 82 mph), the mode on-time should fall within 11/60 to 16/60 sec. If the mode on-time over 24 hr from a detector is outside the expected range, it is indicative of a detector error or transient event. The latter can be addressed by looking at the results from several days or avoiding results on days with known incidents, e.g., snow. If the location is known to have many hours of recurring congestion, the test can be modified to exclude congested traffic. Detector errors, on the other hand, may be as simple as the sensitivity of the detector differs from what is otherwise expected to chronic errors in the loop sensor.

The frequency plot from Fig 1B is repeated for each lane, both upstream and downstream loops, for each of the four loop sensor units. The resulting mode on-times are shown in Fig 2 As expected, the mode on-time generally increases as one moves to the outer lane (lane numbers increase as one moves outward). Because the loops are in good working order at this

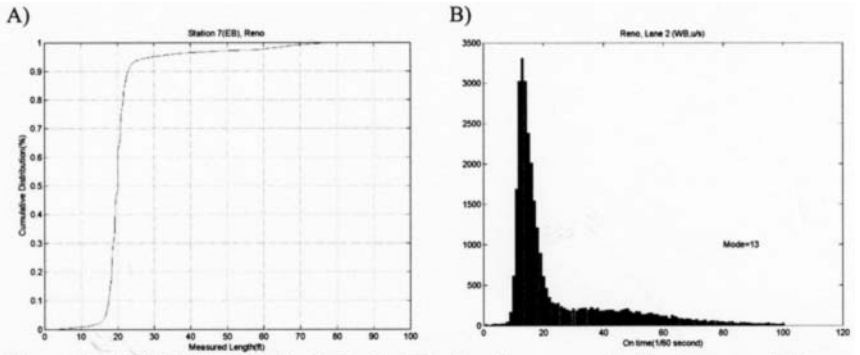


Figure 1, (A) Distribution of individual vehicle lengths measured in five lanes over 24 hr period, (B) distribution of on-times over 24 hr at one loop detector.

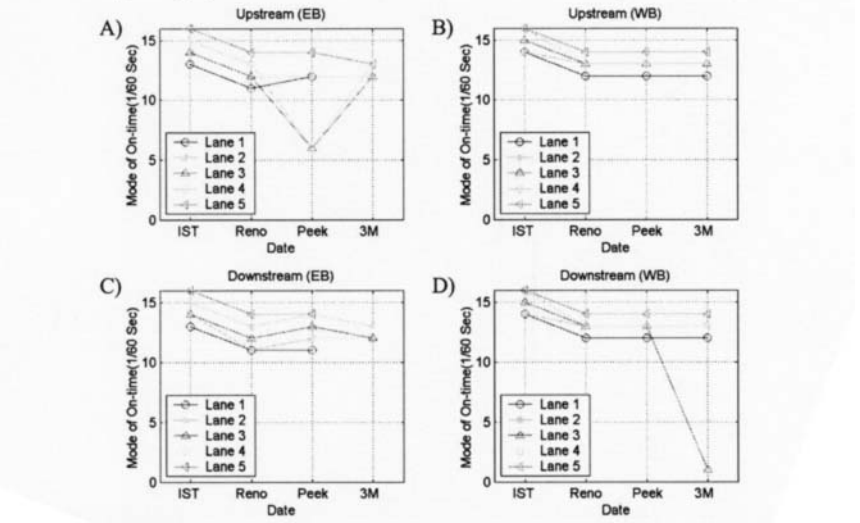


Figure 2, Mode of on-time over 24 hr at each loop detector, (A) & (B) upstream loops in given lane, (C) & (D) downstream, (A) & (C) eastbound, (B) & (D) westbound.

station, almost all of the sensor units yield mode on-times within the expected range. Notable exceptions are lane 3 and 4 for Peek in Fig 2A and lane 3 for 3M in Fig 2D.

Manual Validation

The results from this mode on-time test are compared to manual validation at the downstream loop in lanes 1-4 eastbound. Coifman (2006) looked at the performance of the four sensor units using concurrent video recorded at this station and 30 min of data were manually reduced for each sensor across the four lanes using Videosync (Caltrans, 2005). The time periods were selected to ensure that 15 min would be from free flow and 15 min from congestion. In almost all combinations of sensor and observable lane, fewer than 2% of the on-time measurements during free flow had any errors. The only notable exception is Reno lane 2, which had an error rate around 10% and these were primarily flicker and short transient non-vehicle actuations. This error is barely evident in Fig 2C, and it is not likely that one could detect its presence from the mode on-time alone. In this case the low on-times are not frequent enough to significantly impact the mode on-time and the mode on-time remains within the expected range. The error is, however, evident in the on-time distributions, Fig 3 shows the frequency plot of on-times in lanes 2 and 3 from the Reno data. The extra peak near zero in both lanes reflects the flicker and short on-times found in the manual validation. To catch the extreme errors due to flicker and locking up, the rest of paper presents minimum on-time and maximum on-time test.

Under Minimum On-time test

The minimum on-time test catches short on-times due to flicker or truncated on-time arising from detector errors. Since on-time is proportional to vehicle length and inversely proportional to speed, the minimum on-time corresponds to the shortest possible vehicle at the maximum feasible speed. This study uses a 20 ft vehicle at a maximum speed of 85 mph, which translated into an on-time of 10/60 sec. Fig 4 shows the percentage of vehicle with on-times less than 10/60 sec for each loop detector over 24 hr (note that the scale of vertical axis is different in each of the plots). Most of the loop detectors have fewer than 5% of the measurements below the threshold. Perhaps not surprising, loop detectors with low mode on-time generally have a large percentage of vehicles with on-times less than 10/60 second. Fig 4B-C show that Reno and 3M have a relatively high percentage if actuations under the min on-time test.

Over Maximum On-time test

The test catches the other extreme of on-times, i.e., large on-times due to the detector stick-

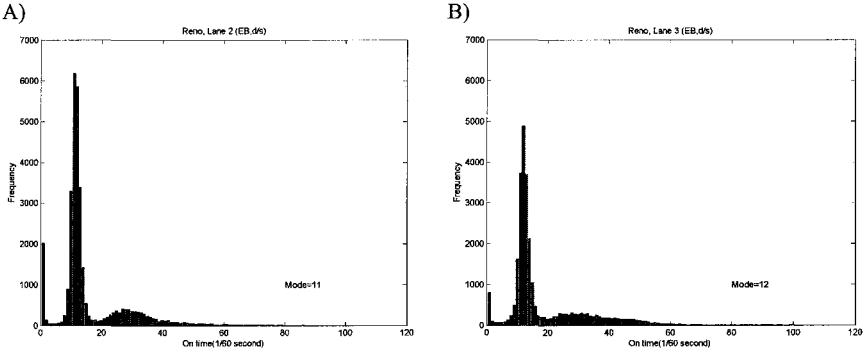


Figure 3, Frequency of on-times during 24 hr at two loop detectors, both indicating a high number of short on-times. similar to manual observations.

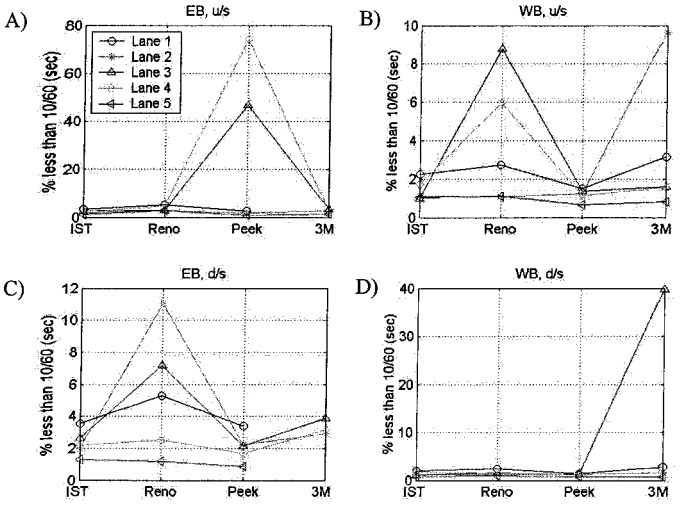


Figure 4, Percentage of vehicles with on-times under min on-time, (A) & (B) upstream loops, (C) & (D) downstream, (A) & (C) eastbound, (B) & (D) westbound.

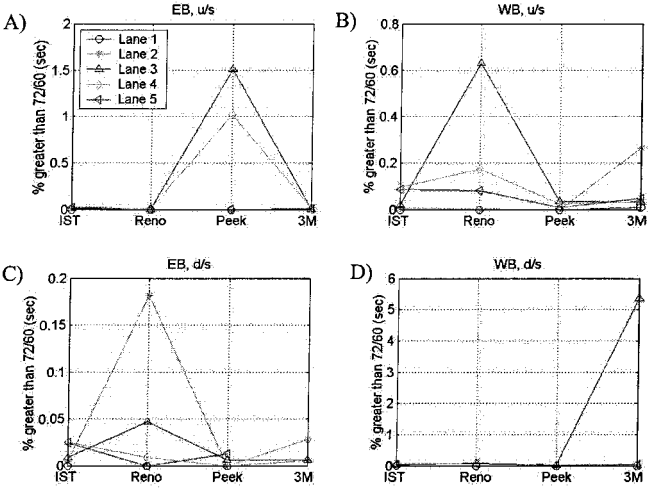


Figure 5, Percentage of on-times greater than max on-time, (A) & (B) upstream loops, (C) & (D) downstream, (A) & (C) eastbound, (B) & (D) westbound.

ing-on. The test is restricted to free flow conditions since during congestion the on-times can be a large simply because the vehicles are moving slowly over the detector. If congested conditions are defined as speed below 45 mph, the maximum feasible on-time during free flow is obtained from the longest feasible vehicle, 80 ft, traveling at the threshold speed, 45 mph, i.e., 72/60 sec. Most on-times during free flow should be less than this maximum on-time. If a large percentage of on-times exceed the maximum on-time it would indicate that the detector is sticking on. Fig 5 shows the percentage of vehicles with on-times greater than maximum on-times for each loop detector (again, the vertical scale changes from one plot to the next). Most of loop detectors are far below 1% in this test. Compared to Fig 2, the trend of the results are similar to those of the mode on-time test.

Conclusions

Vehicle detectors are not always accurate and may be completely non-operational, so it is necessary to identify non-performing and potentially inaccurate detectors. This paper presents the mode on-time test as one measure of performance of loop detectors. The test is applied individually to each loop in each lane and the mode on-time is found over 24 hr. Because most vehicles are passenger cars, most of the time the traffic is freely flowing, and the sample is large, this mode is expected to fall with a small range of on-times, i.e., *11/60 to 16/60 sec.* The mode on-time was compared to manual validation. Because the loops are in good working order at this station almost all of the sensor units yielded mode on-times within the expected range. A few loops had problems that resulted in short on-times, but the errors were not frequent enough to impact the mode. So in addition to the mode on-time, minimum on-time and maximum on-time test were also applied to catch extreme errors due to flicker and detectors sticking-on. The minimum on-time test caught the detector errors due to flicker, not found by the mode on-time test. The maximum on-time test caught the same loops as the mode on-time test using a different criteria.

References

Caltrans (2005). "VideoSync Portable Roadway Detector Evaluation System" http://www.dot.ca.gov/research/operations/videosync (accessed on December 9, 2005).

Chen, L., and May, A. (1987). "Traffic detector errors and diagnostics" Transporation Research Record 1132, Transportation Research Board, Washington, D.C., 82-93.

Cleghorn, D., Hall, F., and Garbuio, D. (1991). "Improved data screening techniques for freeway traffic management systems." Transportation Research Record 1320, Transportation Research Board, Washington, D.C., 17-31.

Coifman, B. (1999). "Using dual loop speed traps to identify detector errors", Transportation Research Record 1683, Transportation Research Board, Washington, D.C., 47-58.

Coifman, B., Lyddy, D., Skabardonis, A. (2000). "The Berkeley Highway Laboratory-Building on the I-880 Field Experiment", *Proc. IEEE ITS Council Annual Meeting*, pp 5-10.

Coifman, B., and Dhoorjaty, S. (2004). "Event data-based traffic detector validation tests", ASCE Journal of Transportation Engineering, Vol 130, No 3, 2004, pp 313-312

Coifman, B. (2006). "Vehicle Level Evaluation of Loop Detectors and the Remote Traffic Microwave Sensor," *ASCE Journal of Transportation Engineering*. Vol 132, No 3, pp 213-226.

Jacobson, L., Nihan, N., and Bender, J. (1990). "Detecting erroneous loop detector data in a freeway traffic management system." Transportation Research Record 1287, Transportation Research Board, Washington, D.C., 151-166.

Neelisetty, S (2004), *Detector diagnostics, data cleaning and improved single loop velocity estimation from conventional loop detectors*, Thesis, Ohio State University.

Effect of Speed Feedback Device on Speeds in Interstate Highway Work Zones

M. V. Chitturi[1] and R. F. Benekohal[2]

[1] Doctoral Student, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 N.Mathews Ave, Urbana, IL 61801; PH (217) 333-8988; FAX (217) 333-1924; email: mchittur@uiuc.edu
[2] Professor, Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, 205 N.Mathews Ave, Urbana, IL 61801; PH (217) 244-6288; FAX (217) 333-1924; email: rbenekoh@uiuc.edu

Abstract
Providing speed feedback to motorists may improve their compliance with speed limits in work zones. Speeding has been identified as one of the major reasons for crashes in work zones and there is a need for effective speed reduction strategies in construction zones. Speed photo enforcement with speed feedback system is a promising technology for speed management in work zones. A component of the photo enforcement system is the speed feedback display device that has some effects in reducing the speeds of motorists in work zones. Data were collected on an interstate highway work zone in three phases and statistical analysis was performed to determine the temporal effect of the speed feedback device. It was found that feedback device was effective in reducing average speed by 4.4 mph in a few days after the deployment. After three weeks of operating the feedback system, the average speed was further reduced by 2.3 mph. Both these reductions were statistically significant. Thus, at this location, providing speed feedback reduced the speeds of the motorists not only immediately after the system was deployed, but also three weeks after deployment.

Introduction
Noncompliance of motorists with the speed limits in work zones could result in potentially hazardous situations for the motorists and workers. The number of persons killed in motor vehicle crashes in work zones nationwide, has risen from 872 in 1999 to 1028 in 2003 and more than 40,000 people are injured due to motor vehicle crashes in work zones (FHWA). Speeding and inattentive driving have been identified as the two major reasons for work zone crashes by a Wisconsin Department of Transportation study (Dixon and Wang, 2000).

Daniel et al performed speed limit compliance studies at highway construction work zones in the state of Georgia and reported that although the drivers reduce their speeds in the vicinity of active work zones, the speeds are still significantly higher than the posted speed limits. In a work zone on an interstate highway, Benekohal, Wang, Orloski and Kastel (9) found that 65% of autos and 47% of the trucks exceeded the speed limit of 45 mph at the work activity area. Benekohal, Orloski and Hashmi (1990) conducted a survey of drivers who had just driven through a work zone and reported that although 79% of the drivers said the speed limit was

reasonable only 59% of the drivers complied with it. Therefore effective speed reduction strategies need to be deployed in work zones.

Speed limit enforcement by police is effective in reducing the speeds of the motorists in work zones. Benekohal, Resende and Orloski (1992) studied the effects of police presence on vehicular speeds in a work zone on a rural interstate highway. They found that police presence was effective in reducing the speeds of all vehicles. However, this option is expensive and requires extensive manpower. Therefore, automated or photo enforcement technologies are increasingly being used by law enforcement agencies to reduce traffic violations, improve safety and make the enforcement effective, safer and less labor-intensive.

Although the final objective of this study is to evaluate the effectiveness of photo enforcement, a part of this study is to evaluate the effectiveness of radar-based speed feedback devices in work zones on interstate highways. This is a crucial step in order to evaluate the net benefits of photo enforcement when it is used in conjunction with speed feedback devices. The proposed photo enforcement in Illinois would have a speed feedback sign mounted on top or close to the van. So through this paper we have been able to quantify the approximate effect of the speed feedback component of photo enforcement. The reduction due to actual enforcement (issuing citations) will be in addition to this reduction. Some studies have been performed to determine the effectiveness of speed feedback signs in work zones in Georgia (Dixon and Wang) and Texas (Fontaine, Carlson and Hawkins). But these evaluations were done on state highways and U.S. routes, not on interstate highway work zones. Also no such studies have been reported in Illinois.

Data Collection and Reduction

The data were collected by Illinois Department of Transportation (IDOT 2004) in a work zone on Interstate Highway I-70 which is a 4 lane divided highway with 2 lanes in each direction. The passing lane in the west bound direction was closed using the standard lane closure format of Illinois Department of Transportation (IDOT). The display trailer was placed at a point 800 ft downstream from the first sign showing the reduced speed limit of 55 mph. This first sign was 4200 ft upstream of the beginning of the lane closure taper.

HI-STAR portable traffic analyzers that utilize vehicle magnetic imaging technology were used to record the vehicle speeds, vehicle type, ADT information, surface temperature and precipitation information. These devices were placed at a point 1900 ft downstream from the speed feedback device. This ensured that the drivers who chose to reduce their speeds after seeing the display device had ample opportunity to reduce their speeds. The speeds in the driving and the passing lanes were obtained separately in 30 min intervals. The vehicle types for the corresponding 30 min intervals were also obtained.

The data was collected in three phases. In all the three phases the location of the Hi-STAR devices remained the same.

- Phase 1: In this phase the speed feedback device was not deployed in the field. This data served as the base for evaluating the reduction due to the device. The data was collected for 33 hours and this yielded the speeds of 16,079 vehicles.
- Phase 2: The objective of this phase was to evaluate the initial impact of the speed feedback device. Data was collected for 31 hrs 30 minutes (17,501 vehicles).
- Phase 3: Data was collected three weeks after the deployment of the speed feedback device in the work zone. This was done to ascertain the long-term impact of the device. Data was collected for 58 hours (32,233 vehicles).

The data were obtained from the HI-STAR devices using the Highway Data Management (HDM) Software. It was found that there were instances of zero speed corresponding to zero volume and some unusually low speeds such as 20 mph were recorded. It was observed that such low speeds occurred in the passing lane (the lane that was closed downstream) in the morning hours around 6AM. It was surmised that these speeds are of the work zone workers' vehicles that were slowing down to park in the median. In order to avoid making any erroneous conclusions, such spurious data were eliminated from the datasets before performing any analysis.

Data Analysis and Findings

For the purpose of comparison, the speeds from one 24 hr period of phases 1 and 2 and two 24 hr periods of phase 3 were used. The average speed and the number of vehicles observed on these days are shown in Table 1. It can be observed that the volumes were not the same during the four days. It is common knowledge that volume of traffic affects the speed of the traffic. Therefore, the volumes of vehicles in each half-hour period of the four days were plotted against time and compared (shown in Figure 1).

Table 1. Average speed and volumes of vehicles in the 3 phases

Phase	Average Speed (mph)	No of vehicles
Phase 1 day 1	64.9	10,201
Phase 2 day 1	60.5	11,834
Phase 3 days 1 and 2	58.2	24,567

From Figure 1 it can be observed that the temporal variation of volumes in the four days is very similar. The maximum volume observed in any half-hour duration of all the four days is 461 vehicles or 922 vph. According to the Highway Capacity Manual (HCM 2000) the speed on a freeway remain relatively unchanged until the volumes reach 1300 vph. Also the maximum difference in the volume (in any half-hour) between the phases is 154 vehicles (or 308 vph). Therefore the variation in speed due

to the difference in volumes can be taken to be negligible. Between phases 1 and 2 the only known change was the introduction of the speed feedback device and between phases 2 and 3 any difference can be attributed to the long-term effect of the speed feedback device. Therefore any speed difference between the phases can be attributed to the immediate or long-term effect of the speed feedback device.

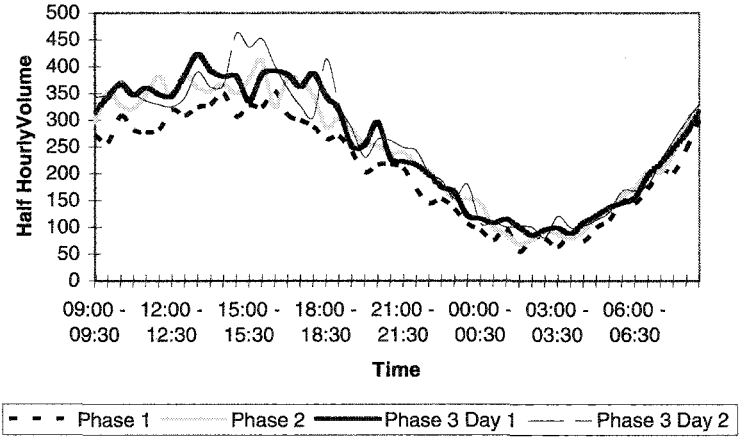


Figure 1. Variation of Volume with Time in Phases 1, 2 and 3.

From Table 1 it can be concluded that the speeds have reduced due to the speed feedback device. However in order to verify that the speed reductions are consistent throughout the day, the variation of the average speeds of the vehicles in each half-hour period of the four days are shown in Figure 2. It can be noted from Figure 2, that the speed in Phase 1 is consistently higher than the speeds in phase 2 for any period. Therefore there is a definite and consistent decrease in speeds of the vehicles due to the introduction of the speed feedback device. To determine the long-term effect of the device, speeds of phases 2 and 3 are compared. It should be noted that the speeds on both the days of phase 3 of the data collection are very similar and are mostly lower than the speeds of phase 2. In other words, there does not appear to be any diminishing effect of the device with time. Rather these data show that the speeds have decreased further after having the speed feedback device in the work zone for nearly a month.

In order to verify that the observed differences in speeds are statistically significant, paired t-tests were performed and 95% confidence intervals were used. Paired t-test was chosen because the data were not randomly selected. Rather they were grouped by the time of day. Table 2 shows the results of these statistical tests. The speeds of days 1 and 2 of the third phase were compared and it was found that statistically there is no difference between them. Therefore the average of these two days is used in the other comparisons. When phase 1 is compared to phases 2 or 3 the tests indicated that

the speeds in phase 1 were significantly greater than the speeds in phase 2 or 3. Therefore it can be concluded that the deployment of the speed feedback device was effective in reducing the speeds of the motorists in work zone not just immediately but even after a month of its deployment. Interestingly, it was found that the speeds of phase 3 were significantly lower than the speeds of phase 2. In other words the effectiveness of the display device in reducing the speeds increased with time. Caution should be used in extending this observation to other locations. We propose to collect more data to verify this finding.

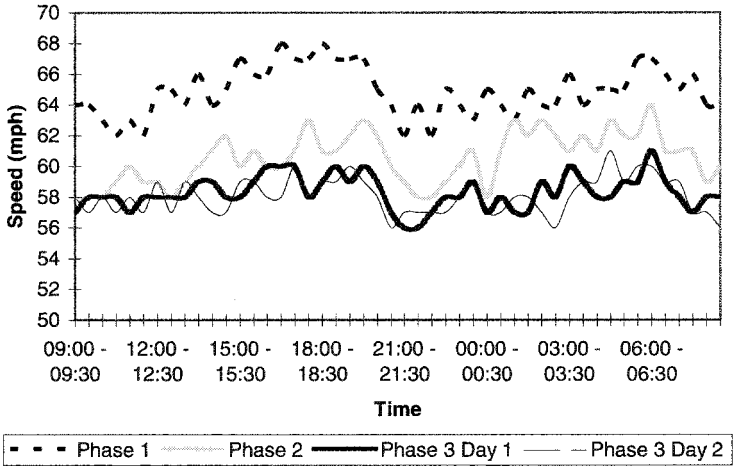


Figure 2. Variation of Speed with Time in Phases 1, 2 and 3.

Table 2. Results of Statistical tests

Test	Mean Difference	Standard Deviation of Difference	Degrees of freedom	t-value	Probability	Conclusion
Phase 3 Day 1 - Day2	0.25	1.14	47	1.5203	0.0676	Phase 3 Day 1 = Day 2
Phase 1 - Phase 2	4.38	1.77	47	17.1246	<0.0001	Phase 1> Phase 2
Phase 1 - Phase 3 Average	6.71	1.21	47	38.3815	<0.0001	Phase 1> Phase 3 Average
Phase 2 - Phase 3 Average	2.33	1.43	47	11.3309	<0.0001	Phase 2> Phase 3 Average

Conclusions

Providing speed feedback may help in increasing the speed limit compliance in work zones. Data were collected in an interstate highway work zone, in three phases to determine the impact of speed feedback device on speeds of motorists in work zones and its long-term impact. It was found that the average speed dropped by a significant 4.4 mph immediately after the introduction of the device and three weeks later the average speed decreased by a further 2.3 mph, which was also statistically significant. Therefore it can be concluded that in this one work zone, the speed feedback device was effective in reducing the speeds of the motorists and its impact does not dissipate with time. More data would be analyzed to validate these findings which are from one site.

References

Benekohal, R.F., Resende, P.T.V. AND Orloski, R.L. (1992). *Effects of police presence on speed in a highway work zone: circulating marked police car experiment.* Report UILU-ENG-92-2020. University of Illinois at Urbana-Champaign.

Benekohal, R F., Orloski, R. L., and Hashmi, A. M. (1990). *Survey of Drivers Opinion About Work Zone Traffic Control On a Rural Highway.* Interim Report FHWA-IL-UI-234. University of Illinois at Urbana-Champaign, U.S. Department of Transportation.

Benekohal, R.F., Wang, Orloski, R. L., Kastel. (1992). *Speed reduction profiles of vehicles in a highway construction zone.* Report UILU-ENG-92-2021. University of Illinois at Urbana-Champaign.

Daniel, J., Clements, M., and Dawkins, J. (1995). *Evaluation of Enhanced Traffic Control of Construction Zones – Final Report.* Georgia Department of Transportation Project 8904, Georgia Institute of Technology.

Dixon, K. K., Wang, C. (2002). *Development of Speed Reduction Strategies for Highway Work Zones.* Georgia Institute of Technology.

FHWA. Work Zone Safety Facts & Statistics. http://safety.fhwa.dot.gov/wz/wz_facts.htm

Fontaine, M.D., Carlson, P.J. and Hawkins, G.H. (2000). *Evaluation of traffic control devices for rural high-speed maintenance work zones: second year activities and final recommendations.* Texas Transportation Institute.

Highway capacity Manual 2000 (HCM2000). Transportation Research Board, National Research Council, Washington, D.C.

IDOT (Illinois Department of Transporation). (2004) *Work Zone Speed Study Using a Visual Speed Notification Device.* Springfield, IL.

Use of Hardware-in-the Loop Traffic Simulation in a Virtual Environment

Ayman Smadi, Ph.D.[1] and Shawn Birst, P.E.[2]

[1]Advanced Traffic Analysis Center, North Dakota State University, P.O.Box 5074 Fargo, ND 58105, PH (701) 231-8101; FAX (701) 231-6265; email: Ayman.Smadi@ndsu.edu
[2]Advanced Traffic Analysis Center, North Dakota State University, P.O.Box 5074 Fargo, ND 58105, PH (701) 231-1063; FAX (701) 231-6265; email: Shawn.Birst@ndsu.edu

Abstract
A recent assessment by the National Transportation Operations Collation revealed unsatisfactory performance in a number of traffic signal operation areas and an overall score of D- for the nation's traffic signals. This poor rating affects commuters nationwide and contributes to traffic congestion.

Numerous traffic simulation models have been developed to analyze more complex transportation facilities. However, these models may have some shortcomings in accurately modeling complex traffic signal controller functions (which could have a drastic effect on operations).

Hardware-in-the loop simulation (HILS), which replaces a simulation's signal control logic with actual controller hardware, allows advanced signal functionality to be tested and analyzed. Moreover, this simulation process opens new research areas in traffic controller functionality.

This paper discusses the development of a controller interface device (CID) by the Advanced Traffic Analysis Center (ATAC) to support HILS. This CID improves upon existing devices in a number of ways, including: it works with the VISSIM traffic simulation model, uses current off-the-shelf technology, more cost effective, easy to setup and use, small in size, and its Ethernet connectivity allows it to communicate with CIDs/controllers virtually anywhere in the world.

Introduction
There has been an increased focus on improving transportation system operations to meet safety and mobility needs through better management of existing capacity. One of the areas at the forefront of this effort is improving traffic signal operations. A recent assessment by the National Transportation Operations Collation revealed unsatisfactory performance in a number of traffic signal operations areas and an overall score of D- for the nation's traffic signals (National Transportation Operations Coalition 2005).

Traffic simulation models assist transportation engineers by providing a comprehensive method for analyzing complex transportation facilities. However, these models have not been fully embraced by the traffic engineering community due to a variety of reasons, including resource limitations, data limitations, and inherent shortcomings in accurately modeling the traffic signal controller functions (which could have a drastic effect on operations).

To assist in modeling more realistic traffic controller operations, hardware-in-the-loop simulation (HILS) was developed to test and evaluate traffic controller capabilities prior to field installation. Using this simulation method, the emulated traffic signal control logic of the simulation model is replaced by actual traffic controller hardware. This ensures accurate signal operations and enables the use of controller-specific functions not available in software simulation models. Moreover, HILS opens new research areas in traffic controller functionality, including low transit signal priority, coordination recovery methods, emergency vehicle and railroad preemption, and complex signalized intersections.

A controller interface device (CID) performs the data exchange between the simulation model and the traffic controller to support HILS. The CID developed by the Advanced Traffic Analysis Center (ATAC) at North Dakota State University (NDSU) improves upon existing devices in a number of ways, including: it works with the VISSIM traffic simulation model, uses current off-the-shelf technology, more cost effective, easy to setup and use, small in size, and its Ethernet connectivity allows it to communicate with CIDs/controllers virtually anywhere in the world.

CID Functionality

The criteria defined by ATAC for the CID included it must be easy to set-up and use, small in size, upgradeable, and economically feasible to build. In addition, the CID must include a NEMA TS 2-2003 serial connection and Ethernet connectivity.

CID Hardware Components

For the CID to function correctly, several hardware requirements had to be met. In addition to meeting the latency requirement of the NEMA TS 2 standard, the following hardware was required:

- 10/100Base-T Ethernet hardware for communicating with the host PC
- RS-232 port for performing the Ethernet configuration
- RS-485 SDLC hardware for communicating with the traffic controller
- Memory for allowing the CID to operate as a stand-alone device

After experimenting with our own designs, an off-the-shelf solution was chosen. This increased the size and power consumption of the CID, but removed the need for creating our own hardware, reducing cost and development time.

Embedded System

A PC/104 platform from Acrosser Technology, Inc. was selected for its compact form factor and an SDLC serial card from SeaLevel Systems, Inc. was

incorporated into the device. The AR-M9850E embedded system from Acrosser was chosen because it included a power supply, a case, an AR-B1423C embedded computer, and enough room for one expansion card. The AR-B1423C was a good match for the CID because it runs at 133MHz, uses inexpensive SODIMM RAM, has 10/100Base-T Ethernet, VGA video onboard, and DiscOnChip (DOC) support.

Memory

To operate as a stand-alone device, the CID requires some non-volatile memory. This stores the CID's operating system and runtime files, allowing it to start up by itself into a fully functional state. A 16 MB M-Systems DiscOnChip solid state hard disk was selected since is it uses very little power, essentially unlimited write cycles, fairly inexpensive, and has error correction and smart buffering. The AR-B1423C main board does not have built-in RAM; therefore, a 64MB SODIMM RAM card was incorporated and is more than adequate. Figure 1 illustrates the CID connected to a traffic controller.

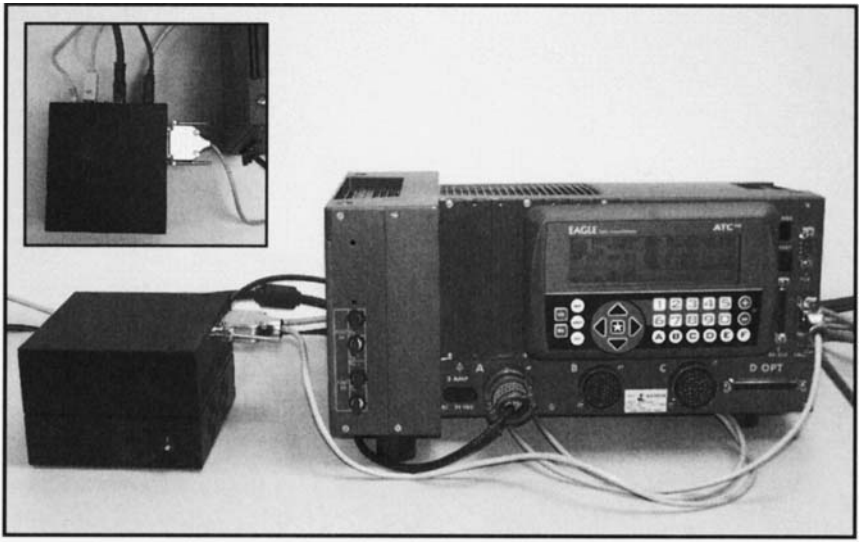


Figure 1. CID connected to a traffic controller.

CID Software Components

The FreeDOS operating system was selected because it is open source, is actively supported, and available for no cost. For the SDLC serial card, custom drivers were required to meet the CID's requirements. Several software programs were developed to run on the CID and a PC, including CIDserv, CIDlink, and SerUpdt Configuration Utility. These programs will be discussed in more detail in the following sections.

CIDserv software
The CIDserv software runs continuously on the CID. The program's major task is to keep the controller out "Diagnostic Flash Mode" by relaying relevant diagnostic information to it. The CIDserv program exchanges status information with a computer through the use of TCP/IP sockets over Ethernet.

CIDlink Software
The CIDlink software was developed to receive and transmit pertinent data with the CID and illustrate this data in graphical form (note Figure 2). CIDlink, which was developed using Java, is used for verification as well as testing since detectors (vehicle, pedestrian, and preemption) may be activated by mouse clicking on the appropriate detector. In addition, the CIDlink software has a "Virtual Cabinet" feature that allows the CID to act as a Malfunction Management Unit (MMU) and a Terminal and Facilities Bus Interface Unit (T&F BUI). It can also run in a "snoop mode" to observe the controller's signal and detector states.

SerUpdt CID Configuration Utility
The SerUpdt CID Configuration Utility is a program written in Visual Basic that allows easy access to the Serial Update function of the CID. Using the RS-232 serial port, the user can set and retrieve the CID's IP address, port number, subnet mask, and gateway. It also allows the user to save and retrieve configuration files on the local PC.

Simulation Interface
To incorporate additional simulation packages, no changes will be needed within the CID's firmware. Interfacing with different software simulation packages will only require modifying the software application residing on the PC.

VISSIM Interface
Interfacing the VISSIM simulation model to the CID is achieved through a dynamic link library (DLL) that runs within VISSIM for manipulating the detector and signal data in the simulation (PTV AG 2006). Running on top of the DLL framework, which was developed by PTV AG, a socket interface sends detector information to the CID and receives signal state information from the CID. One instance of the DLL is created for the simulation and communicates with all of the CIDs in the network.

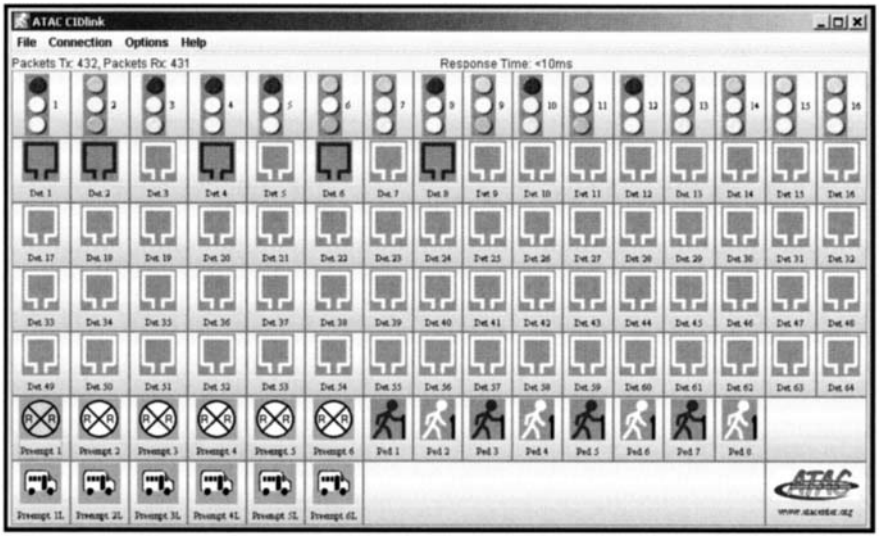


Figure 2. CIDlink program for testing and verification.

Traffic Controller Configuration

The NEMA TS 2 controller needs to be configured correctly for operation with the CID (National Electrical Manufacturers Association 2003). It will need to have the appropriate addresses enabled in order to allow it to send the required frames over its Port 1 or SDLC interface. The CID will receive these frames and then emulate the destination device. The CID emulates of nine cabinet devices, including the four T&F BIU devices, four Detector BIU devices, and the MMU.

CID Operation

The CID replies to diagnostic information from the traffic controller, sends signal states from the traffic controller to the PC, and sends detector signals from the PC back to the traffic controller (note Figure 3).

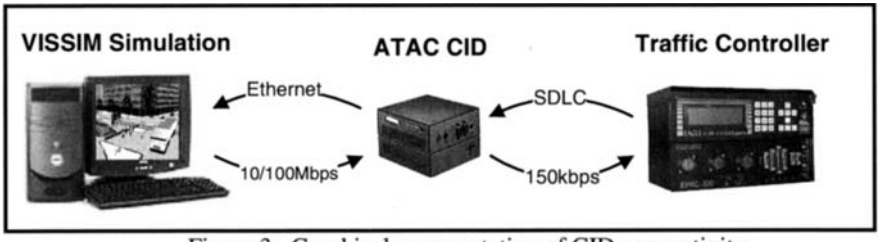


Figure 3. Graphical representation of CID connectivity.

When VISSIM is simulating, the traffic logic is executed at the beginning of each time step. VISSIM calls the control program (SC_DLL.dll) to calculate the signal states for the next time step. The CID will regulate the speed of the simulation by

pausing VISSIM for a specified amount of time, which is based on the number of time steps per simulation second (0.1 to 1.0 seconds). For each time step in VISSIM, the following tasks are performed:

- Detector information is gathered and sent to the CID,
- Detector information is formatted into NEMA TS 2 protocol and sent to the traffic controller,
- Traffic controller responds to the detector data and provides signal states,
- Signal states are sent to the CID, formatted into VISSIM protocol, and relayed to the PC, and
- Signal states are sent to VISSIM and displayed within the simulation network.

Since the CID controls the timing and issues its signal data asynchronously from the detector data, a simulation can be run for an extended period of time and have zero net time error, even in the presence of network delays, VISSIM hardlock checks, PC slowdowns, etc.

Conclusions

Hardware-in-the loop simulation (HILS) allows for more accurate simulation of signalized intersections, especially when advanced controller functionality is tested or analyzed. The CID developed by ATAC improves upon existing devices in a number of ways. Probably the most important attribute of the CID is the Ethernet connectivity, allowing it to communicate with CIDs/controllers virtually anywhere in the world. Therefore, agencies, companies, and universities can share CID/controller units for a various projects.

Acknowledgements

The authors would like to acknowledge the efforts of those involved in development of ATAC's CID, including Andrew Paulsen, Ryan Schumacher, Jason Gates, and Kiel Ova.

References

National Transportation Operations Coalition. (2005). *National Traffic Signal Report Card, Technical Report.*

PTV AG. (2006). *VISSIM 4.00 – 4.10,* Karlsruhe, Germany. http://www.ptv.de/cgi-bin/traffic/traf_vissim.pl>

National Electrical Manufacturers Association. (2003). NEMA TS 2, "Traffic Controller Assemblies with NTCIP Requirements, Version 02.06.

Advantages of Using Innovative Traffic Data Collection Technologies

Ayman Smadi, Ph.D[1] Jason Baker, EIT[2]; and Shawn Birst, P.E[3]

[1]Advanced Traffic Analysis Center, Upper Great Plains Transportation Institute, North Dakota State University, P.O. Box 5074, Fargo, ND 58105; PH (701) 231-8101; FAX (701) 231-6265; email: Ayman.Smadi@ndsu.edu

[2]PH (701) 231-1086; FAX (701) 231-6265; email: Jason.Baker@ndsu.edu

[3]PH (701) 231-1063; FAX (701) 231-6265; email: Shawn.Birst@ndsu.edu

Abstract

Accurate traffic data are essential for supporting a multitude of transportation-related decisions which affect transportation system operations, management, and planning. However, the lack of accurate and timely data is often cited as a common problem for both transportation practitioners and policy makers alike.

This paper discusses the temporary use of advanced sensor technologies for collecting and presenting traffic data for supporting transportation system operations. It illustrates the use of video-based and radar-based traffic detection technologies. The range of applications included in this paper comprises site-specific traffic assessment, traffic signal operations, freeway performance evaluation, and highway work zones. In addition, the paper illustrates the significant value of video data in supporting transportation decisions among policy makers and the public.

The paper covers two classes of advanced traffic detection sensors; the Autoscope Video Vehicle Detection system and Wavetronix SmartSensor Traffic Detector based on Digital Wave Radar. It discusses the potential applications, typical set up requirements, types of data generated by these systems, and the performance of these technologies under different traffic collection environments.

The information contained in this paper is based on various traffic operations projects conducted at the Advanced Traffic Analysis Center at North Dakota State University. These projects range from analysis of school traffic safety to evaluating traffic operations during a major multi-year construction project along Interstate Highway 29 in Fargo.

Introduction

The lack of adequate data is often cited as a major obstacle to effective management of transportation systems, including real-time operations, performance measurement, and future demand forecasts. Although this shortage of data has been around for a number of years, recently several trends have further intensified the need for accurate and timely data. First and foremost among these trends is a national focus on transportation system operations and real-time system management. Second among these trends is increased public scrutiny of transportation projects, resulting in a need to better justify transportation decisions and effectively communicate traditionally technical information to non-technical audiences. Another trend in an era of fiscal restraint is the need for modeling tools that support transportation funding requests and clearly demonstrate expected benefits from transportation investments.

With the increased focus on system operations there has been an increased attention to traffic data collection methods, technologies, and practices. However, traffic data collection remains an under-funded activity by most transportation agencies. That could be due in part to the lack of clear benefit-costs of transportation data. This problem is more serious for small transportation agencies that lack the data collection budgets, staff, and required infrastructure.

Potential data applications may be categorized into three groups: system operations, system planning, and infrastructure management. System operations tend to require more detailed data collected over a relatively short time period (i.e., real-time). Specific traffic operations applications include traffic signal operations, freeway operations, incident management, etc. Traffic monitoring and data collection technologies may be intrusive or non-intrusive. Intrusive systems are usually installed in the road, with loop detectors being the most common type. Non-intrusive technologies have gained popularity since the late 1980s and include video, radar, and ultrasonic (Klein and Mimbela 2000).

Traffic data collection may be permanent (continuous) or temporary. Permanent counting stations record traffic on a continuous basis to support highway performance and management. Temporary traffic data collection is conducted to support a variety of special studies, ranging from traffic signal timing to parking lot operations. Temporary data collection can be a time and resource intensive process. It may also have safety and operational impacts on the crews conducting the data collection and on the road users, depending on the technology used. Temporary traffic data collection will be the main focus of this paper.

This paper illustrates the use of two classes of advanced sensor technologies, video-based and radar-based systems for temporary traffic data collection. The range of applications that may be supported by these sensors includes site-specific traffic assessment, traffic signal operations, freeway performance evaluation, and highway work zones. The paper provides a discussion of potential applications, typical set up requirements, types of data generated by these systems, and the performance of these technologies under different environments.

Issues with Traditional Traffic Data Collection Technologies

Inductive loops and road tubes are among the most commonly used temporary traffic data collection technologies. However, these methods lack the flexibility and cost-effectiveness when used for temporary or special purpose data collection projects. Loop detectors are the predominant detection technology used to support actuated traffic signals. They are also used to collect counts across freeway and arterial sections. However, most transportation agencies generally install a minimum number of loops to support signal operations. In adequate number of and configuration of loops makes it impossible to collect detailed traffic counts. Manual counts are often conducted to supplement automated data collection, i.e. intersection turning movement counts. Because of their installation requirements, loops are infeasible for mobile applications. Road tubes on the other hand are designed for temporary applications. However, they present a number of challenges, including: mixed accuracy for simple axle counts, limited data detail (long recording intervals), difficulty of setup, and potential safety hazards to crew and road traffic.

Mobile Traffic Data Units

In order to meet traffic analysis data requirements, two systems were developed to allow easy and comprehensive traffic data collection in the field (note Figure 1). These systems use two different detection technologies: video and radar.

Video-Based System

The primary component of this system is the Autoscope 2004 Wide Area Video Vehicle Detection System which processes video images to develop traffic volumes, speeds, and classification. The system is housed in a cargo trailer and consists of two video cameras with zoom and pan/tilt devices, 42' telescopic mast, air compressor, and video recorders. The system is powered by a gas generator.

Radar-Based System

This system uses the radar-based Wavetronix SmartSensor. A mounting system was developed to deploy the sensors for various applications. Each sensor is powered by a deep-charge marine battery. The system consists of five sensors and corresponding mounting platforms, which is suitable for highway applications requiring multiple data points. The sensors, power supplies, and mounting platforms are all housed in a standard cargo trailer.

Set-up requirements

The set-up requirements for traffic data collection devices play a major role in how and when they can be used for various scenarios. There are many factors that need to be taken into account as far as setting up the sensors, such as location, complexity of the user interface (software and hardware), the ease of the deployment, and the type of roadway facility being analyzed.

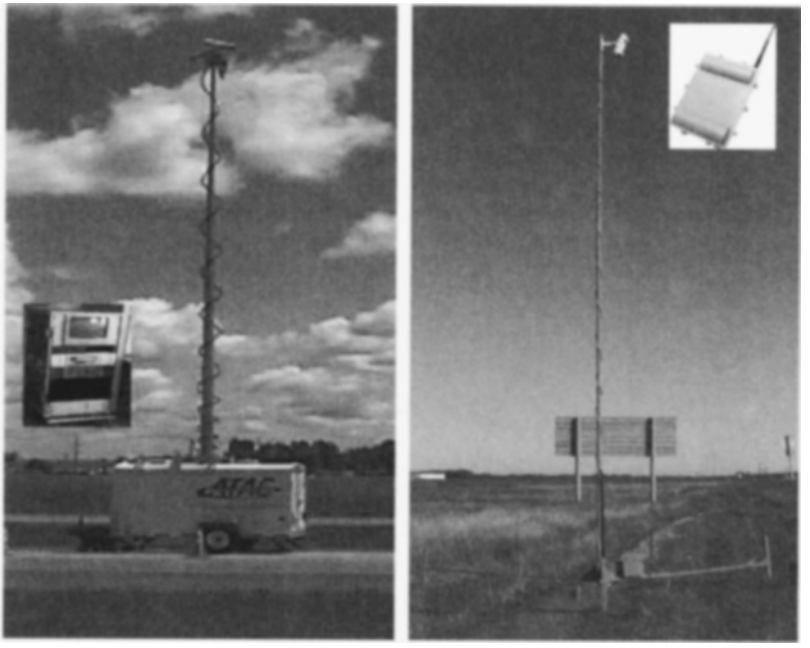


Figure 1 Video-Based (left) and Radar-Based (right) Traffic Data Systems

The Wavetronix SmartSensor may be mounted on a free-standing base or to an object adjacent to the study location (Wavetronix 2004). The height and offset must be taken into account when deploying the sensor. Recommended mounting height range from 12ft to 30ft and offset from 25ft to 35ft. The sensor can be operated in a forward-fired and side-fired mode (best suited for temporary data collection). The sensor is capable of detecting up to eight lanes of traffic at a range of 200-ft. It has an "auto calibration" feature allowing it to automatically assign lanes after observing traffic.

The set-up requirements for the Autoscope system are a bit more involved. The appropriate detectors must be manually configured on a background image of the video. It is important to note that for the most accurate performance of this system, the camera must be placed in such a way so as to reduce any possible occlusion.

Potential applications

The SmartSensor and Autoscope can be used in a variety of different applications. They can both be used as either permanent or temporary data collection devices. Both of these sensors have been used in various data collection projects, including:

1. Work zone traffic analysis
2. Intersection traffic and turning movement analysis
3. Pedestrian volume counts
4. School traffic operation and safety analysis

5. Interchange traffic operations and safety analysis

Video data have been successfully used to illustrate unsafe behavior, quantify delay, and capture potential operations or geometrical problems. The data collection unit which is housed in a plain cargo trailer is often unnoticeable, so it does not interfere with normal driving behavior. The video is strategically used in presentations to provide compelling evidence about potential problems. To illustrate, we provide two common examples: a school traffic study and an intersection delay study.

Elementary and middle schools experience several traffic circulation and safety problems, due to parents picking up or dropping off their kids at the school. This operation usually happens over a short period of time, also coinciding with school bus and pedestrian operations. The video captures the dynamics of this situation and illustrates that parents may be more responsible for problems, and thus alleviate the pressure on traffic engineer to install unwarranted traffic control devices.

For intersection operational studies and traffic signal warrant analysis, the public often has a different perception about the extent of delay. Although traffic engineers rely on the Manual of Uniform Traffic Control Devices (MUTCD), there are often pressures from the public and political offices to install unwarranted signals. Using video data can greatly enhance how the public receive technical data and accurately illustrates actual delays.

Types of data collected

The SmartSensor and Autoscope sensors are capable of collecting volume, speed, occupancy, and classification. Autoscope can also be set up to detect incidents and measure queues. In addition, raw video data of the location are often as valuable as the quantitative data since they provide valuable insights on traffic operations. Figures 3 and 4 show examples of the types of data collected by the two systems.

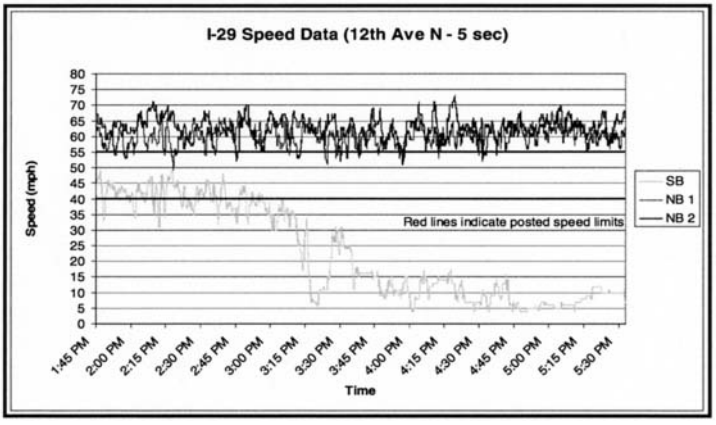


Figure 2 SmartSensor Speed Data

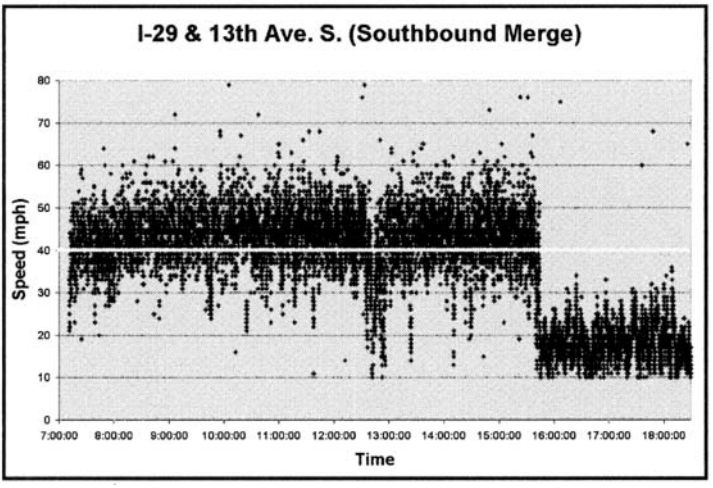


Figure 3 Autoscope Speed Data

Summary/Conclusion

Advances in sensor technologies provide tremendous opportunities for improving traffic data collection. Portable systems that use video or radar detector technologies have been successful deployed and used to support a multitude of applications. In addition to the rich quantitative data collected by these systems, video data offer additional benefits, especially for delivering information to policy makers and the affected/interested public. Video data when used with other analysis tools (i.e., traffic simulation) can be quite powerful in demonstrating current problem areas and justifying proposed solutions.

References

Klein, Lawrence A. and Mimbela, Luz Elena Y. (2000). *Summary of Vehicle Detection and Surveillance Technologies Used in Intelligent Transportation Systems*. Federal Highway Administration, Intelligent Transportation Systems Joint Program Office.

Wavetronix. (2004). *SmartSensor Installation Guide*.

Determining Origin-Destination Flows across a Two-intersection Network from Non-overlapping UAV and Ground-based Imagery

Mark R. McCord[1], Rabi G. Mishalani[2], Benjamin Coifman[3], Yuxiong Ji[4], and Michael V. Iswalt[5]

[1]Professor, Department of Civil and Environmental Engineering and Geodetic Science and Knowlton School of Architecture, The Ohio State University, 470 Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210; PH (614) 292-2388; FAX (614) 292-3780; e-mail: mccord.2@osu.edu

[2]Asscociate Professor, Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, 470 Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210; PH (614) 292-5949; FAX (614) 292-3780; e-mail: mishalani@osu.edu

[3]Asscociate Professor, Department of Civil and Environmental Engineering and Geodetic Science and Department of Electrical and Computer Engineering, The Ohio State University, 470 Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210; PH (614) 292-4282; FAX (614) 292-3780; e-mail: coifman.1@osu.edu

[4]Graduate Research Associate, Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University, 470 Hitchcock Hall, 2070 Neil Avenue, Columbus, OH 43210; PH (614) 688-3761; FAX (614) 292-3780; e-mail: ji.28@osu.edu

[5]Transportation Engineer, Fehr & Peers Transportation Consultants, 2990 Lava Ridge Court, Suite 200, Roseville, CA 95661; PH (916) 773-1900; FAX (916) 773-2015; e-mail: m.iswalt@fehrandpeers.com

Abstract

Origin-destination (OD) flows are important inputs to both off-line planning and real-time traffic management applications. Typically OD flows are estimated from time-consuming and expensive travel surveys or license plate matching studies. These estimates can be updated using link-based observations. The use of Unmanned Aerial Vehicles (UAVs) to monitor traffic networks introduces the possibility of directly determining OD flows for small networks. A methodology that combines non-overlapping imagery is developed and applied to a two-intersection network. The empirical findings show that even with short imaging time, interesting and nontrivial results can be obtained.

Introduction

Origin-destination (OD) flows are important inputs to both off-line planning and real-time traffic management applications. Typically OD flows are estimated from time-consuming and expensive travel surveys or license plate matching studies. These estimates can be updated based on link volume measurements using point-location sensors, travel times, and intersection turning movements (Ashok and Ben-Akiva, 2000; Mishalani et al., 2002).

The use of Unmanned Aerial Vehicles (UAVs) to monitor traffic networks, the motivation of which is discussed in Coifman et al. (2006), introduces the possibility of determining OD flows directly instead of updating historical ones, at least for small networks. This paper presents such a methodology. Furthermore, an empirical study that demonstrates the value of the methodology is presented. Some limitations and their impacts on the results are also discussed.

Data

On July 22, 2003 multiple experiments were conducted on the campus of The Ohio State University using the BAT III UAV technology (MLB, 2005) carrying a payload of two video cameras. The UAV flew at an altitude of 150 m and an air speed of approximately 50 km/h while transmitting video images to the ground station in real-time. The flight lasted almost two hours and data were collected from several facilities. A general overview of the set of experiments is presented in Coifman et al. (2006).

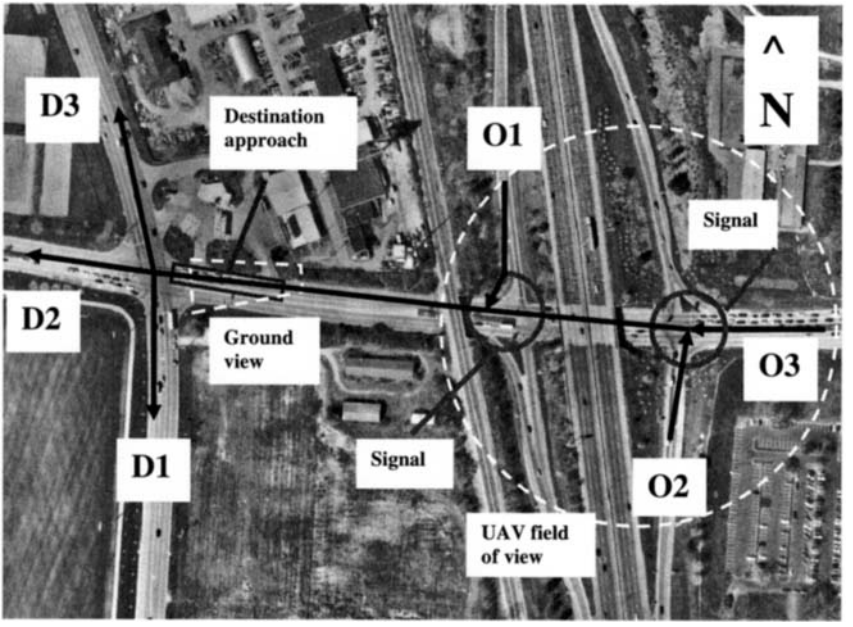


Figure 1. Network and fields of view.

Figure 1 shows a map of the network considered in the study presented here. The UAV's field of view (FOV) is the area inside the dashed circle, the area from where the upstream flows originate. The FOV is limited because the UAV was required to fly below the FAA imposed maximum altitude for remotely controlled aircraft. Therefore, the destinations were captured by video data collected using a ground-based camera viewing the area inside the dashed trapezoid depicted in Figure 1. The camera is pointing westward and, therefore, the vehicles' destinations can be readily observed. Alternatively, another UAV could be used to

achieve this extended coverage. The two fields of view are non-overlapping. Therefore, vehicles could not be directly tracked from origin to destination.

Methodology

The measurement of flows between origins and destinations could theoretically be accomplished through direct one-to-one matching of all vehicles between origins and destinations. However, this task would prove tedious over large samples, and in this case would be very difficult, if not impossible, given the resolution of the imagery and the distance between the non-overlapping UAV and ground-based camera fields of view. Instead, a methodology was developed in which vehicles are matched from the various origins to an approach feeding into the downstream intersection, and subsequently the vehicles are followed through to their respective destinations.

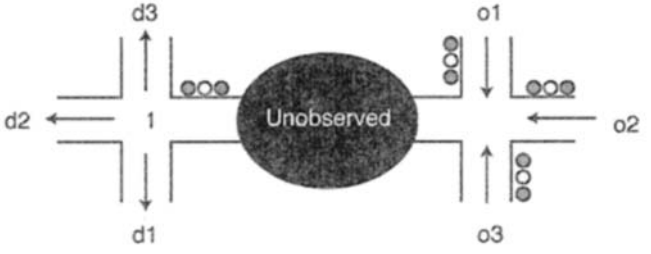


Figure 2 (a). Illustration of the network.

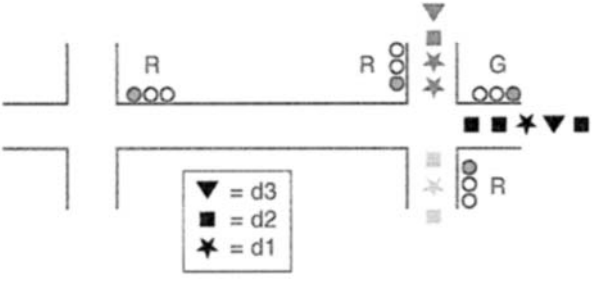
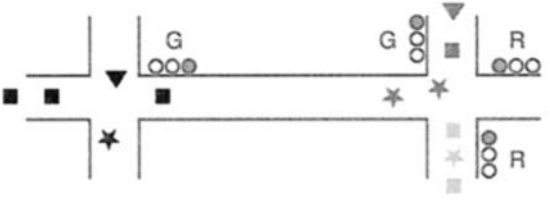


Figure 2 (b). State of the network just as the O2 platoon receives the green light.



The matching methodology avoids the need to directly match individual vehicles, but rather relies on the presence of platoons. In brief, platoons will generally contain vehicles from the same origin, and any unobserved reordering within a platoon from a given origin will not change the proportion of vehicles bound for a given destination. Key to the feasibility of this methodology is the fact that headways between platoons, induced by clearance intervals at the upstream signal, are normally much larger than headways between vehicles within platoons. Large headways mean overtaking between platoons is generally uncommon, and if it does occur, it will only be among the few vehicles at the end of the platoon. Figures 2 (a) through 2 (c) illustrate the process, where each origin is given one shade and each destination one symbol. In this hypothetical example, and as seen in Figure 2 (c), there is a gap after the platoon from O2 has departed the upstream intersection (on the right) before the platoon from O1 is released. Given that overtaking is not likely to occur, at least in an extensive manner, the first-in-first-out (FIFO) assumption is reasonable and, therefore, vehicles can be indirectly matched based on the order in which they leave the UAV FOV and the order in which they enter the ground-based camera FOV.

A computer interface was developed to record the origins, departure times, and lanes in which vehicles depart the upstream intersections (within the UAV FOV). A similar process was repeated for the vehicles at the downstream intersection, from arrival to the destination approach (within the ground-based camera FOV – see Figure 1) to their final destination. Within each intersection the vehicle movements were followed, yielding a map to the particular origin or destination. A few vehicles departed the network at the southbound freeway on-ramp (see Figure 1), which were captured by the UAV FOV. These vehicles were excluded from the analysis. The remaining departing and arriving vehicles were matched using the FIFO-based methodology presented above.

Empirical Results and Methodology Assessment

Table 1 shows the OD flows determined by the above methodology as applied to the empirical data over a period of approximately 8 minutes. The value of the methodology is assessed with respect to three criteria: (1) whether there is a reasonable indication that the resulting OD flows are accurate, (2) whether a small sample based on tracking a few vehicles could produce sufficiently accurate OD patterns, and (3) whether certain sources of errors that might cause the developed methodology to fail are likely to be present and whether they could be detected.

Table 1. Determined OD flow.

	D1	D2	D3	Origin flows
O1	18	32	1	51
O2	11	42	2	55
O3	5	8	3	16
Dest. flows	34	82	6	122

Regarding the first criterion, ideally the region between the two non-overlapping FOVs would be observed so that the determined OD flows of Table 1 could be evaluated with respect to the actual measured OD flows. While a complete set of OD flow measurements is not available, a subset of vehicles (14 out of 122 comprising 11.5%) could be tracked through the network based on physical characteristics. The developed matching methodology correctly associated the destinations to the origins of all the tracked vehicles providing an initial indication that the methodology has potential to produce accurate results.

Regarding the second criterion, the determined flows of Table 1 are first compared to the expected OD flows reflecting the case where there is independence between the origins and the destinations; that is, information on the origin (destination) of a vehicle does not affect its probability of going to a destination (coming from an origin). The flows that would reflect independence are presented in Table 2 and are calculated from the total number of vehicles originating from each origin as imaged from the UAV and the total number of vehicles destined to each destination as imaged from the ground-based camera. For example, the O1-D1 entry in Table 2 is computed as the product of the total flows (122), the estimate of the marginal probability of a vehicle originating from O1 (51/122), and the estimate of the marginal probability that a vehicle is destined to D1 (34/122). The null hypothesis of independence is rejected at the 5% level for pairs O2-D2 and O3-D3, suggesting that even with only 8 minutes of data, the developed methodology is capable of producing interesting origin-destination patterns different from those associated with independence between origins and destinations. In contrast, the OD flows based on the 14 tracked vehicles, which are shown in Table 3, do not allow the rejection of the independence hypothesis for any OD pair and reflect large discrepancies with the patterns in Table 1. Therefore, even though the origins and destinations of the 14 tracked vehicles are accurately captured by the developed methodology, the small sample size associated with those 14 vehicles (11.5%) obscured important OD patterns revealed by the developed methodology, which takes advantage of the entire set of observed vehicles.

Table 2. OD flows assuming origins and destinations are independent.

	D1	D2	D3	Origin flows
O1	14	34	3	51
O2	15	37	3	55
O3	4	11	1	16
Dest. flows	34	82	6	122

Table 3. OD flows of directly matched vehicles and in parenthesis the values scaled up to achieve a total of 122.

	D1	D2	D3	Origin flows
O1	1 (9)	5 (44)	0 (0)	6 (53)
O2	2 (17)	4 (35)	0 (0)	6 (52)
O3	0 (0)	2 (17)	0 (0)	2 (17)
Dest. flows	3 (26)	11 (96)	0 (0)	14 (122)

Finally, regarding the third criterion, the developed methodology could suffer from two important limitations. First, measurement error in the form of missing the detection of vehicles at the origin or destination ends affects the developed indirect FIFO-based matching methodology, thus potentially producing erroneous OD flows. A preliminary analysis indicates that such errors are limited, especially if an undetected vehicle at the origin end can be narrowed down to a subset of 20 vehicles at the destination end, or vice versa. Furthermore, a methodology for detecting the presence of such an error was developed based on monitoring the derived travel times from the origin end to the destination end. Therefore, at the minimum the presence of potential errors in the resulting OD flows could be highlighted.

The second important limitation relates to potential FIFO violations resulting from vehicle overtaking. As discussed in the methodology section, a platoon and its turning movements observed at the downstream intersection can be traced back to a unique origin. Therefore, overtaking that occurs within a platoon from a single origin does not result in any OD flow error. However, if vehicles in one platoon overtake vehicles in the preceding platoon, an incorrect OD flow may result. While inter-platoon overtaking is not expected to be common due to the large headways separating platoons, the quantitative analysis of such situations and the extent of their impact are reserved for future research.

Discussion

Knowing the OD flows in small networks could be interesting for signal coordination, or route or destination choice studies. The developed methodology would be useful for two non-overlapping fields of view of relatively closely spaced intersections. The UAV offers the potential to image an intersection for a short period of time and then travel quickly to another intersection, thereby providing the ability to conduct multiple studies as those illustrated here. While assessing the methodology using a data set covering a longer period of time is worthwhile as part of future research, the empirical results show that even with short imaging time, interesting and nontrivial results can be obtained.

Acknowledgements

This study was funded through a grant to the National Consortium for Remote Sensing in Transportation-Flows (NCRST-F) at The Ohio State University from the US Department of Transportation. The efforts of Steve Morris from MLB in providing the aircraft and operator for the field experiment are particularly appreciated.

References

Ashok, K., and M. Ben-Akiva M. (2000). "Alternative Approaches for Real-Time Estimation and Predication of Time-Dependent Origin-Destination Flows." *Transportation Science*, Vol. 34, No. 1.

Coifman B., M. McCord, R. Mishalani, M. Iswalt, and Y. Ji (2006). "Traffic Flow Data Extracted from Imagery Collected using a Micro Unmanned Aerial Vehicle." *Proceedings of the 8th International Conference on Applications of Advanced Technologies in Transportation*. ASCE, Chicago, Illinois.

Mishalani R.G., B. Coifman, and D. Gopalakrishna (2002). "Evaluating Real-Time Origin-Destination Flow Estimation Using Remote Sensing-Based Surveillance Data." *Proceedings of the 7th International Conference on Applications of Advanced Technologies in Transportation*, K.C.P. Wang, S. Madanat, S. Nambisan, and G. Spring, Eds., ASCE, Cambridge, Massachusetts.

MLB. MLB Company. http://spyplanes.com, accessed on December 8, 2005.

Traffic Demand Reduction Using an Automated Work Zone Information System for Urban Freeway Rehabilitation

Eul-Bum Lee,[1] Chang M. Kim,[2] and John T. Harvey[3]

[1] Research Engineer, Institute of Transportation Studies, University of California at Berkeley, 1353 S. 46th Street, Bldg. 452, Richmond, CA 94804; PH: (510) 665-3637; FAX: (510) 665-3562; email: eblee@berkeley.edu

[2] Graduate Student, Department of Civil and Environmental Engineering, University of California at Davis, EU(III) Rm. 1001, One Shields Ave. Davis, CA 95616; PH: (530) 754-6429; FAX: (530) 752-7872; email: chkim@ucdavis.edu

[3]Associate Professor, Department of Civil and Environmental Engineering, University of California at Davis, EU(III) Rm. 3139, One Shields Ave. Davis, CA 95616; PH: (530) 754-6409; FAX: (510) 231-9589; email: jtharvey@ucdavis.edu

Abstract

In October 2004, a 9 lane-km stretch of badly deteriorated truck lanes on Interstate-15 (I-15) at Devore in Southern California were rebuilt during two 9-day, one-roadbed continuous closures, and around-the-clock (24/7) construction operations utilizing the counter-flow traffic system. A preconstruction traffic sensitivity study anticipated that a nominal reduction of 10 percent of peak-hour traffic demand through the construction work zone (CWZ) would cause maximum traffic delays of as much as 95 minutes, while a 20 percent reduction would reduce delays to as little as 45 minutes. The I-15 Devore Project aimed to achieve a 20 percent reduction in traffic demand by encouraging road users to change their travel patterns; to do this the project implemented both the Automated Work Zone Information System (AWIS) and a proactive public outreach program via the Internet. This paper describes the configuration and performance of the AWIS, which provided road users with estimates of travel times through the CWZ. AWIS did this by using a built-in algorithm based on measured real-time traffic speed and occupancy data. AWIS travel times were displayed on-site on portable and permanent changeable message signs and off-site on the project web site. The AWIS information contributed to reducing more average daily traffic demand reduction (up to 19 percent) thorough the CWZ, producing high traffic volume increases to detour freeways. The operation of AWIS and the public outreach effort contributed to decreasing the maximum weekday peak-hour delay through the CWZ from 95 min. to 50 min. (on average). The outcome of this case study will help transportation agencies and practitioners efficiently design and operate AWIS for highway rehabilitation under high-traffic volume.

Introduction

Traffic Management for Urban Highway Rehabilitation

Most urban highway rehabilitation projects in California and other states are under heavy traffic volumes, and they therefore cause serious traffic disruptions for the communities that use the highways. This results in major inconvenience for the traveling public and for commercial enterprises. A traffic management plan (TMP) for traffic demand control over a rehabilitation corridor becomes a primary issue in these cases. Transportation agencies often combine intelligent transportation systems with TMP in highway rehabilitation projects.

The Arkansas Department of Transportation (DOT) implemented an automated work zone information system (AWIS) at rural work zones on Interstate 40 (I-40). The objective of this implementation was to induce traffic demand reductions by comparing AWIS displays with field measurements of traffic condition and to determine traffic volume backups forming in construction work zones (CWZ) (Tooley et al 2004). Fontaine (2003) reviewed AWIS implementations in six states. He remarked that five of them — Maryland, Iowa, Kentucky, Illinois, and Ohio — either did not achieve successful results or did not evaluate the system's performance in terms of traffic diversion. In Nebraska, the sixth state, the AWIS implementation only achieved an insignificant 3 percent change in the volume of traffic diversion.

I-15 Devore Reconstruction Project

The I-15 Devore Reconstruction Project was located between Interstate-10 (I-10) and Interstate-215 (I-215) in the City of Devore in Southern California (see Figure 1). In October 2004, a 9 lane-km stretch of badly deteriorated concrete truck lanes was reconstructed there; the reconstruction included a new asphalt concrete base and a new concrete slab. Under high traffic volumes, one truck lane in one direction was rebuilt in 210 hours (about 9 days) using a one-roadbed continuous closure, 24/7 accelerated construction, and a counter-flow traffic system.

The I-15 Devore Corridor consists of three or four lanes for each direction and carries approximately 110,000 Average Daily Traffic (ADT), about 10 percent of which is heavy trucks. The corridor has uniquely high commuter peaks on weekdays, as well as high traffic volume on weekends when recreational motorists travel through it between Los Angeles and Las Vegas. The highest traffic volumes northbound (NB) appear on Friday afternoon and the highest traffic volumes southbound (SB) appear on Sunday morning.

While one roadbed was closed completely for the duration of construction activities, the other one was operated with five lanes, including one temporary shoulder lane. Using quick-change moveable barriers (QMB) to provide dynamic lane configuration (three by two lanes), the open roadbed carried traffic in both directions. A counter-flow traffic control system made it possible to provide three lanes in the peak direction during each peak period, with minimal traffic interruption.

According to the preliminary traffic analysis using the Highway Capacity Manual Demand-Capacity (D-C) Model (TRB 2000), a nominal 10 percent reduction in traffic demand through the CWZ (believed to be the most realistic goal achievable in this network geometry) would still produce maximum traffic delays of as much as 95 minutes. Achieving a 20 percent reduction of peak-hour traffic demand could reduce the 95-minute delay to as little as 45 minutes and could smoothly expedite the approach of extended closures with their round-the-clock (24/7) operations. (Lee et al. 2005). To achieve a 20 percent reduction of traffic demand, Caltrans implemented AWIS and a proactive public outreach program (that especially used the Internet) to encourage more road user "no-shows," travel pattern changes (mode and trip time), and diversions to detours.

Research Approach

As reported here, the I-15 Devore research team designed and configured the AWIS hardware and software — including the development of an algorithm for real-time estimation of CWZ travel time. Each AWIS component and its connectivity with a computer server were ensured. Locations of the portable, changeable message signs (CMS) were carefully selected to induce more traffic diversion to detour roads. For validation and performance evaluation of the AWIS, estimated travel times on the I-15 CWZ corridor were compared with travel times measured by probe vehicles. The travel times estimated in AWIS were displayed in real time on the CMS and the project website.

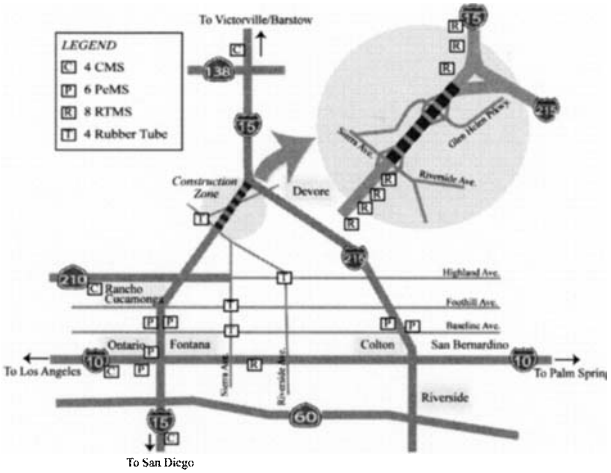


Figure 1. Locations of the Traffic Devices Employed in the Study Area

The project website was developed to provide real-time traffic information about the freeway network to update travelers of traffic conditions along their routes through CWZ prior to departing. A traffic measurement study was performed to capture traffic impacts during construction within the CWZ corridor and on neighboring freeways and local roads, as well as to investigate the contribution of implementing the AWIS. Detailed descriptions of the processes and their outcomes are presented in the following sections.

Automated Work Zone Information System (AWIS)

System Overview

The I-15 Devore AWIS consisted of three major components: traffic monitoring devices (Remote Traffic Microwave Sensor; or RTMS), the portable CMS, and the server station to estimate travel time in the programmed algorithm. According to the per-construction traffic analysis, length of the longest traffic queue was expected to be about 4.8 km in NB traffic. Three traffic monitoring stations for NB and two for SB were located at 1.6-km intervals upstream of the CWZ start. In addition, one traffic monitoring device was located on I-10 EB (eastbound) to check its traffic condition as a main detour freeway route. Three portable CMSs were installed for NB traffic and one was installed for SB traffic. Each was located upstream of the junction with I-15 to enable the travelers to change their routes to detours before entering the I-15 CWZ corridor. Figure 1 illustrates the locations of all the field devices, including traffic monitoring stations and information displaying stations.

The traffic data collected at the monitoring devices were transmitted to the server station for the travel-time estimation. After travel time was estimated from the algorithm, it was evaluated against lane-occupancy data. Travel information was then generated and transmitted to both the portable on-site CMS and the project web site in real time. The traffic information messages were classified into three levels; free flow, intermediate, and severe congestion, depending on the severity of the traffic delay through the CWZ. Figure 2 shows the configuration of the I-15 Devore AWIS.

Performance Monitoring

The travel times estimated by AWIS were compared with the actual travel times measured by the probe vehicles in order to validate the automated system's ability to predict travel time. Overall, the data showed a good match between the AWIS estimates and the travel time measured during both the peak and non-peak construction hours. Two specific dates during construction were chosen and the travel times estimated AWIS were compared with the actual travel time measured by the three probe vehicles for the dates chosen. A comparison of the estimated and actual travel times during the SB construction (the second closure) are shown in Figure 3. The average measured travel time of 7 samples for I-15 SB (peak direction) during morning peak hours was 34.7 minutes (standard deviation (S.D.) 4.5 minutes) and the average estimated was 36.4 minutes (S.D. 6.9 minutes). The average measured travel time of 11 samples for I-15 NB (peak direction) during afternoon peak hours was 51.3 minutes (S.D. 20.4 minutes) and the average estimated travel time was 45.0 minutes (S.D. 17.8 minutes).

The traffic measurement study was also performed before and during construction to investigate the impacts on traffic caused by the extended closures, and to assess the contribution of AWIS implementation. Traffic volumes were collected to quantify traffic reduction on the I-15 CWZ corridor and traffic volume changes on the neighboring freeways and local roads.

The traffic volumes collected by the two RTMS on the I-15 CWZ were used to compare traffic volumes before and after construction. The Average Daily Traffic (ADT) for I-15 SB decreased by 19 percent (ADT 52,900 to 42,700) and the ADT for I-15 NB decreased by 16 percent (ADT 52,500 to 43,900) during weekday closures. Those reductions were larger than Caltrans initially expected and close to the reduction goal that the project planners had hoped to achieve through AWIS implementation. Due to a greater than expected reduction in traffic demand during construction, it was estimated that the weekday peak-hour traffic delay through the CWZ was reduced by as much as half, from 95 to 50 minutes.

The traffic measurement study also showed that traffic impacts of the extended closures were widely spread on the freeway network. The source of these data was the California Freeway Performance Measurement System (PeMS) database; the study extracted traffic flow and speed data on selected locations in the freeway network — I-15 outside the CWZ, I-10, and I-215 — to quantify the impact of the extended closures (PeMS 2004). The ADT (24,000) on I-215 SB during the extended closures was 15 percent larger than the usual ADT (20,800) before construction, and the ADT (85,812) on I-10 EB during the closures was also 10 percent greater than the usual ADT (77,934) before construction. The ADT increase on I-215 SB indicated a shift in traffic there from I-15 SB. The increased ADT on I-10 EB is explained by the diversion of traffic from I-15 NB to I-10 EB, then to I-215 NB.

In order to investigate the traffic impact on local roads, traffic volumes were monitored at the four major intersections on alternate local roads south of the I-15 CWZ for four days before and during the extended closures. The intersection of Lytle Creek and Sierra was the closest to the I-15 CWZ and there was a 7 percent increase (10,761 to 11,489) in ADT. A 2 percent increase (49,177 to 49,972) in ADT was observed at intersections during the extended closures. These increases were attributed to local motorists making short trips shifting to local streets to avoid the I-15 CWZ. Table 1 contains the data showing traffic-flow reduction over the network.

Table 1. Traffic Flow Reductions on the Network

Category		CWZ		Detour Freeways		Local
		I-15SB	I-15 NB	I-215 SB	I-10 EB	Roads
Weekdays	ADT	-19%	-16%	15%	10%	2%
	Peak hrs	-18%	-17%	16%	36%	2%
Weekends	ADT	-12%	-20%	n/a	n/a	4%

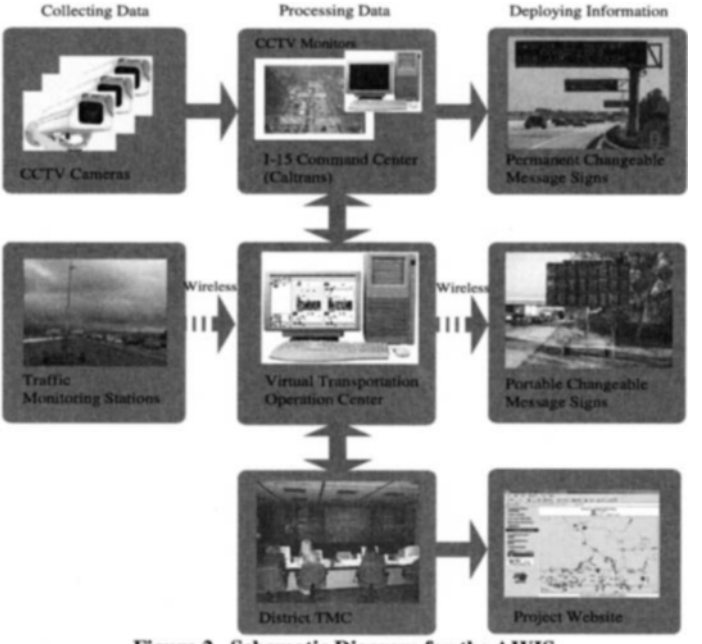


Figure 2. Schematic Diagram for the AWIS

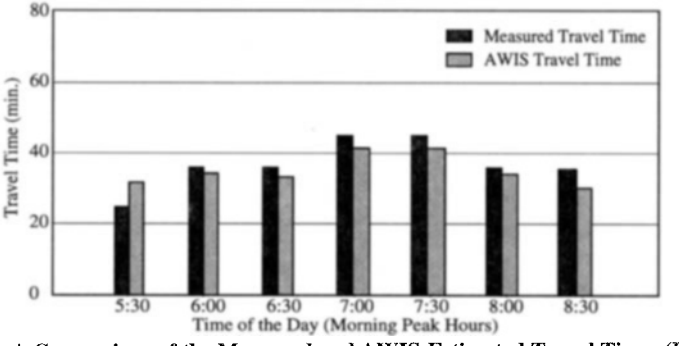


Figure 3. A Comparison of the Measured and AWIS-Estimated Travel Times (I-15 SB, 27km, October 26, 2004)

Conclusions

The I-15 Devore Project successfully rehabilitated 9 lane-km of deteriorated truck lanes within two 9-day extended closures without causing a significant traffic delay, contrary to the initial concerns of the public and of Caltrans. AWIS was successfully implemented and achieved a 20-percent reduction in peak hour demand during construction. The conclusions of this case study are:

1. The I-15 Devore AWIS, in fully automated mode, provided travel time estimates that were close to the actual travel times measured by probe vehicles.
2. According to the traffic measurement study with before- and during-construction data, the ADT volumes on the I-15 CWZ corridor decreased by up to 19 percent for SB and 16 percent for NB and the ADT volumes on alternate roads increased by up to 15 percent.
3. AWIS operation successfully effectuated traffic diversion to alternate roads. It contributed to a decreased traffic delay within the CWZ by half, from an expected 95 minutes to an actual 50 minutes during weekday peak hours. It is estimated that the reduction in traffic demand yielded a road-user cost-savings benefit of approximately $3.8 million.

AWIS operation on I-15 Devore demonstrated that such a program could dramatically reduce traffic demand within the CWZ corridor. The implementation details and guidelines presented in this paper will help practitioners and transportation agencies efficiently configure the AWIS in future urban freeway rehabilitation projects. The outcomes of this study offer potential cost-saving travel benefits for both individual road users and for the public sector by demonstrating ways to lower the cost of travel and the cost of highway reconstruction.

References

Fontaine, M.D. 2003. "Guidelines for Application of Portable Work Zone Intelligent Transportation Systems." *J. of Transportation Research Board,* 1824: 15-22.

Lee, E.B., Thomas, D. K., and Bloomberg, L. 2005. "Planning Urban-Highway Reconstruction with Traffic Demand Affected by Construction Schedule." *J. of Transportation Engineering,* ASCE, 131(10): 752-761.

PeMS. "Freeway Performance Measurement System (PeMS)." July 30, 2004. University of California, Berkeley and California Department of Transportation. *http://pems.eecs.berkeley.edu/Public/index.phtml.*

Tooley, M.S., Gattis, J.L., Janarthanan, R., and Duncan, L.K. 2004. "Evaluation of Automated Work Zone Information Systems." *J. of Transportation Research Board,* 1877: 69-76.

Transportation Research Board (TRB). 2000. "Highway Capacity Manual (HCM)." National Research Council: Washington, D.C.

Modeling Tour Mode and Destination Choice in the Jakarta Metropolitan Area

S. Yagi[1] and A. Mohammadian[2]

[1]Dept. of Civil and Materials Engineering, University of Illinois at Chicago, 842 W. Taylor St., Chicago, IL 60607; PH (312) 413-7447; FAX (312) 996-2426; email: syagi1@uic.edu
[2]Dept. of Civil and Materials Engineering, University of Illinois at Chicago, 842 W. Taylor St., Chicago, IL 60607; PH (312) 996-9840; FAX (312) 996-2426; email: kouros@uic.edu

Abstract

This paper presents a work tour mode and destination choice model within an activity-based modeling framework. The case study is Jakarta, Indonesia as one of the largest metropolitan areas in Asia. The dataset obtained from a large household travel survey provides daily travel patterns and detailed information on household socio-demographic characteristics. Eight most commonly used combinations of travel modes observed in the region are considered: drive alone, shared ride, motorcycle, taxi, motorcycle taxi, transit with motorized access, transit with non-motorized access, and non-motorized transport. In addition, eleven representative destinations are also considered for each tour and are sampled from the 336 traffic analysis zones using the stratified importance sampling method which is based on the distance as well as total jobs as an indicator of the magnitude of attraction in the destination. A wide variety of different types of variables contributed significantly to the model, including variables related to trips, activities/tours, households, individuals, and destination zones. The modeling results suggest that, in the context of Jakarta, choice alternatives, structure of the model, and key variables are quite different from those observed in the developed world. This study is part of a larger effort with ultimate goal to develop a comprehensive activity-based microsimulation modeling system for developing countries.

Introduction

Japan International Cooperation Agency (JICA) conducted "The Study on Integrated Transportation Master Plan (SITRAMP)" (National Development Planning Agency 2004) in the Jakarta Metropolitan Area from November 2001 to March 2004. The overall objective was to identify possible policy measures and solutions to develop a sustainable transportation system in the Jakarta Metropolitan Area with a focus on encouraging public transport usage and improving mobility of people. As such, detailed transportation surveys and analyses were undertaken to prepare a comprehensive long-term transportation plan with the objective to develop and calibrate disaggregate travel demand models to simulate present and future interactions between socio-economic distribution and transportation in the region.

The Household Travel Survey (HTS), among a variety of the surveys conducted, provides the largest and most comprehensive travel data in the region. The dataset covers as many as 166,000 households which correspond to three percent of the entire population, and provides daily travel patterns and detailed information on household socio-demographic characteristics. The ultimate goal of the authors' research on Jakarta data is to develop a novel activity-based microsimulation modeling system to test different transportation policy scenarios for the Jakarta Metropolitan Area. The study will simulate the way individuals schedule their daily activities and travel in an urban region of the developing world. It is hoped that the new

model will contribute to the improvement of the emerging travel demand forecasting techniques and provide an opportunity to evaluate urban transportation policy scenarios.

The activity-based modeling structure used in this study is a system of random utility based disaggregate logit and nested logit models assuming a hierarchy of the model components, with three types of major models, namely, choices of daily activity patterns, times of day, and mode and destination. As such, mode and destination choice is placed at the "bottom" of the hierarchy, and is conditional on decisions at the higher level. This paper presents a work tour mode and destination choice model within an activity-based modeling framework.

Modeling Structure

The main purpose of the study was to estimate a model of joint choice of mode and destination for work tours. The modeling approach is a discrete choice model based on the random utility maximizing principles. It has been shown that the multinomial logit model is the most popular form of discrete choice model in practical applications (Mohammadian and Doherty 2005). Nested logit model, which has been utilized in this study, is a model that has been developed in order to overcome the so-called independence of irrelevant alternatives (IIA) limitation in the multinomial model by modifying the choice structure into multiple tiers. Nested logit models are very commonly used for modeling mode choice, permitting covariance in random components among nests of alternatives. Alternatives in a nest exhibit an identical degree of increased sensitivity relative to alternatives in the nest (McFadden 1978). A nested logit model has a log-sum or expected maximum utility associated with the lower-tier decision process. The parameter of the log-sum determines the correlation in unobserved components among alternatives in the nest (Daganzo and Kusnic 1993). The range of this parameter should be between 0 and 1 for all nests if the nested logit model is to remain globally consistent with the random utility maximizing principle.

The results of the HTS survey show that more than 90 percent of people return home using the same mode as they used for the from-home trips, though the percentages vary depending on modes and purposes. This suggests that from-home trips constrain the modes and destinations of the subsequent segments such as returning-home trips. Therefore, in this study, from-home trips are focused on and used to estimate the key module of the entire work tour mode and destination choice model.

Alternative Settings

Eight most commonly used combinations of travel modes observed in the region are considered. These include auto drive alone, auto shared ride, motorcycle, taxi, motorcycle taxi, transit with motorized access, transit with non-motorized access, and non-motorized transport. As for the destination choice, eleven representative destinations are considered for each tour in order to reduce the computational burden. These destinations are sampled from the 336 traffic analysis zones using the stratified importance sampling method. The strata of destinations are constructed based on the distance as well as a size variable which indicates the magnitude of attraction in the destination, that is, total jobs for work tours. As a result, this sampling method leads to higher probabilities of being selected for zones closer to the origin as well as for zones with higher job accumulation. Actual sampling strata for these 11 representative destination zones are as follows:

- Zone 1, "sampled" from the origin zone;
- Zones 2 and 3, sampled from a distance less than D_1;
- Zones 4 and 5, sampled from a distance between D_1 and D_2 and total jobs less than J;

- Zones 6 and 7, sampled from a distance between D_1 and D_2 and total jobs greater than J;
- Zones 8 and 9, sampled from a distance greater than D_2 and total jobs less than J;
- Zones 10 and 11, sampled from a distance greater than D_2 and total jobs greater than J,

where:

D_1 is the 20th percentile distance from the origin zone to all other work tour destinations;
D_2 is the 60th percentile distance from the home zone to all other work destinations; and
J is the 50th percentile total jobs of all actual work tour destinations.

While the value of size variable, J, stays the same regardless of the origin zones, the values of distance, D_1 and D_2, are different depending on the origin location. Hence, the composition of the above sampling strata for destination choice is also different by the origin zone.

As this is a joint model of mode and destination choice, total number of choice alternatives is presumed to be 88 (i.e., total number of modes multiplied by total number of destination zones). However, frequencies of each mode in relation to the destination strata were investigated. Due to the extremely low frequencies observed from the dataset, it is assumed that motorcycle taxi alternatives are unavailable if the distance to the destination zone is greater than D_2, and non-motorized transport alternatives are unavailable if the distance is greater than D_1. Thus, the maximum number of available alternatives is reduced to 76. Auto (drive alone) alternatives are made unavailable for individuals under 17 (i.e., pre-driving age) and for those who do not have access to any automobiles as indicated in the survey. Additionally, motorcycle alternatives are made unavailable for those who do not have access to any motorcycles.

In the nesting structure, auto drive alone, auto shared ride, and motorcycle; and taxi, motorcycle taxi, transit with motorized access, and transit with non-motorized access are each placed in the second tier under different nests while non-motorized transport is placed as a degenerate branch. Although nests are created for each representative destination zone, log-sum parameters are set to be common for the nests which involve the same mode group.

Explanatory Variables

While the majority of variables have been directly derived from the HTS data, some other zone-based data have also been made available such as generalized travel time computed from the preliminary highway and transit network assignment, and land use composition in each traffic analysis zone computed from the GIS database. Thus, a number of types of variables are tested for modeling. In brief, explanatory variables are classified as follows:

- Trip related variables: generalized travel time, transit walk time, and travel distance;
- Activity/tour related variables: intermediate stop(s), primary/secondary tour, and start times of tours and return trips;
- Household related variables: household income, household member composition, and auto and motorcycle ownership;
- Individual related variables: full/part-time worker, personal income, gender, age, and commuting allowances provided by the employer; and
- Destination zone related variable: identity of origin and destination zones, fraction of land for business use, and service industries job density.

In addition, natural logarithm of the size variable (i.e., total jobs) is included as a destination zone related variable. The coefficient is considered as the scale parameter (Ben-Akiva and Lerman 1985), and the value was estimated as 0.82 in the preliminary model estimation. Following the Ben-Akiva and Lerman study, this sale parameter has been constrained to 1 in the final model. However, this had little effect to the values of coefficients of other variables.

Modeling Results

Results of the mode and destination choice model for from-home trips that is estimated for work tours are summarized in Table 1. The model shows a good fit with the adjusted rho-squared value of 0.38. The log-sum parameters are estimated as 0.73 and 0.69 for private modes and taxi and transit modes respectively, staying within a reasonable range.

Among several types of cost and time-related variables, so-called generalized travel time proved to work best in the model. It is computed from the preliminary network assignment highway or transit network by origin-destination zone pair and by mode, including not only travel times (in-vehicle time and waiting time in the case of transit) but also times that have been converted from monetary costs such as transit fares and highway tolls. Three kinds of time values are used to calculate the generalized time. These are for low, middle, and high income households, which were estimated through a stated preference survey conducted in SITRAMP. Coefficients of the generalized time are estimated separately for auto, motorcycle, taxi, and motorcycle taxi and transit. While the coefficients for auto and taxi have higher absolute values, the coefficient for motorcycle shows the lowest sensitivity to the generalized time. This result seems reasonable because auto and taxi are generally used by middle to high-income people and motorcycle is used by low to middle-income people, as indicated by other income-related variables included in each mode. On the other hand, the coefficient for motorcycle taxi and transit stays somewhere in between, and relatively few income- related variables are included in the utility functions.

The generalized travel time also works as one of the variables that determine the utilities for destination choice. So does natural logarithm of the above-mentioned size variable (i.e., total jobs). As for other destination-related variables, the origin zone dummy has a very high *t*-value, increasing the utility for intra-zonal work tours. Fraction of land for business works positively in the utilities for destination choice, while service industry job density works in the other way around.

In general, the model has captured the key significant variables, including not only trip attributes such as cost, time, and distance of travel, but also various socioeconomic attributes of both households and individuals, some of which are implying interactions among household members. Furthermore, tour or activity-related variables such as existence of intermediate stops, distinction of primary or secondary tours, and start times of the tour or returning segment of the tour also play significant roles in the model. As compared to similar studies conducted for the U.S. cases (Bradley, et al. 1997), the model gives the following characteristic implications in the context of Jakarta:

- Gender plays a more active and distinct role in the model. That is, males have higher utilities of private modes, while females have higher utilities of taxi and transit.
- Income also has a greater influence on mode choice. Generally, utilities of auto and taxi increase as the income becomes higher, while utilities of motorcycle, transit (with non-motorized access), and non-motorized transport increase as the income becomes lower.
- Car competition between household members, which often has an influence on mode choice in the U.S., is not an "issue" in the case of Jakarta. The auto ownership ratio is still low in Jakarta, and whether the household owns an auto or motorcycle is more significant for mode choice rather than actual number of vehicles owned.
- A variety of commuting allowances are sometimes provided by the employer. Availability of such allowance mainly increases utilities of private modes.
- In Jakarta, auto shared ride more often indicates those who do not actually drive but have chauffeurs. As a result, it is characterized as a mode for high-income people.

Table 1 Home-Based Work Tour Mode/Destination Choice Model

Generalized Travel Time (hr)	**coeff.**	***t-stat***
for auto	-1.923	*-16.1*
for motorcycle	-1.184	*-25.4*
for taxi	-1.957	*-6.4*
for motorcycle taxi and transit	-1.640	*-37.0*

Destination Land Use	**coeff.**	***t-stat***
Dummy: origin zone	1.962	*57.9*
Service job density (/ha)	-0.064	*-16.9*
Percentage of land for business use	0.065	*7.1*
Log of size variable (total jobs)	1.000	*constr.*

Alternative / Variable		
Auto Drive Alone		
Alternative-specific constant	4.185	*5.2*
Log of travel distance (km)	0.986	*16.6*
Dummy: tour starts in a.m. peak	0.839	*7.0*
Dummy: tour starts in p.m. peak or later	0.468	*2.8*
Dummy: one-member household	0.881	*2.4*
Dummy: car-owning household	2.639	*15.6*
Log of age of the individual	2.136	*10.1*
Dummy: male individual	1.672	*12.9*
Log of monthly ind. income (mil. Rp)	0.340	*4.1*
Dummy: toll allowance provided	0.823	*2.1*
Dummy: free parking provided	0.662	*2.3*
Dummy: private mode allowance provided	2.434	*13.1*
Auto Shared Ride		
Alternative-specific constant		
Log of travel distance (km)	0.958	*15.8*
Dummy: intermediate stop on way to work	2.296	*11.5*
Dummy: tour starts in a.m. peak	0.509	*4.4*
Dummy: tour starts in p.m. peak or later	1.279	*3.1*
Dummy: return trip starts in late night	0.421	*2.5*
Log of monthly hhd income (mil. Rp)	0.726	*8.7*
Dummy: car-owning household	4.806	*32.5*
Log of age of the individual	2.793	*14.6*
Dummy: toll allowance provided	1.645	*3.9*
Dummy: private mode allowance provided	2.491	*13.2*
Motorcycle		
Alternative-specific constant	15.584	*21.0*
Dummy: tour starts in a.m. peak	0.693	*9.2*
Dummy: infant (age < 5) in household	0.363	*3.6*
Dummy: child (age:5-17) in household	0.264	*3.8*
Log of monthly hhd income (mil. Rp)	-0.724	*-11.6*
Dummy: motorcycle-owning household	1.692	*16.1*
Dummy: male individual	1.675	*17.7*
Dummy: allowance provided	-1.907	*-12.3*
Dummy: free parking provided	0.761	*2.1*
Dummy: private mode allowance provided	3.739	*18.1*

Taxi		
Alternative-specific constant	12.092	*8.5*
Dummy: return trip starts in late night	0.720	*3.0*
Dummy: high-income hhd (>4 mil. Rp/mo)	1.675	*6.7*
Log of age of the individual	1.032	*3.2*
Dummy: male individual	-1.549	*-7.3*
Dummy: full-time worker	-0.459	*-1.8*
Motorcycle Taxi		
Alternative-specific constant	16.363	*19.4*
Dummy: tour starts in a.m. peak	0.645	*6.6*
Dummy: child (age:5-17) in household	0.294	*3.5*
Dummy: male individual	-0.809	*-7.8*
Dummy: part-time worker	0.433	*3.7*
Transit w/ Motorized Access		
Alternative-specific constant	16.411	*19.6*
Log of travel distance (km)	1.221	*31.6*
Dummy: tour starts in p.m. peak or later	-1.645	*-3.6*
Dummy: male individual	-0.903	*-11.2*
Dummy: private mode allowance provided	0.400	*5.4*
Transit w/ Non-Motorized Access		
Alternative-specific constant	19.071	*22.8*
Transit walk time (hr)	-1.009	*-12.8*
Log of travel distance (km)	0.957	*31.7*
Dummy: intermediate stop on way to work	-0.480	*-2.8*
Dummy: tour starts in p.m. peak or later	-1.170	*-5.3*
Dummy: male individual	-0.710	*-11.3*
Log of monthly ind. income (mil. Rp)	-0.531	*-10.1*
Non-Motorized Transport		
Alternative-specific constant	5.965	*10.4*
Log of travel distance (km)	-0.431	*-13.4*
Dummy: tour starts in early morning	-0.823	*-17.4*
Dummy: return trip starts in late night	-0.449	*-7.0*
Dummy: child (age:5-17) in household	0.200	*5.1*
Log of age of the individual	1.069	*17.8*
Dummy: part-time worker	0.641	*12.9*
Log of monthly ind. income (mil. Rp)	-2.408	*-29.3*
Dummy: allowance provided	-1.197	*-13.0*

No. of Observations = 23296	**Logsum:** expected maximum utility	
$L(0)$ = -96955	from private modes	0.735 *78.8*
$L(\beta)$ = -60357	from taxi and transit modes	0.690 *62.8*
ρ^2 = 0.377		

Summary

This paper presented a work tour mode and destination choice model that is developed as part of the comprehensive activity-based modeling system for Jakarta Metropolitan Area. A wide variety of different types of variables contributed significantly to the model, including variables related to trips, activities/tours, households, individuals, and destination zones. Interpretation of the effects of these explanatory variables in the developed model led to several interesting insights. Furthermore, the model has shown that, in the context of developing countries, structure of the model, choice alternatives, and key variables are significantly different from those observed in the developed world. This study is part of a larger effort with an ultimate goal to develop a comprehensive activity-based microsimulation modeling system for developing countries. The results of the study can be used to provide accurate estimates of travelers' behavior which are expected to serve as better inputs for evaluation of different transportation policy scenarios for the region.

Mode and destination choice models for other tour purposes (i.e., school, maintenance, and discretionary activities) have also been estimated. Furthermore, the comprehensive mode and destination choice models include two more modules: one for estimating mode and destination choice for work-based sub-tours, and the other for estimating intermediate stop location choice. For the work-based sub-tour mode and destination choice, the methodology is similar to that for the above-mentioned mode and destination choice of from-home trips, except that the mode used to travel from home to work is a highly significant variable in the work-based sub-tour choice model. Meanwhile, for the intermediate stop location choice, the stratified importance sampling method is adopted again with different strata of stop locations, and separate models are estimated by mode and by purpose of the tour.

Acknowledgement

The authors would like to express their gratitude to the National Development Planning Agency of Indonesia for permitting us to use the survey data in this study. This study was partially funded by Foundation for Advanced Studies on International Development (FASID).

References

National Development Planning Agency (BAPPENAS), Republic of Indonesia, and Japan International Cooperation Agency (JICA) (2004). *The Study on Integrated Transportation Master Plan for Jabodetabek (Phase 2)*, Final Report, prepared by Pacific Consultants International and ALMEC Corporation.

Mohammadian, A., and S.T. Doherty (2005). *Mixed Logit Model of Activity Scheduling Time Horizon Incorporating Spatial-Temporal Flexibility Variables.* Transportation Research Record, No. *1926,* TRB, National Research Council, Washington D.C., pp. 33-40.

McFadden, D. (1978). *Modeling the Choice of Residential Location.* Transportation Research Record, Vol. 672: 72-77.

Daganzo, C.F., and M. Kusnic (1993). *Two Properties of the Nested Logit Model.* Transportation Science, Vol. 27: 395-400.

Ben-Akiva, M. and S.R. Lerman (1985). *Discrete Choice Analysis: Theory and Application to Predict Travel Demand,* Cambridge, Mass.: MIT Press.

Bradley, M., J.L. Bowman, Y. Shiftan, K. Lawton, and M.E. Ben-Akiva (1997). *A System of Activity-Based Models for Portland, Oregon: A Demonstration Project for the FHWA Travel Model Improvement Program (TMIP),* Washington D.C.: U.S. Department of Transportation, December.

An Exploratory Analysis of the Household Travel Behavior and Lifestyle Choices

Y. Zhang[1] and A. Mohammadian[2]

[1]Dept. of Civil and Materials Engineering, University of Illinois at Chicago, 842 W. Taylor St., Chicago, IL 60607; PH (312) 413-7447; FAX (312) 996-2426; email: yzhang31@uic.edu
[2]Dept. of Civil and Materials Engineering, University of Illinois at Chicago, 842 W. Taylor St., Chicago, IL 60607; PH (312) 996-9840; FAX (312) 996-2426; email: kouros@uic.edu

Abstract

It has been shown that the lifestyle is a key factor in shaping household daily activity schedule, travel pattern, and trip generation. Therefore, the analysis of the choice of lifestyle is an important element in explaining household travel behavior and to address a wide range of policy issues related to transportation equity, mobility, and accessibility. This paper aims to explore typical household lifestyles in the US that are transferable to other urban areas for travel demand forecasting. Several variables that appeared to have strong relationships with the lifestyle choices were used in a principal component analysis. Using a TwoStep clustering schema, extracted factors were employed to identify typical lifestyles that shape the observed household activity patterns. Seven distinguishable clusters were identified that present national level standards of living with respect to travel behavior. Furthermore, exploring lifestyles, socio-demographics, land-use variables, and graphical representation of the household location facilitated the process of explaining each lifestyle more precisely.

Introduction

It has been shown that the choice of lifestyle has a direct impact on travel behavior, activity pattern, and trip generation. Transportation analysts are interested in identifying various household lifestyles and their attributes. Krizek and Waddell (2002) clustered the seventh wave (year 1997) of the Puget Sound Transportation Panel (PSTP) data into 9 clusters, which demonstrated distinguishable lifestyles including: Retirees, Single busy urbanists, Elderly homebodies, Urbanists with higher income, Transit users, Suburban errand runners, Family- and activity-oriented participants, Suburban workaholics, and Exurban, family commuters. They clustered 1907 households using five factors generated from variables reflecting travel patterns, activity participation, vehicle ownership and urban form/residential location. While this study offers insight into various lifestyles, its geographical focus has made its applications limited. Zhao et al. (2003) explored the effectiveness of using lifestyle models in determining seasonal resident trip production in Florida. Schwanen and Mokhtarian (2005) analyzed data from commuters in three neighborhoods in the San Francisco Bay area to assess the impacts of residential neighborhood type on distance traveled and travel modes. They also examined the relationships between distance traveled and factors such as attitudes, personality, lifestyle, and mobility limitations.

The analysis of the lifestyle has also very important implications for policymakers as they are interested to address transportation equity issue and to balance mobility, accessibility, and traffic safety for elderly or underrepresented groups. Hildebrand (2003) used lifestyle groups in a simplified activity-based framework to model travel characteristics and behaviors of the elderly.

Furthermore, there are several private companies (e.g., Claritas, ESRI BIS, Melissa Data, etc.) that have developed detailed national level segments of the people for commercial purpose. There are also several information inquiry systems available, by which users can easily get the information of the segments for their zip code. However, these segments are business-based and aim to target potential customers for businesses. It is important to develop a national level lifestyle model that can explain household travel behavior and address a wide range of transportation policy issues. This paper aims to explore typical household lifestyles that are transferable to other US urban areas for travel demand forecasting. This study attempts to cluster about 70,000 household records from the 2001 National Household Travel Survey (NHTS) dataset and to explain their household lifestyles and travel characteristics.

Data Sources

In order to establish a relationship between travel behavior, household characteristics, and land-use attributes, exploratory regression analysis was performed that resulted in the selection of explanatory variables in this study. These variables are in individual, household, and census tract levels. Demographics and socio-economics and land use-related variables were extracted from NHTS and other sources. Pedestrian environment variables (e.g., road density, intersection density, and block size) are estimated by overlapping street network and census tract shape-files in a GIS environment. A transit usage measure has been introduced by matching the proportion of the transit users to the total number of workers in the census tract using the CTPP 2000 dataset. Finally, a congestion factor was extracted from the Travel Time Index Tables of the Urban Mobility study (TTI 2004) and used as proxy to present the congestion level that a typical household experiences at the MSA level.

Analysis and Findings

Principal Component Analysis

Principal component analysis is used to avoid problems associated with the multi-collinearity in the dataset and to reduce number of variables. Table 1 presents the results of the factor analysis. There are totally six factors extracted explaining nearly 72% of the variation in the dataset. All of the eigenvalues of these factors are larger than 1. In order to understand these factor better, each factor is named according to the loading of its variables. A simple comparison of these factors with those presented in Krizek and Waddell (2002) work proves both similarities and differences. Factor analysis of the Puget Sound dataset has resulted in five factors, which were "urban form", "non-work activities", "environmentally benign travelers", "work commute" and "auto-dependent travelers". While broad, these factors do not clearly characterize land use and infrastructure densities. In our study only the factor representing infrastructure densities is called *urban form*. The *auto-reliant non-mandatory activities* can be compared to "non-work activities". However, this study did not find a high correlation between carpool usage and non-mandatory trips as shown in the PSTP study. Furthermore, *mandatory trips* in the current study are comparable to the "work-commute" factor. It can also be shown that the *Non-auto travelers* and *Travel less* factors in the current study are comparable to the PSTS's "environmentally benign travelers" and the inverse of the "auto-dependent travelers" factors.

Cluster Analysis

Several cluster analysis techniques were examined to assign households into homogeneous groups. Finally a TwoStep clustering method was employed as it is efficient, converges fast

TABLE 1 Factor Analysis of Each Lifestyle Dimension

Household Measures	Factors					
	1 Auto-reliant non-mandatory activities	2 Land use factors	3 Urban form	4 Mandatory trips	5 Non-auto users	6 Travel less
No. of private vehicle trips	**0.871**	-0.101	-0.035	0.322	-0.105	-0.069
No. of tours	**0.840**	-0.028	0.003	0.385	0.090	0.127
No. of discretionary trips	**0.788**	0.000	-0.058	0.005	-0.002	-0.041
No. of maintenance trips	**0.768**	-0.016	0.011	0.073	-0.023	-0.053
Employment density-CT level	-0.032	**0.952**	0.221	-0.056	0.122	0.030
Residential density- CT level	-0.034	**0.951**	0.214	-0.069	0.125	0.034
Population density- CT level	-0.036	**0.926**	0.259	-0.066	0.167	0.037
Road length/Area -CT level	-0.022	0.314	**0.853**	-0.119	0.084	0.091
No of intersections/Area-CT level	-0.029	0.313	**0.853**	-0.114	0.098	0.088
Road length/No. of intersections in CT	-0.015	-0.009	**-0.801**	0.098	0.004	-0.060
Travel time index	-0.037	0.134	**0.616**	0.060	0.117	-0.085
No. of drivers	0.205	-0.062	-0.082	**0.823**	-0.096	-0.088
No. of vehicles	0.071	-0.101	-0.186	**0.745**	-0.153	-0.188
No. of household members	0.463	-0.006	-0.002	**0.634**	0.125	0.084
No. of mandatory trips	0.164	-0.038	0.012	**0.602**	0.172	0.314
No. of transit users in the HH	0.001	0.168	0.121	-0.001	**0.931**	0.012
No. of transit trips in the HH	-0.018	0.176	0.126	-0.035	**0.926**	0.003
Sum of trip miles for each HH	0.302	0.016	-0.032	0.169	0.082	**-0.606**
No. of non-motorized trips	0.345	0.179	0.061	-0.067	0.199	**0.581**
No. of carpool users in the HH	-0.021	-0.009	0.014	0.281	-0.041	**0.397**

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
Rotation converged in 6 iterations. CT=Census Tract. Densities are measured in thousands per square miles.
Travel Time Index=Travel time at peak period/travel time at free speed, provided by Texas Transportation Institute.

and has the advantage of automatically identifying the optimal number of clusters using a Bayesian information criterion (Chiu et al. 2001). Our observations in this study confirm the notation in the Krizek and Waddell work that too few clusters would have difficulty in identifying groups while too many clusters "provide diminishing returns in terms of variance explained". Various cluster numbers ranging from 5 to 12 were explored to examine other potential alternatives. Finally a seven cluster scheme was adopted as it identified distinguishable clusters that present national level standards of living with respect to travel behavior.

Assessment of Results

After examining cross-tabulations of variables for each clustering scheme, it became evident that clustering scheme with seven clusters is the preferred alternative. This is due to the fact that all seven clusters can be easily justified and both the values of each variable and GIS location distribution (See Figure 1) make sense. The level of the importance of each variable can be used to explain the significance of a factor in defining a certain cluster.

To get a better understanding of each lifestyle, cross-tabulations are prepared for each lifestyle versus all of the available socio-demographic variables and the original variables that the six factors are extracted from. Each lifestyle is then labeled according to the values of its variables and also its GIS location distribution. These labels are only used to help identifying the nature of each lifestyle and summarized as follows.

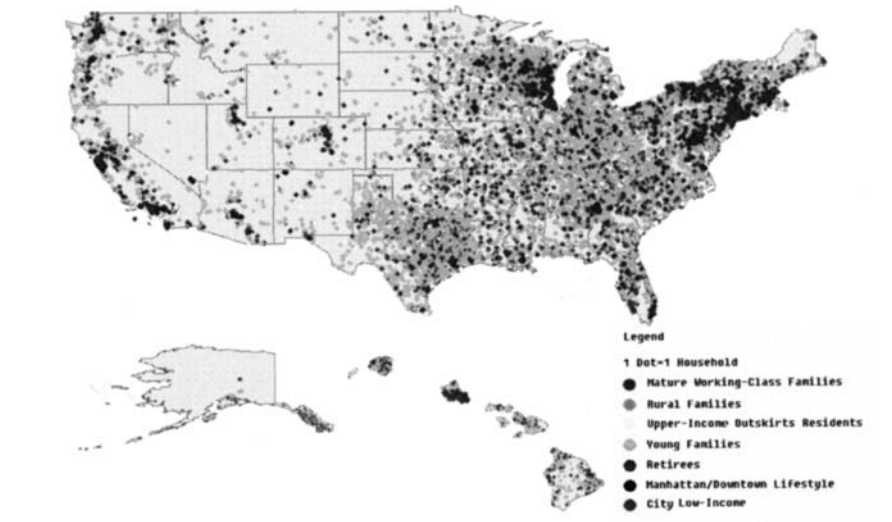


FIGURE 1 Cluster Distributions

Mature working-class families: This lifestyle covers 14% of households in the dataset. They have the biggest household size (3.7), largest percentage of mandatory trips per person (91%), the highest percentage of workers (65%), the lowest percentage of elderly aged 65 and more (3%), the highest percentage of persons aged 16-24 (16%) and 45-54 (18%). They do not use public transit regularly but use carpool most frequently (9% of the household members). They have almost the same average travel distance relative to the whole nation.

Rural families: This lifestyle presents the largest group (26%) in the nation. A large percentage of these households live in rural areas (76%) and 23% in suburban areas. They are of middle size (2.50) but have the highest automobile ownership rate (95%) and highly rely on their cars. However, their travel distance is just slightly higher than the national average.

Upper-income outskirts residents: Household members of this group comprise 17% of the households in national level dataset. This lifestyle presents the richest group with an average income of 4.1 times higher than the poverty level. These are second smallest households with only 2.1 members. Their neighborhoods include outskirts of downtowns. They have very good accessibility (higher road infrastructure density that are not highly congested). They do not tend to use transit and almost never use carpool.

Young families: This group characterizes 16.5% of households and presents a similar location distribution as the mature families. They have more children and almost 62% of them are households with two adults and kids younger than 15 years old. The percentage of children younger than 15 is the highest (35%) and that of people older than 65 is only 6%.

Retirees/Empty nesters: This cluster represents almost 20% of the population with 47% of them being retirees without children. The have the smallest number of household members (1.4), mandatory trips and workers. They also travel shorter distances and highly rely on automobiles. Their average income level is only 2.8 times higher than the poverty level.

Manhattan/Downtown Lifestyle: The group presents the smallest cluster size (2%) with average land use (population, employment and housing) densities that are much higher than other clusters. This lifestyle seems to be more stand-alone than any other cluster. It has the highest percentage of walk and bike usage (1.4 trips per person per day), transit usage (0.53 trips per person per day), lowest vehicle ownership (0.33/person), shortest travel distance, smallest percentage of private vehicle trips, and the most diversified population (only 44% of them are white).

City low-income: Household members of this cluster (almost 5% of the population) tend to live in the outskirts of downtowns, similar like upper-income outskirts residents. However, they have larger household size (2.9), much lower income (almost the same as the retirees), and highest percentage of transit usage (even much higher than the downtown residents). Their automobile ownership rate is very low (0.49/person) but they tend to travel longer distances (higher than the national average). They have the largest percentage of African American population (17.6%) and a high percentage of Asian and Hispanic population.

It is interesting to examine the relationship between the lifestyles and their constitution of different segments of the household lifecycles. Young singles or couples without children tend to live in dense urban area so that they can save time from long commute to work and other activities. As they have limited income, they tend to rent an apartment. As they move to the next stages of their lives, form families, and have children, they tend to need more space and better environment for the kids, so they make decisions to move to suburb, and become young families living in detached houses. They probably turn out to be are very busy, both in work and in taking care of their children. As a result, they tend to travel longer distances with higher frequencies. This lifestyle requires them to drive their kids to many activities. When the children grow up, these young families will change their lifestyle and become mature families with more workers, they will have more mandatory trips and higher income. The new generation gets employed, leaves the parents' house, moves to a dense urban area, and finds a partner. Those mature families move to their next lifecycle -empty nesters- where only retirees or active adults form the household.

Summary and Conclusion

The analysis of the choice of lifestyle is an important element in explaining household travel behavior. This has also very important implications for policymakers as they are interested to address transportation equity issue and to balance mobility, accessibility, and traffic safety for elderly and underrepresented groups. Using the 2001 National Household Travel Survey (NHTS) dataset, principal component analysis and a TwoStep clustering scheme were employed to identify the seven most distinguishable lifestyles. In order to explore each category more precisely, cross-tabulations of the lifestyles versus socio-demographics, land-use variables, and location of the household were examined. This study offers a detailed lifestyle model that can relate travel behavior to various stages of the household lifecycle from early formation to the retirement ages. This can be used in many applications including travel demand forecasting and urban transportation policy analysis.

The effects of the neighborhood characteristics were also investigated in this study. An important finding of the analysis was the study of the Manhattan lifestyle as a stand-alone cluster among other groups. This cluster can speak to the power of the built environment. It is evident that given a chance to act differently, travelers will adapt and behave differently. It was also shown that only Manhattan lifestyle and city low-income have high transit usage. This implies that both very high land-use densities and poor economic attributes will result in

higher public transit use. By contrast, upper-income outskirts residents whose environment has similar accessibility and land use densities almost never use public transit. These findings can have many important implications in urban design and transportation planning.

This study is part of a larger effort to evaluate travel behavior and lifestyle choices with application in travel data transferability. It is assumed that households with similar lifestyle and socio-economic characteristics have similar travel behavior. This will probably allow us to transfer travel data from a national level data sources (e.g., NHTS) to a small region where travel survey data is not available. Further research in this area is underway to examine the transferability of the travel related characteristics as function of the household lifestyle to other geographical regions.

Acknowledgement

The authors would like to express their gratitude to the Federal Highway Administration (FHWA) for supporting this study.

References

Chiu, T., D. Fang, J. Chen, Y. Wang, and C. Jeris, (2001). A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment. *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining.*

Claritas, PRIZM NE Social Group, http://www.claritas.com , last accessed July, 2005.

ESRI Business Information Solutions, Community Tapestry: The Fabric of America's Neighborhoods, http://www.esribis.com/tapestry, last accessed July, 2005.

Federal Highway Administration (FHWA), (2005), National Household Travel Survey (NHTS), version 4 datasets, http://nhts.ornl.gov/2001/index.shtml, last accessed July, 2005.

Hildebrand, E.D. (2003), Dimensions in Elderly Travel Behaviour: A Simplified Activity-Based Model using Lifestyle Clusters, Transportation, 30(3), pp 285-306.

Krizek K.J. and P. Waddell, (2002). Analysis of Lifestyle Choices: Neighborhood Type, Travel Patterns, and Activity Participation, Transportation Research Record 1807, 119-128.

Melissa Data, Prizm Code Descriptions, http://www.melissadata.com/Lookups/PrizmDes.asp, last accessed July, 2005.

Schwanen, T., and P.L. Mokhtarian (2005). What if You Live in the Wrong Neighborhood? The Impact of Residential Neighborhood Type Dissonance on Distance Traveled, Transportation Research, Part D: Transport and Environment, 10 (2), pp 127-151.

Texas Transportation Institute, (2004), Urban Mobility Study - Key Mobility Measures, http://mobility.tamu.edu/ums/congestion_data/tables/national/table_1.pdf.

U.S. Census Bureau, Poverty Thresholds 2001, http://www.census.gov/hhes/poverty/threshld/thresh01.html, accessed Jul, 2005.

Zhao, F., L-F Chow, M-T Li, and A. Gan. (2003), Development of FSUTMS Lifecycle and Seasonal Resident Trip Production Models for A. Florida Urban Areas, Final report to the Florida Department of Transportation, Department of Civil and Environmental Engineering, Florida International University, 172 pages.

Development and Analysis of a an Internet-based Travel Survey

Michael Anderson[1], Jose Ortega, Sampson Gholston and Steven Jones

[1] Department of Civil and Environmental Engineering, The University of Alabama in Huntsville, Huntsville, AL 35899; PH (256) 824-5028; FAX (256) 824-6724; email: mikea@cee.uah.edu

Abstract

The increasing development in technological capabilities and resources is pushing the transportation industry to implement new and better technologies to improve techniques for collecting travel behavior data for transportation planning applications. Travel surveys are the most common procedure used for this task and the development of more efficient ways to gather data is a major concern for transportation agencies. Improvement of data collection methods would produce great benefits for policy makers and taxpayers because of potential savings in time, cost and resources. The ever-increasing availability of the Internet creates an opportunity for a potentially more efficient, less costly means of collecting travel behavior data. An efficient Internet-based travel survey has the potential of being a very effective means of collecting data for inclusion into transportation planning models (e.g., trip generation, distribution). This project examines the usefulness of an Internet survey through a pilot study in Huntsville, AL. The study specifically addresses the data (types, range, etc.) obtained from the Internet travel survey and efficacy of applying the collected data for transportation planning applications.

Background

As the world entered the 21st century, society's necessity to use more accurate techniques to be able to predict future situations increased. In the transportation industry, forecasting and predicting are essential abilities directly affecting planning processes and projects. All transportation projects must be planned to satisfy both present needs and future demands. To accomplish this goal, current data related to the project's environment need to be collected and future data need to be forecasted. The data collection aspect of a transportation project is where travel surveys take center stage. Transportation planning and forecasting has relied on travel surveys since the mid 1940s as the method used to collect travel data (Alder and Rimmer, 2002). As more research is directed toward the improvement of questionnaires, travel surveys become more direct, cost effective, dynamic and utilized (Alder and Rimmer, 2002). However, the basic travel survey methodology has not experienced any drastic improvements and past techniques are still used.

The reliance on outdated methods to collect travel survey data limits their use and effectiveness. The incorporation of technology can drastically improve the process. Technological tools are numerous and could be used to enhance travel questionnaires. One major technological tool is the Internet, which has the ability to reach almost all people and if applied as a survey medium, could potentially improve survey management, convenience and capacity (Alder and Rimmer, 2002). Other technologies, such as geographic coding, could also be incorporated into the survey's design in order to make the questionnaire easier for respondents to answer and less complicated for interviewers to analyze. Consequently, with the intention of investigating the usefulness and promising application of these technologies, the University Transportation Center of Alabama (UTCA), the Alabama Department of Transportation (ALDOT) and the University of Alabama in Huntsville (UAH) sponsored a project to test if Internet surveys can be used to collect travel data, ultimately upgrading the current trip generation models and friction factors used in the State. The household demographics and travel data needed for the trip generation models include the average number of trips per household per day.

The scope of this paper includes the implementation of Internet features in a travel survey and the verification of the results with respect to data quality. Issues such as response rate, quality of responses and characteristics of respondents are examined in order to investigate the efficacy of the Internet survey.

Literature review

Travel surveys provide the data necessary to develop trip generation models (Stopher and Metcalf, 1996). The need for a large quantity of accurate data is vital to developing relationships to convert socio-economic characteristics to travel needs. Consequently, one of the most important issues in the initial stage of the modeling process is the selection of the travel survey methodology, as a good selection could save time and money. Using the Internet as a tool to enhance travel survey capabilities could open new opportunities and advantages to trip generation modelers.

An FHWA report states that to collect household and daily trip information about a sample population, the current surveying techniques rely basically on the combination of two methods, telephone and mail back (Richardson et al., 2000). In these types of surveys, households are randomly chosen, contacted by phone and mailed a travel dairy to each trip-making member of the family. The use of telephones is widely accepted since at least 93% of the households nationwide have telephone service (Alder and Rimmer, 1999).

Internet applications are currently used for several types of surveys such as health related studies, user satisfaction analysis, marketing and advertising, etc. The basic layout for the development of an Internet survey consists of the design of the web page and the survey questionnaire. The respondent's task is to complete and submit the questionnaire. A database is created to store the submitted data, which can be then analyzed. Richardson et al. state some of the advantages of using the Internet, such as high-speed Internet access, e-mailing of survey materials and better response time (Richardson et al., 2000). Hence, some consequences are greater choice for respondents of when and where to be interviewed, use of interactive tools and graphics and use of multimedia methods that could simplify complicated survey questions and make it much easier and faster. Also, improving the appearance and understandability of survey instruments increases the response rates. In the past years

Internet usage has increased considerably, over 65 percent of the U.S. population (Alder and Rimmer, 2002) and its integration into daily life in America has increase, thus further enhancing its potential use a data collection tool. Internet survey methods, however, are still considered limited because the sample of Internet users are not considered to be demographically representative of the entire population (Alder and Rimmer, 2002).

For some Internet based travel surveys, various techniques have been used. In Las Cruces, NM, e-mailing was employed for the distribution of questionnaires in combination with a $2 cash incentive (Alder and Rimmer, 2002). Likewise, the distributions for the surveys in the Indiana, Central Florida, Volusia County, and Chattanooga were sent by email and included various cash rewards (Alder and Rimmer, 2002).

However, there are also difficulties for posting web surveys. These include errors in the source code for the questionnaire or issues dealing with the setup of the web page design. These obstacles have to be taken care and revised in order to provide the user with the appropriate web page layout that will enable him/her to accurately resolve the questionnaire and also facilitate data retrieving. Nevertheless, Internet travel surveys are becoming gradually more popular because of their low cost, simple usage, easy access, software availability, faster response rates, faster intercommunication and interaction, high geographical technologies and administrative capabilities, etc (Alder and Rimmer, 2002).

To facilitate the correct development of an Internet travel survey, the following recommended list of necessary items has been developed (Alder and Rimmer, 1999):

- Household structure
- Number of adults and children
- Number of licensed drivers
- Socio-economic status
- Trip information
- Trip start and end times
- Trip purposes
- Travel mode
- Location of trip origin/destination

The survey developed for this paper includes the above items, but disregards the travel mode option because it only focuses on auto trips made by the household members, as the existing travel model for Huntsville only contains automobile trips. Also, the survey will generate table summaries of the data inputs with the aim of alerting the respondent and making him/her double-check their answers.

Huntsville, AL Internet Travel Survey

A case study for the Huntsville, AL metropolitan area is presented with information about planning, pretests, advertising, observations and project supervising. After collecting a sample of household trip data, the inputs were validated to determine is a representative sample from the general population could be obtained.

The first step in the development of the survey was to design an introduction page that presented information about the surveys sponsors and researchers as well as key concepts, objectives and general instructions. The user was presented with background information to introduce him/her to the survey's topic and allow an easier and faster response. Beyond the introduction page, the survey was divided into two parts. The first was used to collect information about household characteristics. Information included the number of people in

the household by gender, age, and race; annual household income; number of vehicles by type, manufacturer and model year. The second part of the survey involved the collection of number of trips made by each vehicle in the household during a normal weekday. In addition, the respondent was asked to enter the number of passengers, trip starting time and purpose. The survey was developed using Microsoft Frontpage attached to a Microsoft Access database and the graphics for the maps were developed in ArcGIS.

The collection of trip data was designed to allow for two different methods of collecting the same information. The first method required respondents to type physical addresses for the starting and ending locations for all trips. The addresses were saved in a database and required post-processing, or geocoding of the trips, to geographically locate origins and destinations. This method caused some difficulties for respondents since often specific addresses of destination points are not known. Therefore, the preferred method involved graphical selection of starting and ending trip locations from an interactive map. The geographic coordinates of the selections were automatically saved in the database.

Entering trip information using the interactive maps required the user click the starting and ending location for each trip. If either the start or end point were located in the downtown area, a larger downtown map appeared to help the user locate the specific location. After selecting the trip origin and destination, the respondent was asked to input the # of passengers, trip starting time and purpose, and then select submit to input the trip. Then, the window would refresh with the origin of the next trip previously identified, as it was the destination of the previous trip. Thus, the user only had to input the destination and the trip information for all subsequent trips. Once the user had entered all the trips for the first vehicle of the household, he/she had to select "next vehicle" to enter the second vehicle trips or "DONE" if he/she was finished with the survey. For visualization purposes, the Huntsville city map and the downtown map included features such as major roads and streets, parks, hospitals, recreational areas and airports. For instance, the University of Alabama in Huntsville, Huntsville Hospital, Grissom High School, and the Redstone Arsenal were included. The idea behind including such elements was that the respondent will identify and use these elements as references for locating his/her trip's origin and destination.

The survey was advertised in The Huntsville Times newspaper and was featured on the WHNT Channel 19 Evening News at 6:00 pm and 10:00 pm. After a period of two weeks from the time the advertising of the survey location, a total of 198 responses were collected. Invalid and erroneous and test entries were immediately flagged using the IP address table. The total number of responses decreased to 175 after deleting the incorrect inputs.

To validate information gathered from the demographic section of the survey, Table 1 shows the number of responding households, by income group and compared to the actual census data for the Madison county area obtained from the Huntsville Chamber of Commerce's web site (HSV Chamber, 2005). Note that the number of respondents was reduced as the survey allowed for a skip/no answer response.

Table 1. Number of households per income group

	Demographics: Income						
	Less than 15,000	15,000 - 30,000	30,000 - 50,000	50,000 - 70,000	70,000 and above	skip/not answer	Total
Huntsville case study							

Respondents	9	13	17	22	69		130
Percentage, %	6.9%	10.0%	13.1%	16.9%	53.1%		100.0%
Census data Madison county							
Number of households	16,212	26,732	17,919	21,890	27,332		110,085
Percentage, %	14.7%	24.3%	16.3%	19.9%	24.8%		100.0%

Table 4.1 shows the response distribution of the number of households per income groups. The household cross section obtained from the Internet travel survey shows only 6.9% of the survey respondents claimed a household income less than $15,000, while the actual percentage of Madison County households is 14.7%. Households with high annual income, above 70,000, were responded at a greater percentage than households. This result in not necessarily novel as lower-income families may not have permanent or easy access to the Internet. However, a positive factor from these results is the knowledge that the higher income households in Huntsville are willing to respond to an Internet travel survey using a low cost survey mechanism and without the promise of cash incentives. From this, the notion of using the Internet to reach a certain segment of the population with limited resources is promising.

Furthermore, a zip code distribution graph was developed to assess how effective the survey was in achieving a spatial cross section. Figure 2 presents a dot density plot of the zip code distribution. Most of the responses came from zip codes 35758, 35761, 35803 and 35806 but users with zip codes from other counties were also reached, such as 35058 from Cullman, AL and 38488 from Taft, TN.

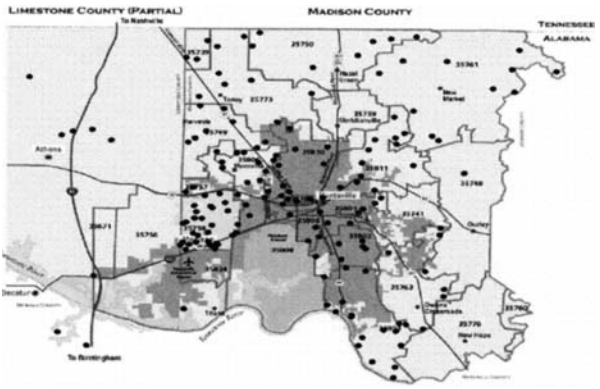


Figure 1. Zip code distribution dot density plot

After evaluating the demographic characteristics of the survey's users, the trip inputs are analyzed to further examine the survey's performance. Since trip related data and household characteristics were asked in different sections of the survey, the results and conclusions about their success were related but independent. Table 2 shows trip inputs for a single person household with one vehicle where the user correctly followed the instructions. Unfortunately, many people who responded to the initial part of the survey did not complete

this section in its entirety, limiting the quality travel data necessary for trip generation modeling.

Table 2. Correct trip inputs

Id	Time	Passengers	Purpose	Vehicle #
101	08:00:am	1	School	1
101	09:30:am	1	Home	1
101	3:00:pm	1	School	1
101	05:30:pm	1	Work	1
101	08:00:pm	1	Home	1

Conclusions

On the whole, the survey was moderately successful. Despite the low number of responses, the demographic analysis turned out to be promising and the Internet survey proved to be effective when collecting this type of data. The main reasons for limited travel data are the format of the web pages and the time required to complete the entire survey. Once these features are revised, an Internet travel survey could effectively collect both household and trip related data.

In order to better deploy the Internet travel survey, various aspects could be improved to enhance its performance. Such aspects consist of a web page with better design, a more organized layout and more powerful capabilities with respect to trip features. Similarly, a more aggressive advertising plan could be used. Cooperation from city organizations should be implemented in order to increase public involvement. Similarly, involvement from MPO's and city counseling groups would be beneficial since feedback and recommendations could be obtained.

References

Adler, T. and Rimmer, L. (1999) SBIR Final Project Report Computer-Based Intelligent Travel Survey System: CASI/Internet Travel Diaries with Interactive Geocoding. USDOT. 1999.

Adler, T. and Rimmer, L. (2002) SBIR Phase II Final Report: Computer-Based Intelligent Travel Survey System. USDOT.

HSV Chamber. (2005) Chamber of Commerce of Huntsville/Madison County: *http://www.huntsvillealabamausa.com/new_exp/newexp_toc.html* Accessed October 3, 2005.

Richardson, A.J., Lee-Gosselin, M. and Griffiths, R. Travel Surveys. (2000) USDOT. Committee on Travel Survey Methods.

Stopher, P. and Metcalf, H. (1996) Methods for Household Travel Surveys, NCHRP Synthesis of Highway Practice 236.

Transportation Network Analysis to Access Community Resources by Older Populations

Srinivas S. Pulugurtha[1], Barbara A. Hirshorn[2], Vanjeeswaran K. Krishnakumar[3] and John E. Stewart[4]

[1]Department of Civil Engineering, University of North Carolina, Charlotte, 9201 University City Boulevard, Charlotte, NC 28223-0001; Tel. (704) 687-6660; Fax (704) 687-6953; Email: sspulugu@uncc.edu

[2]The Center on Aging, University of Nevada, Las Vegas, 4505 Maryland Parkway, Las Vegas, NV 89154-5084; Tel. (702) 895-2039; Email: Barbara.Hirshorn@unlv.edu

[3]The Louis Berger Group Inc., 500 Amigo Court, Suite 100, Las Vegas, NV 89119; Tel. (702) 736-6632; Fax (702) 736-0704; Email: vkumar@louisberger.com

[4]Institute for Families in Society, University of South Carolina, Columbia, SC 29208; Email: jstewart@gwm.sc.edu

Abstract

This paper discusses the need for, and illustration of, a methodology for accessing community-based resources by older resident populations. The methodology involves: 1) identifying census blocks groups with older, non-institutionalized populations that may be "at high -risk" or "at moderate -risk" of a decreased ability to reside in a non-institutionalized community-based setting; 2) geocoding locations of community-based resources (e.g. health and social services, law enforcement and emergency services, and local government offices) key to the well being of older residents; 3) conducting a transportation network analysis; 4) identifying the best travel path for older residents to access various needed community-based resources. "At high -risk" and "at moderate -risk" census blocks groups are identified as a function of three demographic characteristics: the relative proportion of the population 65 and older in a census block group (1) with income below the poverty threshold, (2) with one or more disabilities, and, given those two factors, (3) living alone. Network analysis tools in standard GIS software are used to identify the best travel path to access community resources from "at high -risk" and "at moderate -risk" census block groups. The identified "best travel path" is based on travel distance, travel time, traffic volumes, number of intersections along the path, and the number of crashes involving older residents. The methodology is illustrated using data from a project that used GIS methods to indicate the physical relationship of older sub-populations to needed community resources in a section of the Las Vegas, Nevada metropolitan area.

Introduction

Census Data 2000 show that more than 12 percent (an estimated 35 million) of the persons in the United States (US) are 65 years or older (Census Bureau, 2001). One in seven licensed drivers in the US is 65 years or older (Sivak et al., 1995). Advancements in the field of medical science, nutrition, and improved health standards, enabling people to live longer, are expected to have a direct impact on the growing older population so that, by 2030, nearly 20 percent of the US population (an estimated 70 million persons), and one in five licensed drivers, will be 65 years of age or older. The majority of this population will be non-institutionalized and will need to travel to access community resources vital to maintaining

well being and autonomy -- recreational facilities, the Social Security Administration office; mental health services; hospitals; urgent care centers; employment and training facilities; and retail establishments (e.g. vision and dental care providers, pharmacies, and groceries).

The geographic distribution of community resources and access to these resources pose transportation-related challenges for many older residents. Geographic Information System (GIS) methods provide spatial and aspatial analyses for responsible agencies to use to respond to these transportation-related distribution and access issues. In this paper, standard GIS methods were used to develop and illustrate a process to improve access to resources for older, non-institutionalized residents.

Methodology

The methodology involves the following steps:

Step 1: Identify "At High -risk" and "At Moderate -risk" Elderly Population Block Groups

Increasing old age often is associated with increased frailty and a diminution of income. It is a time of life that also may be marked by an attenuation of one's social support network – a chronic issue in communities at any point along the urban-rural continuum. This issue takes on additional resonance in a Sunbelt in-migration "boom town", such as Las Vegas, where many have moved during middle or old age for employment opportunity or retirement living (Frey and DeVol, 2000; Litwak and Longino, 1987).

To obtain a measure that identifies relative levels of vulnerability among non-institutionalized older populations in each census block group, a GIS layer is created representing the distribution of the older population according to defined and relevant characteristics that put older residents "at -risk" of not accessing needed community resources and, therefore, a compromised capacity to maintain personal autonomy and well being in a non-institutionalized, community-based environment. Using Census 2000 data, the following critical parameters are estimated to compute "at -risk" values for each census block group:

1) Percent of the population 65 years of age and older below the poverty threshold (a primary risk factor)
2) Percent of the population 65 years of age and older with one or more disabilities (a primary risk factor)
3) Percent of the population 65 years of age and older living alone, a secondary risk factor when accompanied by other characteristics indicating vulnerability (such as low income or disablement)

A threshold value of one standard deviation above the mean was used for each of the above parameters. Based on the estimated values for each parameter for each census block group, the older population in a particular census block group is categorized as "at high -risk", "at moderate-risk", or "at low -risk". If the estimated values of all three parameters exceed their corresponding threshold value, then older population vulnerability is characterized as "at high -risk" for that census block group. If the estimated values on either or both of the primary risk factors ("1" and "2" above) exceed their corresponding threshold values, then that block group is designated as "at moderate -risk" for older population vulnerability. If a block group falls below the threshold value on both of the primary risk factors, then it is designated as "at low -risk" (Hirshorn et al., 2004).

Step 2: Create GIS Layers Representing Community Resources

The method of deriving spatial coordinates for a specific location based on street name / reference street name, street address or mile-post is called "geocoding" (ESRI, 2002). Since,

the study area is in an urban area, the information in the attribute table of the street network is used to locate the addresses for geocoding based on street name / reference street name or street address. The output of a geocoding process is either a "shapefile" or a "geodatabase" feature class of points. The geocoding feature available in standard GIS software can be used to geocode the community resources data into a GIS format. Using this feature, layers representing community resources are developed.

Step 3: Conduct Transportation Network Analysis and Identify Best Path

The focus of this step is to conduct a transportation network analysis to study the accessibility of community-based resources from "at high -risk" and "at moderate -risk" census block groups. Transportation network analysis has been routinely used by engineers and planners to formulate mathematical models describing the evolving network flow conditions of traffic, evaluate stochastic conditions that arise in network design, and to improve the performance of the system. A network analysis also helps identify spatial gaps between particularly "at –risk" older populations living in the study area and critical community resources. In this analysis, two or three "shortest alternate routes" are selected for each origin-destination. High speed roads and congested sections with greater variance in speed between vehicles (such as freeways) generally are not preferred by older residents and, therefore, are ignored in the analysis.

The selection of path or a mode of travel depends on factors related to the older individual (poverty level, disabilities, living alone) and the transportation network (travel distance, travel time, traffic conditions, number of crashes, and number of conflict points). In particular, these factors may greatly influence the decision by an older person to use transit, para-transit, or a ride as the mode of travel. If the older person does not have a choice or would prefer to drive, the network analysis can help the person select the "best path" for travel between an origin and a destination. The "best path" is the one that is a combination of travel time, the number of crashes, and the number of conflict points faced by the individual. The number of conflict points, representing the number of times path of traffic movement crosses, is greater in number at intersections (in particular, signalized intersections) as compared with mid-block locations. So, the fewer the number of signalized intersections along the travel path, the lower is the number of times a driver may have to make critical safety decisions.

The following factors were considered important in selecting the "best path".

1) Length of the route (miles)
2) Speed limit (miles per hour)
3) Travel time (minutes)
4) Traffic conditions (vehicular volume)
5) Number of crashes
6) Number of signalized intersections

Travel time for a link is computed by dividing the length of the link by the posted speed limit. The travel time for each link are then summed to compute the travel time for the path (TT).

$$TT = \sum_{i=1}^{i=n} \frac{L_i}{V_i}$$

where,

n = number of links in the path
L_i = length of the i^{th} link (miles)
V_i = posted speed limit of the i^{th} link (miles per hour)

Traffic volumes can be obtained from annual highway traffic reports. Tools were developed to estimate the number of crashes using crash data from the responsible local agency. The number of signalized intersections or all intersections can be obtained from the network database.

The values for each of these variables are normalized to the same scale to obtain a score for each category. Such a normalizing procedure addresses the challenge of combining disparate components (Pulugurtha et al., 2005). The individual scores for each component are normalized to a 0-100 scale. Then, the individual scores for each category are added to compute the final score for the path.

Different weighted scores can be given for different categories in estimating the final score. Heavy traffic volume, longer travel distances, lower speed limits, and higher numbers of signalized intersections have direct impacts on the path travel time. However, crashes and the various negative consequences associated with them have an impact on the individual's life and family. Thus, a weight of "two" is given to number of crashes and "one" each for travel time, traffic volumes, and number of signalized intersections when calculating the final score for each alternative route.

$$\text{Score TT} = \frac{\text{TT}}{\text{Maximum TT}} \times 100$$

$$\text{Score TV} = \frac{\text{TV}}{\text{Maximum TV}} \times 100$$

$$\text{Score CR} = \frac{\text{CR}}{\text{Maximum CR}} \times 100$$

$$\text{Score SI} = \frac{\text{SI}}{\text{Maximum SI}} \times 100$$

$$\text{FS} = \text{Score TT} + \text{Score TV} + 2 \times \text{Score CR} + \text{Score SI}$$

where,

Score TT is the score for the average travel time of the path,
Score TV is the score for the traffic volume of the path,
Score CR is the score for the number of crashes in the path,
Score SI is the score for the number of signalized intersections in the path,
Maximum TT is the maximum of TT values for all the alternative paths,
Maximum TV is the maximum of TV values for all the alternative paths,
Maximum CR is the maximum of CR values for all the alternative paths, and
Maximum SI is the maximum of SI values for all the alternative paths.

Based on the analysis, two paths are selected for each origin-destination pair – "shortest path" and "best path". The "shortest path" is the one with the minimum route length, and the "best path" is the one with the minimum final score computed using the above equations.

Case Study

Recent trends show that Nevada ranked first nationwide in the percent increase of the population 65 and older between 1990 and 2000. More than 70 percent of the population in the State of Nevada lives in the Las Vegas metropolitan area. Epitomizing this change is the geographic focus of this study – a section of the "east-side" valley in the Las Vegas metropolitan area (from here on referred to as "the study area"). The study area is bounded by Sahara Avenue on the north, Boulder Highway – US 95 on the east, Russell Road on the South, and Paradise Road on the west and contains 65 entire census block groups as well as

parts of 17 contiguous census tracts.. The area is home to a growing number of older people who are "aging in place" in neighborhoods that are aging themselves (Hirshorn et al., 2004).

To indicate some of the spatial gaps between older populations and community resources that are key to maintaining their well being and autonomy, a network analysis was performed showing the travel distance for older residents in "at high -risk" and "at moderate -risk" areas to some of these important destinations. Table 1 shows the network analyses results from the "at high -risk" census block group to one of the selected community resources (University Medical Center emergency room). The selected path, distance in miles, average speed limit in miles per hour, average travel time in minutes, number of crashes, number of signalized intersections, normalized scores for each category, final index value, and the final selected path are shown in the table. "S" and "B" under the final path column in the table indicate "shortest" and "best" path, respectively. Maryland Parkway is the shortest path from the "at high -risk" census block group to the University Medical Center emergency room. However, Eastern Avenue, because of its higher speed limit and fewer numbers of crashes along the route, is the best route from the "at high -risk" census block group to University Medical Center emergency room. Figure 1 shows 1) census blocks groups according to risk level, and 2) the shortest and best path from the "at high -risk" census block to University Medical Center emergency room.

Conclusions

Census blocks groups were categorized into "at high -risk", "at moderate -risk", and "at low -risk" groups based on the relative proportions across census block groups of the population 65 and older below the poverty threshold and with disabilities, and, given those two indicators of vulnerability in old age, living alone. GIS methods were used to geocode locations of community resources. A transportation network analysis was conducted to identify the shortest and best path, to assist older residents in deciding the mode and path to travel to their destination. The selection of the path is based on path travel time, traffic volume, number of crashes, and number of signalized intersections. Crashes were given more weight because of the impact they could have on an individual's life. In addition to helping older residents access needed community resources, the methodology and results can be used by organizations to identify reachable locations for siting community resources. The methodology is robust and has the potential to be applied to other urban settings and other transportation-related problems, as well.

Acknowledgements

The authors acknowledge Nevada Housing and Neighborhood Development (H.A.N.D.) for its financial support for the project upon which this investigation is based.

Table 1. Network Analysis Results.

Paths	Distance (Miles)	Average Speed Limit (mph)	Average Travel Time (min)	# of Crashes	# of Signalized Intersections	Annual Average Daily Traffic	Score TT	Score CR	Score SI	Score TV	Final Index	Final Path
			TT	CR	SI	TV						
Eastern - Path 1	6.08	41	9	4,012	7	36,000	88	74	88	90	85	Best
Maryland - Path 2	5.39	32	10	4,871	7	38,000	97	89	88	95	92	Shortest
Paradise - Path 3	6.22	36	10	5,455	8	40,000	100	100	100	100	100	

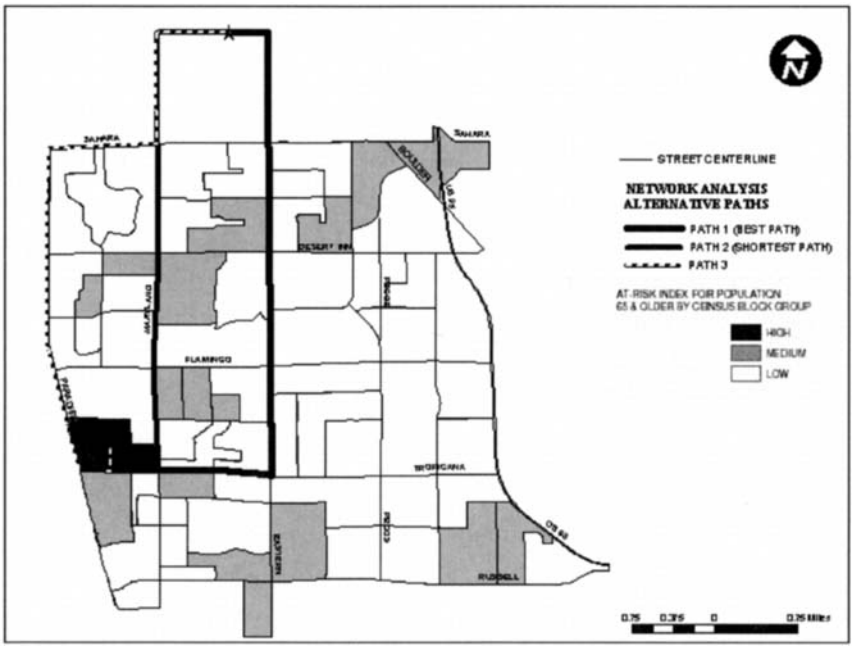


Figure 1. "At-risk" Block Groups and Illustration of Transportation Network Analysis.

References

Census Bureau (2001) U. S. Census 2000 Population Tables and Reports. U.S. Census Bureau Website http://www.census.gov/main/www/cen2000.html, 2001.

ESRI. Arc/Info Documentation – Version 8.2. Environmental Scientific and Research Institute, Inc., Redlands, CA, 2002.

Frey, W. H., and DeVol, R.C., (2000). America's Demography in the New Century: Aging Baby Boomers and New Immigrants as Major Players. The Milken Institute, Santa Barbara, California.

Hirshorn, B.A., Pulugurtha, S.S., Krishnakumar, V., Stewart, J.E., Wongchavalidkul, N. (2004). Aging on the 'East Side' of the Las Vegas Valley: A GIS-Based Analysis of Older Residents and Their Community Context. Las Vegas, Nevada: The UNLV Center on Aging and the UNLV Transportation Research Center.

Litwak, E. and Longino, Jr., C.F. (1987). Migration Patterns Among the Elderly: A Developmental Perspective. The Gerontological Society of America, Vol. 27, No. 3, 266-272.

Pulugurtha, S S., Nambisan, S. S., and Uddaraju, M. (2005) Methods to Rank High Pedestrian Crash Zones. (Paper Number 05-0723). CD-ROM Preprint, 84th Annual Transportation Research Board Meeting, Washington, DC, January 9-13.

Sivak, M., Campbell, K. L., Schneider, L. W., Sprague, J. K., Streff, F. M., and Waller, P. F. (1995). The Safety & Mobility of Older Drivers: What We Know & Promising Research Issues. UMTRI Research Review 26. Ann Arbor: University of Michigan Transportation Research Institute.

An Activity-Based Model of Travel Demand in the Jakarta Metropolitan Area

S. Yagi[1] and A. Mohammadian[2]

[1]Dept. of Civil and Materials Engineering, University of Illinois at Chicago, 842 W. Taylor St., Chicago, IL 60607; PH (312) 413-7447; FAX (312) 996-2426; email: syagi1@uic.edu
[2]Dept. of Civil and Materials Engineering, University of Illinois at Chicago, 842 W. Taylor St., Chicago, IL 60607; PH (312) 996-9840; FAX (312) 996-2426; email: kouros@uic.edu

Abstract

The goal of the study is to develop a comprehensive activity-based modeling system in the context of developing countries, providing accurate estimates which are expected to serve as better inputs for evaluation of different transportation policy scenarios. The case study is Jakarta, Indonesia as one of the largest metropolitan areas in Asia. The modeling system primarily adopts a tour-based structure in which the tour is used as the unit of modeling travel instead of the trip, preserving a consistency in destination, mode, and time of day across trips. The model is developed using the available activity diary survey and household travel survey data. It is a system of disaggregate logit and nested logit models assuming a hierarchy of the model components, with three major models (choices of daily activity-travel pattern, times of day, and mode and destination) with two additional sub-models (choices of mode and destination for work-based sub-tours, and intermediate stop locations). In Jakarta, a variety of urban transportation policy scenarios such as extension of bus rapid transit, area pricing, and license plate restriction are currently being examined, discussed, or implemented. It is hoped that the new models developed in this study will contribute to better understanding of urban travel behavior and improvement of the travel demand forecasting models that will result in better evaluation of urban transportation policy scenarios in the region.

Introduction

Activity-based travel demand modeling assumes that travel is derived from the demand for activity participation. It has been shown that activity-based modeling framework can analyze travel as a daily pattern of individuals' behavior that is related to differences in their lifestyles and activity participation. As a result, travel needs to be studied in a broader context of activity scheduling. Despite advances in academic research on activity-based modeling, the majority of the metropolitan planning organizations (MPOs) in the United States are still using conventional regional models based on the four-step modeling paradigm with numerous variations and enhancements. However, a growing number of MPOs either have already developed and applied models of the new type or have at least made a decision to start development of a new model, sometimes in parallel with maintenance and enhancement of the existing four-step model (Vovsha, et al. 2003). It seems that activity-based models are more popular in developed countries than in developing countries while transportation related problems such as congestion and air quality issues are most serious in developing world.

The goal of the study is to develop a comprehensive activity-based modeling system of which structures and control factors may be significantly different in the context of developing

countries, providing accurate estimates which are expected to serve as better inputs for evaluation of different transportation policy scenarios. The study simulates the way individuals schedule their daily activities and travel in an urban region of the developing world. The case study is Jakarta, Indonesia as one of the largest metropolitan areas in Asia with a population over 21.6 million people and 5.6 million households.

In the Jakarta Metropolitan Area, a variety of urban transportation management policies are currently being examined, discussed, or implemented. For example, the "3-in-1" scheme, in which only high-occupancy vehicles are allowed to use central arterial roads during peak periods, has recently been more widely and strictly enforced along with the introduction of a new bus rapid transit (BRT) using the centermost lane of the arterial roads as a dedicated "busway." More new policies regarding transportation control measures such as area pricing, parking pricing, and license plate restriction are now being discussed for implementation as well as expansion of the BRT system. In addition, raising the fuel price is a current controversial issue that is being considered and implemented by the Indonesian government. As such, it is hoped that the proposed new models will contribute to the improvement of the methodology of travel demand forecasting and evaluation of urban transportation policy scenarios in the region. This paper describes the general framework and methodology of activity-based model that has been applied in this study along with some preliminary results of the analyses.

Data Description

It is worth noting that the activity-based model requires a much richer and more detailed dataset than those required for the conventional travel demand modeling systems. Fortunately, the Jakarta Metropolitan Area has excellent sources of survey data including a regular full-scale household travel survey (HTS) and an activity diary survey (ADS) conducted in 2002. The HTS provides the largest and most comprehensive travel data in the region, covering as many as 166,000 households (3% of the population). On the other hand, the ADS provides a series of daily household (both in-home and out-of-home) activities on the basis of every 15 minutes for 24 hours over four consecutive days, covering around 4,000 individuals that are linked to the corresponding observation in the HTS by a unique identifier. Both surveys cover activity-travel information of individuals of all generations including children who are 5 years or older (Yagi and Mohammadian, 2005). The activity-based modeling system is developed using these two datasets that provide a unique opportunity to conduct numerous research works.

Modeling Framework

A "trip" is defined as a travel between two activities representing the trip purpose (Home to Work, Home to School, etc). The term "purpose" in this study is used to present the activity performed at the trip end. Furthermore, each trip record is coded with travel mode (auto, motorcycle, transit, etc.). A "tour", on the other hand, is defined as a chain of trips which start from a base and return to the same base. One or more activities (i.e., purposes) are involved in the course of a tour. A tour has been considered a home-based tour if it starts from home and ends at home.

Bowman and Ben-Akiva (2000) developed a model in which they consider the daily activity-travel pattern as a set of tours. Activities are prioritized based on the purpose of the activity, with work activities having the highest priority, followed by work-related, school, and all other purposes. Within a particular purpose, activities with longer durations are

assigned higher priorities. The tour of the day with the highest priority activity (i.e., the "primary" activity) is designated as the "primary" tour and others are designated as "secondary" tours. This modeling framework represents the daily activity-travel pattern by three attributes: the "primary" activity, the "primary" tour (i.e., number, purpose, and sequence of activity stops), and the number and purpose of "secondary" tours. The tours are represented by the time of day (five time periods), destination (TAZ: traffic analysis zone), and mode. A prototype of the model was tested in Boston, and it was fully applied in Portland, Oregon (Bowman, et al. 1998; Bradley, et al. 1999).

For this study, the modeling system primarily adopts a modified version of similar tour-based structure in which the tour is used as the unit of modeling travel instead of the trip, preserving a consistency in destination, mode, and time of day across trips. The overall modeling structure is depicted in Figure 1. It is a system of random utility based disaggregate logit and nested logit models assuming a hierarchy of model components, with three types of major models, namely, choices of daily activity-travel patterns, times of day, and mode and destination in the hierarchy. Lower level choices are conditional on decisions at the higher level, and higher level decisions are sensitive to the lower level choices through expected utility logsums, where applicable. Before processing the final output, two additional sub-models are added to this framework in order to determine additional characteristics of tours including destinations and modes of work-based sub-tours and locations of intermediate stops. Each model is described in the next section.

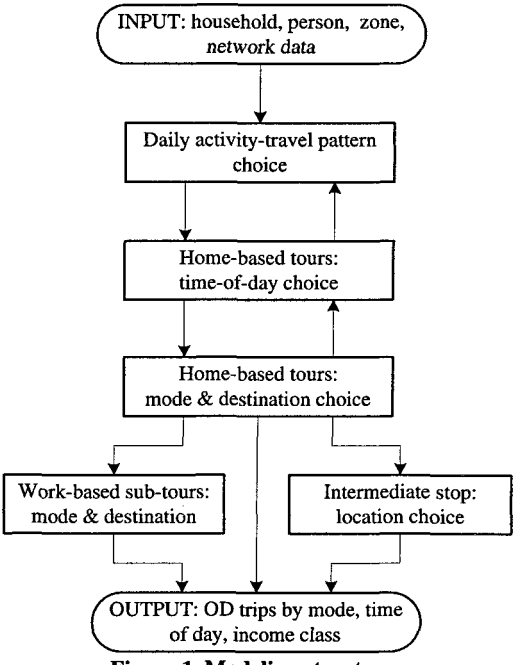


Figure 1. Modeling structure.

Model Description

Daily Activity-Travel Pattern Choice

Daily activity-travel patterns except for the stay-home-all-day option are defined by primary activity, primary tour type, and number and type of secondary tours. The most important out-of-home activity of the day that often has the longest duration is designated the primary tour and all other tours are designated as secondary. All activities in a tour are ranked by priority, with work and work-related, school, maintenance (e.g., shopping, banking, visiting doctor, etc.), and discretionary (e.g., social and recreational activities, eating out, etc.) purposes, in this order. Thus, primary activities or purposes are classified as *home*, *work*,

school, *maintenance*, and *discretionary* for the sake of modeling.

Primary tour type is defined by presence and sequences of stops and presence of work-based sub-tours in a tour. The primary tour type classification depends on whether it is a simple tour from home to a destination and back to home again with no intermediate stops, or a tour with at least one intermediate stop for another activity on the way from home or on the way back home; and whether it includes at least one work-based sub-tour (for *work* tours only). With regard to intermediate stops, for simplicity, activity purpose (e.g., *maintenance* or *discretionary*) is not considered in this model.

On the other hand, secondary tour type is defined by activity purpose, and it is classified into *maintenance* and *discretionary* activities. Neither number nor sequences of intermediate stops in a secondary tour are considered in the model. Number of secondary tours is counted from the dataset and classified into 0, 1, and 2 or more secondary (*maintenance* or *discretionary*) tours. Combination of the primary activity, primary tour type, and number and type of secondary tours brings about a choice set of 105 daily activity-travel alternatives, including 1 *home* pattern, 48 primary *work* tour patterns, 24 primary *school* tour patterns, 24 primary *maintenance* tour patterns, and 9 primary *discretionary* tour patterns.

The most notable difference in relative frequency of the activity-travel pattern alternatives between the HTS and ADS datasets may be the relative frequency of the home pattern. In the HTS, the percentage of stay-home-all-day is nearly four times as high as the one derived from the ADS. In addition, alternatives of simple primary tour type with no secondary tours have higher percentages in the HTS. The main cause for such differences may be that relatively short-distance tours, especially those made by non-motorized mode of transport tend to be overlooked in the HTS. Although modeling based on the HTS database could eventually have no major impact on the forecast of motorized trips, the ADS database has been selected for model estimation of daily activity-travel patterns.

The model for daily activity-travel pattern choice, which is placed at the top of the entire modeling system, has a two-tier nested logit structure, with a choice of whether to *travel* or stay at *home* all day in the upper tier and a choice of out-of-home activity-travel pattern alternatives defined by primary tour activity, primary tour type, and number and type of secondary tours in the lower tier under the *travel*. On the other hand, *home* alternative is a degenerate branch with only one stay-home-all-day alternative. This model is estimated using many variables explaining attributes of the household and the individual. Logsum variables from the lower level, that is, time-of-day choice for *work*, *school*, *maintenance*, and *discretionary* are also included in the model.

Time-of-Day Choice

In order to model the time-of-day choice, a day is divided into five time periods, namely, early morning (3:00 – 6:30), a.m. peak (6:30 – 10:00), midday (10:00 – 16:00), p.m. peak (16:00 – 19:00), and night (19:00 – 3:00). These five time periods are distinguished considering not only characteristic hourly traffic volume but also Jakarta's unique "3-in-1" traffic regulation hours. Alternatives are created by combining the time period to leave home to start the tour and the time period to leave the main destination to start the returning leg of the tour. Assuming that there are no tours that last over night, 15 time-of-day combinations or alternatives are identified.

The time-of-day choice is a multinomial logit model with 15 alternatives, and it is estimated

separately for each of the four purposes using the HTS database. In addition to the variables describing attributes of the tour, household, and individual, the logsum variable from the lower-level mode and destination choice is included in the utility functions. Since a variety of public transport is available throughout the Jakarta Metropolitan Area from early morning till late evening, the logsum variable has been applied to all combinations of time periods.

Mode and Destination Choice

Mode and destination choice is placed at the "bottom" of the hierarchy, and is conditional on decisions at the higher levels, that is, choices of daily activity-travel patterns and times of day. For mode and destination choice, from-home trips are focused on and used to estimate the entire tour mode and destination choice model, because these trips constrain the modes and destinations of the subsequent segments such as returning-home trips. The results of the HTS show that at least over 90 percent of people return home using the same mode as they used for the from-home trips, though the percentages vary depending on modes and purposes.

Eight most commonly used combinations of travel modes observed in the region are considered. These include auto drive alone, auto shared ride, motorcycle, taxi, motorcycle taxi, transit with motorized access, transit with non-motorized access, and non-motorized transport. As for the destination choice, 11 representative destinations are considered for each tour in order to reduce the computational burden. These destinations are sampled from the 336 TAZs using the stratified importance sampling method. For each purpose, the strata of destinations are constructed based on the distance as well as a size variable which indicates the magnitude of attraction in the destination. Size variables have been set as total jobs for *work*, total students at school place for *school*, total service industry jobs for *maintenance*, and the sum of service industry jobs and households for *discretionary* activities. As a result, this sampling method leads to higher probabilities of being selected for zones closer to the origin as well as for zones with larger potential of corresponding attraction.

This is a joint model of mode and destination choice. Excluding some mode and destination alternatives with extremely low frequencies observed in the HTS database, the maximum number of choice alternatives is presumed to be 76. The model has a two-tier nested logit structure. Auto drive alone, auto shared ride, and motorcycle; and taxi, motorcycle taxi, transit with motorized access, and transit with non-motorized access are each placed in the second tier under different nests while non-motorized transport is placed as a degenerate branch. Although nests are created for each representative destination zone, log-sum parameters are set to be common for the nests which involve the same mode group. This model is estimated using variables describing attributes of the tour, household, and individual as well as skims from the actual network such as generalized travel time, transit walk time, and travel distance.

Additional Sub-Models

Furthermore, two more sub-models are included in the activity-based modeling system, namely, mode and destination choice for work-based sub-tours (in *work* tours only) and intermediate stop location choice. For the work-based sub-tour mode and destination choice, the methodology is similar to that of the above-mentioned mode and destination choice for from-home trips, except that a dummy variable indicating whether the same mode was used to travel from home to work is highly significant in this sub-model. Meanwhile, for the intermediate stop location choice, the stratified importance sampling method is adopted again with different strata of stop locations, and separate models are estimated by mode and by

purpose of the tour. These two sub-models are estimated without returning logsums to the upper, major part of the modeling system.

Summary

This paper describes a comprehensive multi-tier activity-based modeling system for Jakarta, consisting of random utility based disaggregate logit and nested logit models. While the majority of explanatory variables have been directly derived from the ADS and HTS data, some other zone-based data have also been defined such as generalized travel time computed from the highway and transit network, and land use composition in each TAZ computed from the GIS database. Thus, a number of types of significant variables are included in the modeling system.

Results from the activity-based micro-simulation model will be fed into a link-based transportation network assignment program in the form of OD trips by mode, and time of day. Outputs from the network assignment model representing vehicle miles/hours of travel and average travel times can be used to evaluate various policy scenarios or economic impacts of different alternatives. Additionally, outputs by link such as volumes, average speeds, and average congestion levels will be used as part of the input data to the emission model.

Like many other metropolitan areas in developing countries, Jakarta is facing a transition from conventional infrastructure and highway construction to lower-cost but environment-friendly urban transportation management. It is hoped that the new models developed in this study will contribute to a better understanding of urban travel behavior and improvement of the methodology of travel demand forecasting. It is also hoped that the study to provide a powerful decision-making tool that can be used to better analyze and evaluate urban transportation policy scenarios not only in the Jakarta Metropolitan Area but also in other urban areas of developing countries.

Acknowledgement

The authors would like to express their gratitude to the National Development Planning Agency of Indonesia for permitting us to use the survey data in this study. This study was partially funded by Foundation for Advanced Studies on International Development (FASID).

References

Bowman, J.L., and M.E. Ben-Akiva (2000). "Activity-Based Disaggregate Travel Demand Model System with Activity Schedules." *Transportation Research Part A*, Vol. 35: 1-28.

Bowman, J.L., M. Bradley, Y. Shiftan, K. Lawton, and M.E. Ben-Akiva (1998). "Demonstration of an Activity Based Model System for Portland." Paper presented at the 8th World Conference on Transport Research, Antwerp, Belgium.

Bradley, M., J.L. Bowman, and K. Lawton (1999). "A Comparison of Sample Enumeration and Stochastic Microsimulation for Application of Tour-Based and Activity-Based Travel Demand Models." Presented at the *European Transport Conference*, Cambridge, U.K.

Vovsha, P., E. Petersen, and R. Donnelly (2003). "Experience, Tendencies, and New Dimensions in the Application of Activity-Based Demand Modelling Systems for Metropolitan Planning Organizations in the United States." Paper presented at the 10th International Conference for Travel Behaviour Research Conference, Lucerne, Switzerland.

Yagi, S. and A. Mohammadian. (2005). "An Exploratory Analysis of Intra-Household Joint Activity-Travel Tours in the Jakarta Metropolitan Area". Proc. of the *6th Transportation Specialty Conference of the Canadian Society of Civil Engineers (CSCE)*, Toronto, Canada.

Trip Rate Analysis for a Metropolitan City in Indian Context

G.J. Joshi[1] and B.K. Katti[2]

[1]Lecturer, Civil Engineering Department, S.V. National Institute of Technology, Surat – 395 007; PH (91) 261 – 2223371; FAX (91) 261 – 2228394; email: gj_svnit@yahoo.com
[2]Former Professor Emeritus, Civil Engineering Department, S.V. National Institute of Technology, Surat – 395 007; PH (91) 261 – 2211967; FAX (91) 261 - 2228394

Abstract

The major thrust of urbanization in India is on metropolitan cities due to their acquired status of growth magnets providing ample employment opportunities. This has resulted in simultaneous densification and expansion in spatial dimension effecting changes in the socio economic characteristics of city households, land use and travel pattern. Eventually the impact has been on urban trip rates with respect to the socio – economic stratification of urban society. The present paper displays trip rate pattern in an Indian metropolitan city of Surat in Gujarat state. The study highlights trip rate patterns for different travel purposes and modes under the influence of socio – economic characteristics with reference to city geography.

Introduction

Morphological changes in the urban areas followed by increased population have major influence on travel pattern. Estimation of trip rate and trip generation considering such changes reflecting population travel behaviour forms important basis for planning effective transportation systems for the cities. In developing countries like India urban population is scattered across the vast spectrum of socio – economic status giving rise to varied demand for travel in terms of trip generation for various discretionary and non – discretionary purposes as well as choice of mode. The study observations for fast growing metropolitan city of Surat in India reflect on the considerable influence of household socio – economic status on the purpose and mode based trip generation rates.

Urbanisation Impact

India has witnessed large-scale industrialisation in pursuit to achieve higher economic growth in post independence era. Urban population has increased from 62 million to 285 million with major thrust on the metropolitan cities during this period. Improved socio – economic condition in urban areas has resulted in increase in travel demand and subsequent increase in two wheeler and car population due to inadequate and inefficient public transport supply (Sachdeva, 2004).

Study Area and Surveys

The study area; Surat is the second largest city of Gujarat state in India on west coast. The city located in South Gujarat region has an area of 112.28 sqkm and population of 2.43 million (2001). The city population has increased by 62% in last decade (1991 – 2001). Faster growth of this industrial city has created huge gap between transport supply and demand in terms of imbalanced mode share. Field surveys are carried out by dividing city into 6 macro level units of sectors and 32 micro level units of zones. The impact of radial spatio – temporal growth of the city is examined by delineating CBD, inner fringe and outer fringe areas. Disaggregation of households based on monthly household income (MHHI in INR) into six groups like economically weaker section (EWS), low income group (LIG), lower middle income group (LMIG), higher middle income group (HMIG), higher income group (HIG) and very high income group (VHIG) has been adopted in the present study (see Table 1).

Table 1. Socio – economic groups

Income group	MHHI (Rs.)	Income group	MHHI (Rs.)
EWS	< or = 4,500	HMIG	10,500 – 16,000
LIG	4,500 – 7,000	HIG	16000 – 24000
LMIG	7,000 – 10,500	VHIG	More than 24,000

The home interview survey in all 32 zones had been carried out by stratified random sampling technique to cover all the sections of society.

Trip Rate: Household Impact

Household size, work – participation ratio, income, and vehicle ownership are the most influential factors and explain the trip rate pattern for a metropolitan city. The household travel studies conducted by CRRI in five Indian cities have revealed the fact that the vast varieties of socio – economic groups with the variations in the household income and ownership of vehicles significantly influence daily trip rates (Reddy, 1993). Such observations are also made by Suryanarayana et. al (1990) and Pillai (1993) in India. Generally, more than 95% of the trips are reported to be home based (Government of India, 1998). Observations on trip rates at three levels: household factors, trip purpose and mode choice in typical metropolitan city Surat in India are as under.

Household size, Income and Vehicle Ownership Impact. The household size in the study area is found to vary considerably among different income groups from 4.53 for EWS to 5.97 for VHIG. Mean household size, working members and vehicle ownership are given in Table 2. The household size and therefore the number of working members and schooling members are observed to increase from lower income groups to higher income groups reflecting joint family systems and show trip rates in increasing order. Every fourth household in EWS category possesses a vehicle whereas each household of higher income group owns at least two vehicles.

Table 2. Mean socio – economic variables and household trip rates in Surat

Income group	Household size	Working members	Schooling members	Vehicle ownership	Daily trips/hh*
EWS	4.53	1.27	1.17	0.28	8.1
LIG	5.02	1.50	1.43	0.76	9.5
LMIG	5.45	1.80	1.36	1.18	10.5
HMIG	5.85	1.96	1.49	1.74	11.3
HIG	5.96	2.00	1.58	2.29	11.3
VHIG	5.97	2.10	1.63	3.06	12.5
City	5.32	1.64	1.42	1.25	10.26

* Inclusive of all modes and purposes

Compared to an EWS household a middle-income household performs 25% more trips while a very high-income household makes 50% more trips in a day.

Purpose Impact. The travel being a derived demand is stimulated by the activity purpose. The proportion of travel for non – discretionary purposes like work and schooling predominates over the discretionary purposes like shopping, recreation, and other. Present study scans deeper in the discretionary purpose trips by frequency based disaggregation. The shopping behaviour of households is analysed by weekly travel (W) and monthly travel (M). Similarly, the recreation and other purpose trips are analysed for frequent (F) and occasional travel (O). Table 3 shows purpose based trip rates for six income groups.

Table 3: Purpose based trip rates in Surat city

Income group	Daily trip rates (tpcd)						
	Work	School	Shopping		Recreation & other		Total
			W	M	F	O	
EWS	0.769	0.517	0.397	0.016	0.048	0.038	1.785
LIG	0.833	0.549	0.370	0.019	0.076	0.056	1.903
LMIG	0.921	0.481	0.331	0.020	0.109	0.066	1.928
HMIG	0.921	0.500	0.320	0.020	0.110	0.070	1.941
HIG	0.932	0.495	0.290	0.022	0.089	0.072	1.900
VHIG	0.983	0.519	0.320	0.029	0.153	0.104	2.108
City	0.871	0.512	0.349	0.021	0.090	0.062	1.905

The work trip rate increases with income and its share in total trip rate varies from 43% in EWS to 49% in HIG households. The school trip rate in all the income groups is around 0.5 tpcd (trip per capita per day) with relatively higher share in lower income groups.

It is observed that the lower income group households make more trips in a week while upper income groups prefer monthly shopping. Trip rates for frequent and occasional recreation and other purpose travel are observed to increase with household income in general. Across the income groups, the share of compulsory trips of work and school purposes is found in the range of 70 - 75%. CIRT, Pune has made observations of 82% such trips in 13 Indian cities (Bagade, 2001). Work purpose trips in the range of 36% - 48% and 24% - 45% education purpose trips are reported for southern Indian cities of Hyderabad, Vishakhapattanam, Vijaywada and Warangal (Chandrashekhar, 2005).

Modal Impact. The motorised modes of travel (MT) in the study area include auto rickshaws (IPT), bus, car and two wheelers while walk and bicycle are the non – motorised modes (NMT). In most cities in India two – wheelers i.e. scooters and motorcycles comprise more than 70% of total motorised vehicles (Sachdeva, 2004). The income group wise daily person trip rates by different modes are given in Table 4. The influence of socio – economic class is conspicuous in choice of mode. Modal share in city trips is equally notable. The lower income group households of EWS and LIG have shown very high trip rate by NMT modes where as it reduces to almost half for the higher income group households. Two wheeler trip rate contributes more than 50% of motorised mode trip rate while trip rate by bus, as a travel mode is less than 4%.

Table 4: Modal trip rates (tpcd) in Surat

Income group	Auto	Bus	Car	Two wheeler	MT	NMT	Total
EWS	0.220	0.008	0	0.174	0.402	1.383	1.785
LIG	0.285	0.031	0	0.416	0.732	1.171	1.903
LMIG	0.284	0.036	0.025	0.612	0.957	0.971	1.931
HMIG	0.306	0.066	0.063	0.828	1.263	0.678	1.932
HIG	0.205	0.057	0.116	0.951	1.329	0.571	1.900
VHIG	0.278	0.05	0.284	1.087	1.699	0.409	2.108
All	0.268	0.036	0.048	0.562	0.915	0.990	1.905

The per capita trip rate for motorized modes for 25 Indian cities is observed to be in the range of 0.391 tpcd to 1.493 tpcd (Government of India, 1998), in case of Surat it is 0.915 tpcd. Two – wheelers carry 46% of work trips and 47% of recreation and other purpose trips. 80% of shopping trips, 65% of education trips, and 43% of work trips are performed by non – motorized modes. Bus carries only 0.8% of work trips and 5% of education trips reflecting on very poor state of public transport. A study on modal shares of journey to work by low-income households in Delhi has revealed 60% trips by NMT and 34% trips by mass transport system of bus and rail (Tiwari, 2001). Earlier study on modal share in selected corridors of Indian cities have shown

mass transport carrying more than 70% trips for Mangalore, Cuttack and Coimbtore whereas Surat had 24% trips by public transport (Reddy, 1993). Cars and two – wheelers share was 33% while 20% trips were carried by IPT. It may be observed that there is sharp decline in the public transport trip share in the study area.

Trip Rate: Spatial Impact

Primarily relative spatial distribution of home and activity ends along with the city size, shape and form influence the trip rate and travel for the given socio – economic character of the population. Closer work place encourages lunch trips and community living in ghetto generates frequent social purpose trips. Size of the city is observed to influence the trip rate (Dharmarajan, 1990 & Pillai, 1993). The shape and form have considerable bearing on the proportion of non – home based trips compared to home based trips. The local service level for shopping is shown to influence the number of home based shopping as well as personal business trips (Kalenoja, 1999). The number of non – home based shopping and personal business trips is higher where the local service level is low (Kalenoja, 1999). The present study examines the spatial trip rates on segmental, sectoral and zonal levels. The effects of temporal city expansion can be realised through segmental analysis whereas sectoral pattern reflects on the effect of macro level land use pattern.

Segmental Impact. The study area is disaggregated on spatio – temporal segments of CBD, inner fringe and outer fringe areas. Population density in CBD, inner fringe and outer fringe is 454 ppha (persons per hector), 301 ppha and 178 ppha respectively. Work trip rates of 0.903 tpcd and 0.882 tpcd observed for inner and outer fringe areas are higher than that for CBD (0.803 tpcd). Similar is the trend in case of school, shopping as well as recreation and other purpose trips. Variations in socio – economic characteristics are mainly responsible for this.

Sector Impact. Sectors are the municipal administrative zones. They constitute macro level units of city land use and settlement pattern. It is observed that North and East sectors have highest work trip rates of 0.928 tpcd and 1.01 tpcd respectively reflecting mixed land use induced proximity between home and work place. The trip rate of 0.848 tpcd in West sector is due to higher number of working members in a household. While weekly shopping trips are varying in a narrow range of 22%, monthly trip rates indicate wide variation.

Zonal Impact. The micro level analysis of trip rate reveals work trip rate varying from 0.572 tpcd to 1.088 tpcd under the influence of residential and work place location behaviour. It is observed to be higher for two reasons one; proximity to work place and two; more number of employed persons in a household. School trip rate is observed to vary from 0.327 tpcd to 0.655 tpcd. Zones with better socio – economic scenario as well as accessibility have shown higher school trip rates. Vast variation in zonal weekly shopping trip rate is observed. Zone 18 is observed to have 0.047 tpcd weekly shopping trip whereas zone 29 has 0.487 tpcd. Similarly, monthly shopping trips are observed to vary from 0.006 tpcd to 0.126 tpcd. Frequent trips for

recreation and other purpose travel are observed higher in magnitude in case of zones in inner and outer fringe in the range of 0.11 tpcd to 0.262 tpcd.

Conclusion

Urbanisation in India has considerable impact on travel pattern and urban transport system. Present study is useful in realization of disaggregate urban travel behaviour. Trip rates analysed in the study are considered as inputs to the MLR and ANN models and also in formulation of transport policies. Household characteristics particularly of income level have significant bearing on the trip rate. Purpose wise and mode wise rates are further under the influence of socio – economic characteristics of households. Segmental growth pattern and land use at sectoral level do matter in urban trips.

References

Bagade, M.V. (2001). "Mobility of masses – issues and options." *J. Transport Management*, 25(4), 291 – 300.

Chandrashekhar, B. P., and Prasad, C.S.R.K. (2005)."Urban public transportation commuter perceptions on APSRTC bus services." *Urban Transport J.*, IUTI, 6(1), 60 – 69.

Dharamrajan, K. (1990). "Urban transport in India." *Proc., National Conference on Transportation System Studies*, I.I.T. Delhi, New Delhi, 269 – 281.

Government of India, MOUAE and RITES (1998). *Study on traffic and transportation policies and strategies in urban areas in India*, Final Report, New Delhi.

Kalenoja, H. (1999). "Spatial differences in the trip generation and travel behaviour – empirical observations in Tampere region." <http://www.tft.lth.se/kfbkonf/5Kalenoja.pdf> (Nov.10, 2005).

Pillai, K. S. (1993). "Land use and transportation integration." *Proc., Urban Transportation National Seminar*, Indian National Academy of Engineering, New Delhi, II-65 – II-71.

Reddy, T. S., and Bhatia, N. L. (1993). "Urban transportation systems planning for cities in India." *Proc., Urban Transportation National Seminar*, Indian National Academy of Engineering, New Delhi, II-33 – II-43.

Sachdeva, Y. (2004). "Urbanisation and urban transport." *Urban Transport Development Training Course Material*, IUTI, New Delhi, 1- 20.

Suryanarayana, Y. and Mukhopodhyay, D. (1990). "A case study of mobility levels of population groups in Ahmedabad city by factor analysis." *Proc., National Conference on Transportation system Studies*, I. I.T. Delhi, New Delhi, 94 – 99.

Tiwari, G. (2001). "Urban traffic management – a case study of Delhi." *Urban Transport J.*, 2(2), 12 – 30.

Discrete Choice Modeling with Partially Missing Information from a Revealed Preference Survey

Quentin K. Wan and Hong K. Lo[1]

[1] Department of Civil Engineering, Hong Kong University of Science and Technology, Hong Kong, China; PH (852) 2358-8742; FAX (852) 2358-1534; email: cehklo@ust.hk

ABSTRACT

Survey data, collected to calibrate discrete choice models, are often with partially missing information on some attributes. Ways to fill in missing data are needed in order to apply standard calibration procedures; however, this may result in seriously biased estimates. In this paper, we propose a calibration approach that takes advantage of the information of the observed attributes as well as to infer the missing data statistics. In doing so, we show in a simulated dataset that the calibration can produce fairly good estimate and correctly account for tendency of bias between users and non-users.

1. INTRODUCTION

Mode choice is a core issue in transportation planning and modeling and forms the basis of travel demand modeling. Indeed, mode choice analyses in the form of random utility maximization (RUM) models abound in the literature as well as in transportation applications. RUM models express the choice probability of a specific mode as a function of its indirect utility, which is in turn derived from the perceived level-of-service characteristics of the mode. Examples include the logit and probit models; more advanced ones include the generalized extreme value (GEV) and mixed logit (ML) models (e.g., Ben-Akiva and Lerman, 1985; Train, 2003). These advanced models, undoubtedly, aim to more accurately model travelers' mode choice behavior.

Surveys are conducted to extract information from travelers to calibrate the discrete choice models, including the perceived characteristics of all available modes. This is achievable in stated preference (SP) surveys but generally not in revealed preference (RP) surveys. Most often, only details of the chosen transport mode can be collected from travelers in an RP survey; the perceived characteristics of the unselected modes are simply missing, or censored. Thus, it raises the issue of compatibility in the calibration.

Current practice handles the issue in two different ways. The first one fills in the true attribute values (e.g. out-of-pocket cost), or the expected values (e.g. travel time), for all

modes. However, this approach takes no account of the considerable information collected from respondents. The other approach is to fill in surrogates, or imputed values, whenever the attribute is missing. Common imputation methods include filling in true or expected attribute values, or the observed means, for the surrogates. The second approach, in contrast to the other approach, does attempt to make use of available information. However, it does not consider possible perception biases. Therefore, the use of surrogates in the place of missing information is questionable.

How should a model be calibrated when some attributes are missing from the empirical data set? And what is the quality of estimate from the current calibration practices? In this paper, we will study the problem of missing data in discrete choice analysis. In Section 2, we discuss the incompatibility issue of RP data in model calibration. With an illustrative simulated case of binary logit model, we demonstrate the biased estimation of existing approaches in Section 3. We will next propose a calibration procedure for RP survey data with partially observed and partially missing information and show the improvement over current practices in Section 4. In the final section, we draw some concluding remarks and suggest future researches. As a result, this study should provide transport practitioners and academics a better tool to understand and model traveler choice behavior with RP survey data.

2. MISSING INFORMATION IN MODEL ESTIMATION

Model calibration refers to recovering the true utility parameters of a correctly specified random utility model. In RUM models, the utility is random and is usually specified as the sum of a systematic component and a random error term, i.e. $U_{ni} = V_{ni} + \varepsilon_{ni}$, where U_{ni} is the utility individual n associate to alternative i in the choice set, V_{ni} and ε_{ni} are the systematical and random components respectively such that $E(U_{ni}) = V_{ni}$ and $E(\varepsilon_{ni}) = 0$. In most applications, linear-in-parameter utility function is specified. These parameters show how, and in what extent, different attributes affect an individual's preference and hence mode choice decision. As a result, it is always the aim to obtain unbiased and efficient estimates of these parameters. Maximum likelihood (ML) estimation is the most popular technique since it provides, with the advent of computing, an elegant platform applicable for all discrete choice models. Moreover, it yields estimates which are consistent and asymptotically efficient, and provides estimates of asymptotic covariance matrix that make hypothesis tests on the estimates feasible.

An issue relating to the information and estimation of discrete choice analysis arises. In developing estimation methods, an implicit assumption is commonly made to meet the precondition of full information. The full information consists of, for each individual, the choice among a set of alternatives, the alternatives' attribute values and the individual's socioeconomic characteristics. However, revealed information is seldom full because of the limited time to finish all questions on all attribute of all modes, and the cognitive burden to report on unfamiliar attributes of other modes not chosen. When a study considers many modes and attributes, it is impractical to collect full information from RP surveys. Instead, in a practical field survey it is common to only record the chosen mode, the attributes values of this mode and the socioeconomic characteristics. Erdem *et. al.* (1999) reported a similar case in which prices were recorded for brought

items but not the other commodity. This may translate into quite a portion of the full information being censored. RP data with partially missing data can not be used as it is. Procedures to fill in the missing values are commonly carried out to provide a set of full information in order to apply the available estimation methods.

3. BIASED ESTIMATION OF EXISTING CALIBRATION APPROACHES

In this section, we demonstrate that existing calibration approaches (*see, e.g.*, Train, 2003) would produce biased estimates. A simple example suffices for such purpose. We consider a binary logit example between modes A and B based only on a single attribute without significant effect due to the individuals' socioeconomic consideration. The choice probability of individual i in a taste homogeneous population is

$$P_i(A) = 1 - P_i(B) = P_i = \frac{e^{V_{iA}}}{e^{V_{iA}} + e^{V_{iB}}} = \left[1 + e^{-\beta_0 + \beta_1 (x_{iB} - x_{iA})}\right]^{-1} \quad (1)$$

where β_0 is the alternative specific constant (ASC) of mode A relative to mode B, β_1 is the generic utility parameter of the single attribute x. When all the attributes are known, we can compute the choice probabilities in (1). Hence the likelihood of a sample is

$$L^* = \prod_i P_i^{y_i} (1 - P_i)^{1 - y_i} \quad (2)$$

where $y_i = 1$ if i chooses mode A, and zero otherwise. The ML estimates of β_0 and β_1 are then obtained as the solutions to $\partial L^* / \partial \boldsymbol{\beta} = 0$.

In order to compare different calibration approaches, we arbitrarily simulated a dataset. We choose to use simulated dataset because this provides us knowledge of both observed and missing attributes for performance evaluation of different approaches. On the other hand, we can also control for the underlying choice mechanism (i.e. according to binary logit) and choice set consideration in the simulated dataset. This avoids model misspecification and all biased estimations are due to the variation of censored attributes. Different distribution assumptions for the attributes are used in two data generation (DG) processes – normal and lognormal distributions. In the normal data generation process the attributes are generated from independent normal distributions such that $x_A \sim N(15, 2.25)$ and $x_B \sim N(20, 4)$. Similarly, the same mean and variance are used in the lognormal data generation process. The realized choices (i.e. values of y_i) are generated according to (1) evaluated at $\beta_0 = -1$ and $\beta_1 = -0.2$ in both data generation processes.

Existing calibrations fill the missing attributes with surrogates to allow probabilities in (1) be computed. We consider four different surrogate schemes, denoted as *M1* to *M4*. In *M1*, any missing attribute is replaced by its true mean. This is possible if an independent source offers such information, *e.g.* the published fare. Erdem *et. al.* (1999) documented such approach in marketing research. *M2* is known as unconditional means imputation (Little and Rubin, 1987) which fills with the observed mean from the RP data. These surrogates are always available but may not fit well to those who are actually not choosing the mode and so not covered in the RP data. The multiple imputation (MI) method is used in *M3* and *M4*. MI is an approach to obtain an estimate as the mean of calibrated parameters from several independent pseudo-complete datasets in which each

fills any missing data by imputed values (Little and Rubin, 1987). In this study, the imputed values are random variables drawn from normal distribution with respective means and variances from the RP data. Each value reported under *M3* is the mean of 5 imputations; and is the mean of 20 imputations in *M4*. The likelihood function in MI is

$$L_{\mathrm{MI}}^{*} = \prod_{i} \left\{ \prod_{m} L^{*}\left(\boldsymbol{\beta} \middle| \mathbf{Y}, \mathbf{X}^{\mathrm{RP}}, \mathbf{X}^{m}\right) \right\} \tag{3}$$

where $\mathbf{Y}$ is the set of y_i, $\mathbf{X}^{\mathrm{RP}}$ is the observed information from RP survey, and $\mathbf{X}^{m}$ is the mth imputation dataset. Likelihood (3) is the usual likelihood except it is now conditional on the expanded (synthetic) dataset with missing values replaced by their imputed values. Table 1 shows the utility parameter estimates of these approaches based on a sample size of 200 individuals. We also include in the column of *M0* the ML estimate with complete data, *i.e.* with both the observed data in RP survey as well as those censored. Clearly such estimates can not be obtained from the reality. It is included so as to account for the finite sample effect of ML estimation, which recovers the true parameters in infinite sample size but not necessarily so in small sample size.

From Table 1, we note that complete case calibration *M0* provides reasonably close estimate to the true parameters for the simulated dataset. This suggests the sample size is not too small for the binary logit model (1). Assuming that external source provides additional information on the true mean of attributes, *M1* replicates parameters as good as *M0*. We also note that *M1* performs better when attributes are normal instead of lognormal, because of the special role of normal distribution in ML estimation theory. In more realistic case when we only have mean statistics from observed sample rather than the true mean, we use them as surrogates in *M2*. However, the generic utility parameter β_1 is basically zero, which means that the attribute x does not explain any choice mechanism at all. Furthermore, the ASC β_0 simply restores the choice probability in binary logit to the market share in RP data. Therefore, the model is not informative. This is due to the lack of variability in data needed to calibrate a discrete choice model. Finally, the MI methods *M3* and *M4* hardly provide satisfactory estimates. In any case, few imputations in *M3* indeed lead to estimates with reversed sign. *M4* with more imputations does not suffer the same problem but still results in very poor estimates. Since more imputations beyond that in *M4* only provides a very marginal improvement in estimate, MI is not a proper method here because of the *non-ignorable missingness* in our scenario (Little and Rubin, 1987). The above analysis demonstrates that most existing practices are unsatisfactory and misleading without full dataset. The only acceptable approach (*M1*) requires additional information from external source, often a common condition as in most suggested methods when there are missing data (*e.g.*, Schafer, 1993).

Table 1. Calibration results of existing approaches

Calibration approach		*M0*	*M1*	*M2*	*M3*	*M4*
Normal DG	β_0	-1.106	-1.051	0.020	0.138	-0.040
	β_1	-0.221	-0.217	0 [Note]	-0.024	-0.013
Lognormal DG	β_0	-1.166	-1.223	0.040	0.157	-0.019
	β_1	-0.236	-0.257	0 [Note]	-0.024	-0.012

[Note] Calibrated values are insignificant and fail to pass null hypothesis.

4. PROPOSED CALIBRATION APPROACH

Calibration usually does not acknowledge that the happenstance of censored attribute itself reveals some information. Attributes may be missing simply by choice behavior of a traveler, or being regarded far worse than it actually is – bias on its attribute exists among some travelers opt not to use the mode. In that regard, it is reasonable to think of the observed and censored attributes are indeed coming from different distributions, and the two together build up the distribution of the attribute among the population. We can also think from the other side of the coin. We can recognize the observed data is actually a realization from the population in a random sampling, hence we can write

$$x = y_i x^{RP} + (1 - y_i) x^{mis} \tag{4}$$

where x, x^{RP} and. x^{mis} refer to the random variables of an attribute in population, RP data, and among missing group, respectively. As before y_i is the choice indicator, therefore, $\mathrm{E}(y_i)$ is the market share of the concerned mode. In the proposed calibration, we assume normal attributes in all 3 forms, and so we can associate the mean and variance of x to those of x^{RP} and. x^{mis}. We can modify the likelihood in (3) to

$$\tilde{L}^*(\theta) = \prod_i \left\{ \phi_{\mathbf{X}}\left(\mathbf{X}^{\mathrm{RP}} \middle| \theta\right) \prod_m L_i^*\left(\boldsymbol{\beta} \middle| \mathbf{Y}, \mathbf{X}^{\mathrm{RP}}, \mathbf{X}^m, \theta\right) \right\} \tag{5}$$

where L_i^* is the likelihood of choice model as before; $\phi_{\mathbf{X}}(\cdot)$ is the likelihood of observed attributes, where $\phi_{\mathbf{X}}$ denotes the normal density function of population distribution of attribute; θ is the vector of likelihood parameter, which includes $\boldsymbol{\beta}$ and mean and variance of x^{mis}. This approach explicitly considers the innate information comes within observed attributes, *i.e.* their likelihoods. Thus, it should produce more reliable estimates.

We apply the proposed approach to the simulated datasets from normal and lognormal DG processes in Section 3. To account for the finite sample size effect, the consistent *M0*-estimates are used as the reference for comparison. In Table 2, η denotes the ratio of percentage discrepancy of the calibrated parameters to that of *M1*-estimates, and roughly speaking, a smaller η has a better convergence to a consistent estimator. The calibration result indicates the proposed approach outperforms *M1*. We anticipate this unless the attributes is deterministic with no variation as implicitly assumed in *M1*. The improvement is due to the consideration of the biases between users and non-users. In the calibration, the means and variances of censored attributes enter into the likelihood as parameters, and so we can also check if it can replicate the actual distribution of censored attributes. Table 3 definitely shows the differences in biases. Though not being accurate, it correctly predicted the trend of bias. This would help better understand choice behaviour due to incorrect projection of market condition. In the proposed approach, the normal distribution is assumed. Therefore, it is desirable to check for its robustness. In Tables 2 and 3, we see fairly good results for lognormal data as in normal data. This shows the possible robustness of the proposed approach.

Table 2. Calibration results of proposed approach

Calibrated Parameters / η	*Normal DG*	*Lognormal DG*
β_0	-1.151 / 81.8%	-1.164 / 3.5%
β_1	-0.219 / 50.0%	-0.228 / 38.1%

Table 3. Calibrated distribution parameters of censored attributes [Note]

Censored attribute	*Normal DG*		*Lognormal DG*	
	Mean	S.D.	Mean	S.D.
x_A	15.416 (15.147)	2.254 (1.351)	15.412 (15.134)	2.228 (1.389)
x_B	21.207 (20.490)	3.479 (2.068)	21.115 (20.535)	3.327 (2.152)

[Note] Sample statistics from respective data generation process are shown in brackets for comparison.

5. CONCLUSIONS & FUTURE RESEARCH

Precise calibration of discrete choice models is the cornerstone of its credibility in mode choice analysis. RP survey is typically conducted to collect information on choice behavior to be used for calibration. In practice, limited resources lead to simple and short RP interviews which result in many missing, or censored, data. Filling in these missing data is necessary for model calibration. We showed that standard approaches of using surrogate mean, or multiple imputation, might produce significantly biased estimates, sometimes with incorrect signs. We proposed an approach in which we make use of the information of the observed data to help recover partial information of the censored attribute. In a simulated dataset, we showed that the approach produced fairly good utility parameters and correctly accounted for tendency of biased view among the non-users. Possible robustness of the approach is also seen in the case study.

For simplicity, we have assumed normal distribution in the core of the proposed approach. However, further examination is needed for extending the use of other distributions. A closer look at the biases between the user and non-user group should provide a better understanding and perhaps a better design for calibration. In this study, we considered a binary logit model with one attribute; we anticipate greater difficulty in extending it to more sophisticated discrete choice models as well as more complicated choice interactions in the reality. Therefore, specific optimization procedures should be developed for calibration.

ACKNOWLEDGEMENTS

This study is sponsored by the Competitive Earmarked Research Grants HKUST6083/00E and HKUST6161/02E of the Hong Kong Research Grant Council.

REFERENCES

Ben-Akiva, M. and Lerman, S.R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand,* MIT Press, Cambridge, MA..

Erdem, T., Keane M.P. and Sun, B. (1999). "Missing price and coupon availability data in scanner panels: Correcting for the self-selection bias in choice model parameters." *Journal of Econometrics*, 89(1-2), 177-196.

Little, R.J.A. and Rubin D.B. (1987). *Statistical Analysis with Missing Data*, Wiley, NY.

Schafer, D.W. (1993). "Likelihood analysis for probit regression with measurement errors." *Biometrika*, 80(4), 899-904.

Train, K.E. (2003). *Discrete Choice Methods with Simulation*, Cambridge University Press, NY.

Computer-assisted Scheduling and Dispatching Systems for Paratransit Transportation: An Assessment of Agency Readiness in Planning for Statewide Deployment

P. Metaxatos[1] and A.M. Pagano[2]

[1]Urban Transportation Center (M/C 357), College of Urban Planning and Public Affairs, University of Illinois at Chicago, 412 South Peoria Street, Suite 340, Chicago, IL 60607; PH (312) 996-4713; FAX (312) 413-0006; email: pavlos@uic.edu
[2]College of Business Administration (M/C 243), University of Illinois at Chicago, 2218 University Hall, 601 S. Morgan St., Chicago, IL 60607; PH (312) 996-8063; FAX (312) 996-0773;email:amp@uic.edu

Abstract

Computer-assisted scheduling and dispatching (CASD) systems have arguably shown to facilitate operational improvements and coordination in paratransit. However, the level of realized positive impacts depends on the specifics of each implementation plan and overall deployment strategy, as well as the attitudes of potential users. In addition, the concerns of paratransit providers about implementing new software given their level of technology readiness need to be addressed. This paper examines the state of practice technology readiness of paratransit providers in the State of Illinois. It is shown that for most operators, implementing CASD systems would require developing new computer skills and training existing personnel.

Introduction

As populations in the United States age, the need for public paratransit services is increasing. Federal funding for such services has increased as well under the expectation that states will spend those transportation grants efficiently and effectively. Accordingly, transportation providers will need to consider whatever efficiencies can be gained through operational improvements and coordination with other providers. Both these objectives can be facilitated through the implementation of computer-assisted scheduling and dispatching (CASD) systems.

The deployment of such systems has shown measurable efficiency, effectiveness and quality of service gains and confirmed expectations that CASD systems provide improvements in dispatching, scheduling, on-time performance and increased passenger satisfaction (Alfa, 1986; Bennett, 1994; Stone et. al, 1994; Chira-Chavala et al., 1997; Pagano et al., 2002; Metaxatos and Pagano, 2004). However, the level of realized positive impacts depends on the specifics of each implementation plan and overall deployment strategy, as well as the attitudes of potential users (Kikuchi, 1988; Pagano et al., 2003).

With increasing numbers of local CASD system implementations, state departments of transportation have only recently realized the need for statewide strategies regarding future deployments of such

systems. In this regard, at the planning stage of statewide CASD deployment programs, the concerns of paratransit providers about implementing new software given their level of technology readiness need to be addressed.

This paper examines the state of practice and technology readiness of paratransit providers in the State of Illinois. It is shown that for most operators, implementing CASD systems would require developing new computer skills and training existing personnel.

Methodology

Nineteen agencies were chosen for on-site visits. They were located throughout the state of Illinois, and were chosen based on whether they received federal (5311) funding and the size of their fleet. At least two to three providers were chosen from each of the size groups. In addition, we considered the ways in which various organizations including hospitals, retirement homes, and taxi companies provided transportation service. We included a provider from each category.

We developed several survey instruments (see Pagano et al., 2003 for details) to determine the operations performed by each individual agency and how they were performed. First, a management survey was used to obtain information from the perspectives of the supervisors and other management personnel. This survey instrument included questions related to management's familiarity with computers, their daily responsibilities, their opinions on service coordination, and their level of comfort with the idea of a CASD system.

Secondly, a dispatcher survey was created to evaluate the knowledge and skill levels of the call takers and schedulers within the agencies. These surveys included questions related to the scheduling and dispatching personnel's day-to-day activities, their technical sophistication and enthusiasm toward a CASD program.

Finally, a technology survey was used to conduct an assessment of the agency's use of technology. The computing facilities and other technology in each organization were inventoried, and the level of computer complexity used in each organization was noted.

Analysis of Agencies

Agency Characteristics: The fleet size ranged from 31 to 58 vehicles in large agencies, from 11 to 30 vehicles in mid-size agencies, and from 1 to 10 vehicles in small agencies. The average fleet size of all agencies surveyed was 16 vehicles, and the average number of trips dispatched per day was roughly 300.

Smaller agencies reported providing anywhere from 30 to180 one-way trips per day, and only one or two demand responsive trips per agency per day. The larger agencies provided anywhere from 600 to 1000 subscription trips per day, with typically 25% of the trips being demand responsive. Medium sized operators had actually a very small focus on the transportation element of their business.

Services Provided: Subscription service and contract routes constitute the majority of service provided. Demand responsive service constitutes 25% of the typical providers' business. All of the agencies provided service within one county or a neighboring county in their service area. Larger regional operators have a greater service area.

All agencies provide a call taking and scheduling procedure. Typically, for the subscription service, the run sheets are developed a week in advance. The demand response calls are dealt with differently (see Pagano et al., 2003 for details). Agencies reported having, on average, two new customers per week. The process often generates a lot of unnecessary and duplicate paperwork.

Reporting, Scheduling and Dispatching: A closer look at the way agencies perform their operations revealed three types of operations. The first type, the most complex, uses CASD software to carry out all operations. Basically, operators fill out the screens, and the system then inserts the schedule and dispatches the run. This is done in batch mode for subscription trips and "on the fly" for demand-responsive trips. Contact with the drivers is maintained through radio and/or mobile phones. Report generation is a button-click away, since all the information is stored in the CASD system's database and all the necessary reports are built into the system. Currently there are only two agencies that were visited that were using this type of system, and one of them is a taxi service.

The second type uses both computers and manual processes. Computers are mainly used for storing information about customers, trips, and vehicles. Scheduling and dispatching are done manually. Reports are generated easily since the data is stored in electronic format. However, some reports, that require summarized data or span longer periods of time, might take more time to prepare because there are no readily available reports built into the system. Communication with drivers is mainly done with two-way radios and mobile phones or payphones.

In the last type, operations are performed manually. In this case, operators take the same information, but use ledgers or "run" sheets to store client information. At the end of each day the scheduler/dispatcher uses this information to generate drivers' manifests, which will be picked up by the drivers the next day. Reports usually take longer to prepare. The process created a large amount of overlap and redundancy. However, communication methods are the same as in the previous two cases.

Billing: Primary billing is a significant issue for transit agencies. Some operators work only with one or two sponsors (funding agencies), while others have as many as thirty. Handling billing manually is manageable for the first category, though a computerized system would make it more efficient and perhaps more accurate. For the second category, billing must be an automated process. Some operators bill manually, keeping ledgers and possibly using word processors or spreadsheets to send them. However, most of them use accounting applications or billing modules of CASD software.

Advance Notice Requirement for Scheduling: All providers preferred to receive requests at least one day before travel was needed, although most were flexible with this policy if short notice trips could be fitted into the day's departures. Larger agencies tended to discourage last-minute requests, suggesting

that in more populous areas, the level of demand was fast approaching what operators can meet with their existing equipment and scheduling systems.

All but the smallest organizations reported having multiple phone lines available to handle more than one customer call at the same time. The larger organizations typically had a holding or queuing system for the calls. A few of the smaller systems reported having only one line available, with one of those having no voice mail or any way for the operator to return a missed call. Most offices with computers available typically had most non-driving employees using computers at some point in their workday.

Technological Preparedness

Software Used: All but one organization reported at least one computer in use at the office. The highest number reported was twenty in the largest operator, with a typical operator using two or three computers. Most offices with computers had some form of computer networking available. Word processing and spreadsheets are the most commonly used applications.

Computer Support: Small providers typically had no specific arrangements for external computer support, often relying on an informal network or volunteers to handle computing needs. The larger operators tended to have some sort of contracted support available, which was either through their corporate office or through a licensed software/hardware distributor. Ten out of the nineteen agencies had access to either one or two in-house computer experts who could address most support needs, or were able to contact a neighboring college student for further assistance. None of the interviewees reported any trouble getting computer support when needed.

Computer Skills/Expertise: Most of the agencies used computers for billing and reporting, at the very least. Three managers reported that employees had very minimal computer skills. For the rest of the agencies, the managers were generally satisfied that the employees' skills matched current job requirements, and a few were very complimentary of their employees' skills. The general feeling towards the future implementation of a CASD system is positive, because most are willing to be attend further training sessions in order to become familiar with computers and the system.

Unsurprisingly, respondents with higher levels of computer expertise were more likely to have a higher level of acceptance of new systems. Those who reported a low level of computer expertise tended to also think negatively about CASD systems, and were also less favorable toward attending training sessions. Of the 19 agencies surveyed, eleven reported a strong disinclination towards implementing CASD.

Managerial Computer Expertise: Implementing a CASD system is a complex undertaking, which in most cases results in re-engineering the call taking, scheduling, and dispatching processes. Completely (or significantly) changing such processes requires management support (sometimes called executive sponsorship) to result in a successful transition. Therefore, it is important to understand management's capacity to support change.

To evaluate this capacity to support change, managers were asked whether they had overseen computer system implementations, and if they did, what was the outcome. Sixty percent of the interviewees had overseen or participated in implementations related to computer systems, but only half of these were convinced that they could successfully manage any other implementation. The remaining 40% have never overseen any such implementation. For those who have overseen implementation, they expressed problems with the software that was installed.

Of all the operators surveyed, 30% would be able to manage an implementation, and 70% would need some external guidance in order to guarantee success. This support could either come from either IDOT or a third party (university, consultant, etc.).

Hardware and Software Needs: Paratransit software is expensive. Agencies using CASD systems often lease it because they do not have sufficient funds to buy it outright. Considering the potential funds available, they were asked how would they spend these. Not surprisingly, of the agencies having a CASD system, most would purchase software or would invest more in hardware. If the funds were to be spent on software they would invest in the same software they have (or upgrades). One agency customized Microsoft Outlook for the use of scheduling, maintenance, and billing. It was quite useful to them.

Other agencies would use the funds as well, but it was not clear how they would spend the money, nor did they seem interested in purchasing a CASD system. Among these, several expressed their interest in CASD systems. They also asked for help and guidance from IDOT (or external entities) since they did not have enough expertise.

Technology Assessment

The level of technological sophistication varied widely among agencies. The more sophisticated agencies used computers not only for constructing driver manifests and customer information, but also for services provided by the agency such as counseling services and maintenance. At the other end of the spectrum were the less sophisticated agencies. Since transportation of people with disabilities and the elderly is such a small component for the majority of these agencies' day-to-day activities, computers were rarely used. Most of the records were kept manually and frequently updated.

Based on a technology assessment inventory (Pagano et al., 2003, Appendix IV), an evaluation of an agency's 'technology readiness' was made to assess the need to upgrade the hardware, if CASD systems will be installed. The technology readiness was scored on a number of attributes related to the surveyed operator's technological capabilities including: (1) number of computers used in the agency, (2) computer configuration (RAM, HDD, drives, monitor, etc.), (3) operating system, (4) networking capabilities (existing network), (5) internet capabilities (modem, high speed), (6) available peripherals (printers, scanners, etc.), and (7) overall technological preparedness. The scoring procedure used factor analysis the details of which we omit because of space limitations. Three categories of readiness were then identified: poor, fair, and good.

Agencies classified as 'poor' need major hardware upgrades because the existing one is either outdated or must be used for other purposes. Agencies with a 'fair' score need upgrades such as operating systems, network cards, and new monitors. Agencies given a 'good' score are technologically up-to-date needing only minor changes (proper network configurations, setting up servers, etc.) prior to deploying the CASD software. It is interesting to note the high positive correlation between operation size and technological sophistication. Most of the agencies (except for two) that fall in the 'good' category have more than 300 trips/day. The two exceptions have between 100 and 300 trips/day.

Conclusions

This paper presented findings from a survey of operators and revealed the following themes: (1) For most operators, implementing CASD would require developing new computer skills and training existing personnel; (2) Most agencies providing paratransit services in rural areas are not centered around transportation; (3) Reporting is the bane of paratransit agencies; (4) CASD implementation should allow data sharing between CASD systems and legacy accounting systems; and (5) CASD system installations must be preceded by computer upgrades. Before a successful statewide deployment of CASD systems is attempted, these issues that impact the daily operations of interested paratransit providers should be addressed.

References

Alfa, A.S. (1986). Scheduling of Vehicles for Transportation of Elderly. *Transportation Planning and Technology*, 11(3), 203-212.

Bennett, S. (1994). A Decision Support Approach to The Computerization of Some Operational Decision Problems in Community Transport. *Transportation Planning And Technology*, 18(4), 307-330.

Chira-Chavala, T., Gosling, G.D., and Venter, C. (1997). *Advanced Paratransit System: An Application of Digital Map, Automated Vehicle Scheduling and Vehicle Location Systems.* Research Report, University of California, Berkeley, UCB-ITS-RR-97-1.

Kikuchi, S. (1988). Development of Computerized Vehicle Scheduling System for Specialized Transportation: Delaware Case. *Transportation Planning and Technology*, 12(2), 151-170.

Metaxatos, P. and A.M. Pagano, A.M. (2004). Efficiency, Effectiveness Impacts of a Computer-Assisted Scheduling and Dispatching System Implementation. *Journal of the Transportation Research Forum*, 43(1), 77-90.

Pagano, A.M., Metaxatos, P. and King. M. (2002). Effect of Computer-Assisted Scheduling and Dispatching Systems on Paratransit Service Quality. *Transportation Research Record 1791*, 51-58.

Pagano, A., Metaxatos, P., Holeman, E., Mora, V., Morreale, A., and Stanis, K. (2003). *Strategic Plan: Computer Assisted Scheduling and Dispatching System.* Final Report. Illinois Department of Transportation, May 2003.

Stone, John R., Nalevanko, Anna M., Gilbert, Gordon 1994). Computer Dispatch And Scheduling For Paratransit: An Application of Advanced Public Transportation Systems. *Transportation Quarterly*, 48(2), p. 173-183.

Time-Dependent Transport Network Design and Urban Development

Barbara W.Y. SIU and Hong K. LO[1]

[1] Department of Civil Engineering, Hong Kong University of Science and Technology, Hong Kong, China; PH: (852) 2358-8742; Fax: (852) 2358-1534; email: cehklo@ust.hk

Abstract

In the long run, there will be changes in the population, socio-economic factors, land-use, etc. within a region. To cope with these changes, it will be necessary to improve the transportation network and alter the tolls, if any, accordingly over time. In this study, we extend the current technique in network design as a static, one-time event to an approach that explicitly captures the time dimension in the network design problem. Thus, we can make optimal decisions on project initiation time, scaling, and phasing in accordance to the design objectives. Through a small network example, we demonstrated that a well-planned, sequential upgrade plan performs better than any one-time network improvements.

1. Introduction

Transportation infrastructure plays an important role the development of a region. It is important to address these questions: Whether and to what extent a region should upgrade its transportation infrastructure? Traditionally, we rely on the technique of transport network design to answer this question. Equally important is the question of timing and phasing. When is the appropriate time to implement these projects? Shall we build a full-blown network for future generations or should we progressively upgrade the network? The temporal dimension of transport network design and improvement is an important consideration.

In the long run, there will be changes in population, socio-economic factors, land-use, etc. within a region. To cope with these changes, it will be necessary to improve the transportation network and alter the tolls accordingly over time. In this study, we extend the current technique in network design as a static, one-time event to an approach that explicitly captures the time dimension within the network design problem. Thus, we can make optimal decisions on project initiation time, scaling, and phasing according to the design objectives.

We formulate this problem as a time dependent Network Design Problem (NDP) via the technique of bi-level programming. In this formulation, the upper level optimizes for a certain social objective, such as minimization in total travel costs, maximization of reserve capacity for future demand, maximization in consumer surplus, or a mixture of these measures. The lower level captures the time-dependent

traffic assignment conditions. To simply the exposition, we model the entire planning horizon by several discrete phases. Within each phase, we adopt the user equilibrium conditions for traffic assignment. Some pioneering work in this area has been done by Lo and Szeto (2004) and Szeto and Lo (2005).

This study considers that workers, in fixed residential zones, are willing to change their work locations. This assumption is for simplicity of illustration; relaxing this assumption is readily achievable within this modeling framework. Given zonal population and location attraction function, we derive the traffic demand and solve the lower level traffic assignment problem. Having the social welfare measure (e.g. consumer surplus) computed and discounted for time, we optimize the social welfare over the entire planning horizon subject to the equilibrium traffic assignment constraints.

2. Formulation

Consider a planning horizon $(0,T]$, wherein all the links are built at time 0 (initial time) and capacity expansion decisions are considered at discrete time instances $\tau=1,2,...,T$, where the time interval between each τ is n years. Given the population in zone i at time τ, we distribute population to employment zones j according to a location attraction function, which depends on the employment opportunities. Furthermore, the working population selects their work locations so as to maximize their utility as a tradeoff between travel time and job opportunity. By iteratively updating the total trip ends at employment centre j and hence the attraction of zone j, we obtain the equilibrium activity and link utilization pattern. In this study, for simplicity, we assume that zonal employments are unbounded.

2.1 Definition

Consider a transportation network modeled by a directed graph $G(N,A)$, where N is the set of nodes and A is the set of links. We use the superscript τ to denote the variables at time τ. Each link $a \in A$ has a flow-dependent travel cost $t_a=t_a(x_a^\tau,c_a^\tau)$, with capacity c_a^τ, which can be expanded over time. Let I denotes the subset of origin zones (trip-production/residential zones) and J be the subset of destination zones (trip-attraction/employment zones), I and J need not to be mutually exclusive. P_{ij} denotes the set of paths between OD pair (i,j), $i \in I$, $j \in J$. Let O_i be the population at zone i and be the employment in zone j. q_{ij}^τ is traffic demand between zone i and j at time τ out of which $f_p^{ij,\tau}$ is via path p, $p \in P_{ij}$. For each destination zone j, we define a destination cost function $c_j(\tau) = c_j(D_j(\tau))$, which is an increasing function of D_j.

2.2 Stochastic Equilibrium Route Choice Behavior

Each traveler seeks a route p between OD pair (i,j) to maximize his random utility. Each route's utility is assumed to be a i.i.d. Gumbel variate. For simplicity, the logit-based route choice is obtained (Sheffi, 1985):

$$f_p^{ij,\tau} = q_{ij}^\tau \cdot \frac{\exp\left(-\theta t_p^{ij,\tau}\right)}{\sum_{r \in P_{ij}} \exp\left(-\theta t_r^{ij\tau}\right)} \qquad \forall\, p \in P_{ij},\ i \in I,\ j \in J,\ \tau = 0,1,\ldots,T \tag{1}$$

the positive parameter θ measures the sensitivity of route choice to travel cost. The i.i.d. assumption may lead to route flow inaccuracies as are well-known, but can be avoided with more elaborate nested logit structures. Under the logit-based SUE route choices, the expected minimum travel time perceived by an individual traveling between OD pair (i,j) can be expressed as:

$$s_{ij}^{\tau} = -\frac{1}{\theta}\ln\sum_{r\in P_{ij}}\exp\left(-\theta t_p^{ij,\tau}\right) \quad \forall\, i\in I, j\in J,\ \tau = 0,1,\ldots T \tag{2}$$

2.3 Stochastic Equilibrium Location Choice Behavior

Each person, living in a fixed zone i, chooses his job location j with the same the logit-based model as in Section 2.2:

$$q_{ij}^{\tau} = O_i^{\tau}\cdot\frac{\exp\left[-\mu\left(s_{ij}^{\tau}+c_j^{\tau}\right)\right]}{\sum_{k\in J}\exp\left[-\mu\left(s_{ik}^{\tau}+c_k^{\tau}\right)\right]} \quad \forall\, i\in I, j\in J,\ \tau = 0,1,\ldots T \tag{3}$$

the parameter μ is related to the standard deviation of the random error term in (2). O_i^{τ} is the population in zone i at time τ, which grows at g% p.a.. Summing all trips ending at zone j, we obtain D_j^{τ}, i.e., the total employment in zone j:

$$D_j^{\tau} = \sum_i q_{ij}^{\tau}, \quad \forall\, j\in J,\ \tau = 0,1,\ldots,T \tag{4}$$

We postulate that the attraction of zone j is an increasing function of D_j^{τ} and the location cost of zone j is assumed to be

$$c_j^{\tau} = -\alpha D_j^{\tau}, \quad \forall\, j\in J,\ \tau = 0,1,\ldots,T \tag{5}$$

with α as a scaling parameter.

2.4 Time Dependent Measures

2.4.1 Total System Travel Time

The total system travel time at time τ is the dot product of link flows x_a^{τ} and link travel time and converted to monetary unit by value-of-time (VOT) parameter ψ:

$$TST^{\tau} = \psi\sum_{a\in A} x_a^{\tau}\cdot t_a\left(x_a^{\tau}, c_a^{\tau}\right) \tag{6}$$

The total system travel time at present value (TST) is obtained by summing all TST^{τ} discounted to time 0 by discount rate r% p.a:

$$TST = \sum_{\tau=0}^{T}\frac{TST^{\tau}}{(1+r)^{n\tau}} \tag{7}$$

2.4.2 Infrastructure Investment Cost

Suppose capacity of link a at time τ is expanded by y_a^{τ} such that the new capacity is $c_a^{\tau} = c_a^{\tau} + y_a^{\tau}$ which involves an infrastructure cost IC^{τ}:

$$IC^{\tau} = \beta\cdot t_a^0 y_a^{\tau}(1+i)^{n\tau}, \tag{8}$$

with β as the unit construction cost parameter, which is to be multiplied to the extent of widening y_a^{τ} and the length of the road measured by t_a^0 or the free flow travel time on link a. The term $(1+i)^{n\tau}$ captures the inflation, which increases the construction

cost by $i\%$ per year in an n year period. Discounting IC^τ to time 0 and summing, we can calculate the total discounted infrastructure investment cost (IC):

$$IC = \sum_{\tau=0}^{T} \frac{IC^\tau}{(1+r)^{n\tau}} \tag{9}$$

The time dependent network design problem can be solved by a bi-level program. The upper level is a minimization program on the sum of TST and IC as defined in (7) and (9), with y_a^τ as decision variable, for all $a \in A$, $\tau=1,2,...,T$. The lower level is the T-period stochastic trip distribution traffic assignment, solving for x_a^τ satisfying the equilibrium conditions (1) and (3) as well as the non-negativity requirements.

3. Numerical Example

Consider a test network as shown in FIGURE 1. Zones 1 and 2 are residential zones and zones 5 and 6 are employment zones. Suppose we plan the network for 20 years, in two 10-year-periods, thus we have $\tau=0,1,2$ and $n=10$. Also, assume that the population grows at $g=2\%$ per year, inflation i ia 1% p.a. and the discount rate r is 4% p.a.. The logit-split parameters θ and μ in (1) and (3) are 1 and 0.2 respectively and VOT ψ is \$2/ per minute. The parameter α in the location attraction function (5) is set to be 0.001, β in the construction cost function (8) is 2000. The initial population in zones 1 and 2 are 50 and 100 respectively.

We solve the optimization program as described in Section 2. The path and link flow solutions are shown in TABLE 1 and TABLE 2. The temporal changes in OD demands and zonal employments are depicted in FIGURE 2 and FIGURE 3.

To illustrate the strength of the time-dependent formulation proposed to the discrete time network design problem, we can compare the results in TABLE 3. The "Do nothing" scenario is to leave the network unaltered since time 0. Plan 1 is we design for the OD demand obtained for time 1 and do the capacity expansion at time 0; similarly, plan 2 is to design for time 2 at time 0. Definitely, if we build the network to the full scale at time 0, the TST over the whole planning horizon would be the lowest, however, given the inflation and discount rate, the benefit of discounted time value of money is not exploited; in addition, we need to incur additional costs in maintaining an under-utilized network, though this is not modeled in this paper. On the other hand, if the sequential upgrade plan is adopted, we can enjoy the benefit of both lowering the TST as well as reducing the infrastructure expenses in present value terms.

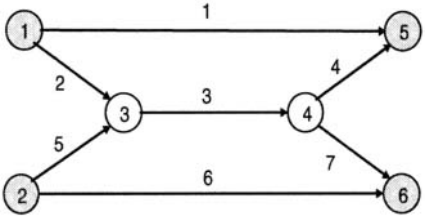


FIGURE 1 The Test Network

TABLE 1 Link Flow Pattern

Route	(O-D)	Link sequence	Time 0 Path Flow	Time 0 Travel Time	Time 1 Path Flow	Time 1 Travel Time	Time 2 Path Flow	Time 2 Travel Time
1	(1,5)	1	28.64	30.48	42.41	32.33	54.87	36.53
2		2-3-4	1.05	33.79	0.02	39.95	0.00	59.71
3	(2,5)	5-3-4	33.86	40.18	48.32	48.41	108.31	79.08
4	(1,6)	2-3-7	20.31	32.58	18.52	33.32	19.42	40.28
5	(2,6)	5-3-7	9.86	38.96	12.27	41.79	0.00	59.65
6		6	56.27	37.22	61.31	40.18	40.29	31.90

TABLE 2 Link Flow Pattern

Link	Free-flow Time	Time 0 Capac-ity	Time 0 Link Flow	Time 0 Travel Time	Time 1 Expan-sion	Time 1 Link Flow	Time 1 Travel Time	Time 2 Expan-sion	Time 2 Link Flow	Time 2 Travel Time
1	30	50	28.64	30.48	0.00	42.41	32.33	0.00	54.87	36.53
2	10	30	21.36	10.39	0.00	18.54	10.22	0.00	19.42	10.26
3	10	80	65.08	10.66	0.00	79.13	11.44	0.00	127.73	19.75
4	10	30	34.91	12.75	1.52	48.34	18.29	25.37	108.31	29.70
5	10	30	43.73	16.77	9.06	60.58	18.69	17.89	108.31	29.64
6	30	50	56.27	37.22	0.00	61.31	40.18	0.00	40.29	31.90
7	10	30	30.17	11.53	0.00	30.79	11.66	0.00	19.42	10.26

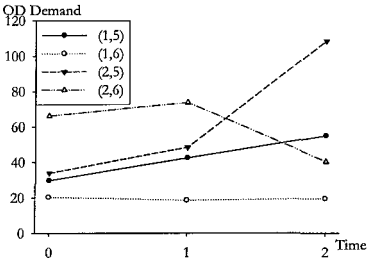


FIGURE 2 OD Demand

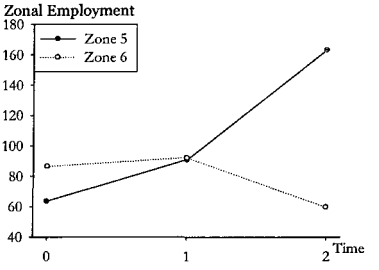


FIGURE 3 Zonal Employment

TABLE 3 Performance of difference schemes

	Do nothing	Plan 1	Plan 2	Optimal Plan
TST	646757	463994	243938	260957
Additional IC	0	21670	108177	65972
Total	646757	485664	352115	326923

4. Concluding Remarks

A time-dependent formulation for the network design problem has been proposed in this study by means of a bi-level program. In the upper level, capacity expansion decisions are considered such that the composite social and economic costs (discounted total system travel time and infrastructure investment) are minimized, whereas the lower lever is a *T*-period distribution assignment problem. Workers choose their work locations based on the perceived minimum travel cost as well as location attractions (which is proportional to employment opportunity of a zone) by a logit-split function, and their route choices are based on the stochastic equilibrium traffic assignment. From the simple network example, we can see the trip OD matrix can change quite substantially over time; hence we can deduce that some currently uncongested links may become congested after several design periods and vice versa. Therefore, we need to carefully spot the most appropriate time and extent of network upgrades. Simply applying a growth factor to the current OD demand or link flow may not suffice in maintaining travel time/congestion at a reasonable level over the planning horizon. In addition, we have shown that with this time-dependent formulation, we can determine the optimal sequential capacity expansion plan which out-performs other discrete time NDP solution plans, this advantage comes from striking an optimal balance between inflation and discount rates as well as building just enough to meet the needs of future traffic.

One remark is that in segmenting the planning horizon into discrete phases, the shorter is each period, the more accurate is the optimal expansion plan, which also requires more computational effort. Finally, we acknowledge that the formulation developed herein is very simplistic: the start-up cost of construction, maintenance cost of infrastructure facilities are not considered; the location attraction function is assumed to depend on employment alone; and there is no limitation on the maximum or minimum zonal employment. In addition, workers are assumed to have fixed residences. All these assumptions can be relaxed, which is the focus of our current studies.

Acknowledgements

The study is supported by the Competitive Earmarked Research Grants HKUST6151/03E from the Research Grants Council of the HKSAR.

References

Sheffi, Y. (1985). *Urban Transportation Networks*. Prentice- Hall, New Jersey.

Szeto, W.Y. and Lo, H. (2004) "Planning transport network improvements over time" *Urban and Regional Transportation Modeling: Essays in Honor of David Boyce*, Chapter 9, 157-176. Edward Elgar.

Szeto, W.Y. and Lo, H. (2005). "Strategies for road network design over time: Robustness under uncertainty". *Transportmetrica*, Vol. 1, No. 1. 47 – 64

Applying cellular-based location data to urban transportation planning

Jiangang Lu[1] , shan Di, Changxuan Pan and Bin Ran

[1] Department of Civil and Environmental Engineering, University of Wisconsin – Madison, Madison, WI 53706; PH (608)262-2524; FAX (608) 262-5199; email: Jiangang@cae.wisc.edu

Department of Civil and Environmental Engineering, University of Wisconsin – Madison, Madison, WI 53706; PH (608)262-2524; FAX (608) 262-5199; email: sdi@cae.wisc.edu

Department of Civil and Environmental Engineering, University of Wisconsin – Madison, Madison, WI 53706; PH (608)262-2524; FAX (608) 262-5199; email: cpan1@wisc.edu

Department of Civil and Environmental Engineering, University of Wisconsin – Madison, Madison, WI 53706; PH (608)262-0052; FAX (608) 262-5199; email: bran@engr.wisc.edu

Abstract

In this paper, we will apply the cellular-based location technology to the urban transportation modeling system (UTMS). We first discuss existing cellular-based location technologies, and then propose a cost-effective and efficient method of using the cells in wireless cellular communications to achieve cellular location information. The cellular location data obtained through this method will be applied to the four-step procedure in UTMS. Comparisons with traditional methods will be made and the advantages of the new method will be presented.

Introduction

Location-based service (LBS) can be defined as an information service that exploits the ability of technology to know where the user is, and to modify the information it presents accordingly. With the development of wireless communication technologies, the cellphone is becoming the most important tool for communication in the United States and the coverage of cellphone users is increasing dramatically, especially in the downtown areas of cities. It is estimated that by the end of year 2005 there will be 200 million cellular subscribers, or 66 percent of the population, in the U.S.. These facts make it possible for us to collect the location information for a large number of cellphone users through the cellular network, and use this information in transportation system planning and management.

With the increase of the coverage of cellphone users among total travelers, the travel behavior information of cellphone users, which we obtain through the communication system, can be used to reflect the travel behavior of all the travelers. Cellular information is available directly from mobile carriers, and it requires less installation cost and labor compared to other method such as conventional loop and resident surveys. In recent years some research has been done based on cellular location information, but most focused on speed estimation and was associated with travel time prediction. Sankar and Civil (1997) proposed a method of using handoffs in wireless cellular communications for monitoring traffic flow and predicting congestion. Akin and Sisiopiku (2002) introduced a way to estimate origin-destination matrices using location information from cellular phones. Unfortunately, this new method is only based on pseudo-cellular-based location data and has only been tested by simulations. Therefore, no previous research has ever used real cellular data, and the application of cellular-based location data in transportation is a very new research area. In our paper, we will extend research in this area to transportation planning fields; specifically, we analyze the real cellular-based location data and integrate them into the urban transportation modeling system (UTMS). As we know,

UTMS is used to predict the number of trips made within an urban area by type, time of day, zonal origin-destination (O-D), travel modes and the routes taken through the transportation network by these trips. The final product of UTMS is a predicted set of modal flows on arcs in a network. Most of the recent work and research in this area, including the analytical method and simulation-based approach, has been focusing on the modeling aspect of UTMS. Strong assumptions have been made to support the implementation of various models. However, such assumptions are hard to be satisfied. Meanwhile, the data needed in transportation planning are generally collected through travel surveys, which are both labor intensive and hard to update. Given the limitations of current transportation planning models, we propose new methods based on cellular-based location data, which can improve transportation planning input data quality, help satisfy assumptions and support the calibration for UTMS model.

Cellular Network and Data Collection

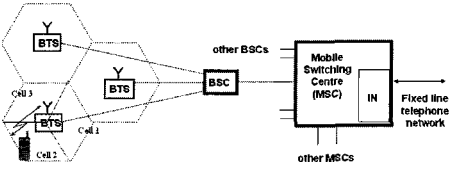


Figure 1. Cellular system structure.

In the cellular system, the cellular network in a city is divided into several areas, each of which is controlled by a Base Transceiver Station (BTS). Furthermore, each BTS is divided into two or three sectors by using directional antennas. Each sector is regarded as a "cell" associated with a unique cell ID across the city cellular system. The upper level of BTS is called Base Station Controller (BSC), and several BSCs are controlled by a Mobile Switching Center (MSC), where we can obtain the cellphone location information. The structure of a cellular system is showed in Fig.1. The coverage of a cell depends on the performance of the BTS, the land-use property, the topology and buildings around the BTS tower. In rural areas, the distance between two towers can be up to several kilometers. However, in urban areas, the radius of cell sector is less than 400 meters, usually between 200 to 300 meters, and in the central city, it's less than 100 meters. Any cellular phone in power-on status is associated with one and only one cell. As cellphone users travel among different cells, the cellphone will automatically scan the control channels and switch to the new cell. In this way, the cell ID can be a description of the geographic location of the cellular user. If there is any cellphone activity, such as power turned on/off by users, or a communication between a single phone and the cellular system, such as calling another phone, sending/receiving text messages, a record will be saved in the MSC server. The records saved in the MSC server include phone ID, cell ID, timestamp, and reason for update. To optimize the cellular system performance, the records in the MSC server will be deleted from the database after the cellphone activity is finished. A probe card is needed to download the data from the MSC server to a local data retrieving server.

Application in Urban Transportation Planning

Trip Generation

A trip generation model is used to determine the number of trips originating and terminating in different zones of the study area. The objectives of trip generation include an estimation of the amount of trips out of a traffic analysis zone (TAZ) into another TAZ, accounting for differences among TAZs due to person, household and business characteristics, and developing functions to predict the future number of trips. In our research, we are interested in how to identify the amount of trips out of or into a specific TAZ. Generally, a TAZ is an arbitrary subdivision of the study area with boundaries including the regional network area and natural or manmade dividers, such as canals and railroads, which naturally limit opportunities for trip crossings. The boundaries of cells are set according to similar rules, and the size of a cell is generally smaller than the size of a TAZ. Therefore, in cases TAZs have been designed in advance. Let P denote the zonal population, C the number of cellphone users and β the percentage of cellphone users in the total population. The relation between P and C can be expressed as $P = \frac{C}{\beta}$. Hence, with the information about the number of cellphone users leaving or entering each zone, we can estimate the number of trips that originate or terminate in each of the TAZs. To obtain information about the quantity of inter-zonal trips by cellphone users, two important characteristics of cellular network systems, shown as follows, need to be paid attention to. First, each cellphone in the cellular network system is corresponding to a pseudo ID. The MSC server keeps collecting updated records for each pseudo ID. Every record includes the pseudo ID and the ID of the cell the cellphone is located at. Second, when some cellphone activity occurs, a message will be sent to the MSC server, which then saved an updated record. Meanwhile, the cellphone also periodically communicates with the MSC server to update its information. The cellular carriers have the option to set the communication cycle. Generally it is set to 2 hours by default in the currently popular GSM cellular network; that is, the MSC server will refresh the information for each cellphone at least once every two hours. We divide one day into several time intervals; each interval is 2 hours. Based on the cell ID information in the MSC server, we can first obtain trip productions and attractions generated by cellphone users in each zone during each time interval. Then we can estimate the total trip productions and attractions for each zone during each time interval, according to cellphone use percentages. We collect this kind of information continuously for a few days and then use statistical methods to analyze the data. Let $CT_i(t)(t = 1,2,\cdots,12)$ and $CA_i(t)(t = 1,2,\cdots,12)$ denote the trip production and trip attraction generated by cellphone users in zone i during the tth time interval. We can estimate the total trip production and attraction in each zone as follows, $T_i(t) = \frac{E(CT_i(t))}{\beta}$ for each $t = 1,2,\cdots,12$, $A_i(t) = \frac{E(CA_i(t))}{\beta}$ for each $t = 1,2,\cdots,12$, where $T_i(t)$ and $A_i(t)$ are trip production and trip attraction for zone i during the tth interval.

Trip Distribution

The purpose of a trip distribution model is to "distribute" or "link up" the zonal trip ends, that is, the productions and attractions for each zone as predicted by the trip generation model in order to predict the flow of trips T_{ij} from each production zone i to each attraction zone j. Traditional trip distribution techniques used for estimating future origin-destination (O-D) flows can be divided into two categories: growth factor (or analogy) methods, and inter-area travel (or synthetic) methods. The gravity model is the most commonly used model; it has been around for over 100 years. In this section, we will compare the cellular-based approach with the gravity model. The gravity model has been

criticized for its behavioral implications; the analogue with physical systems cannot be taken for granted when dealing with human systems. Trip decision is not only related to characteristics of origin zone i, destination zone j and the travel cost. Some trips are pre-determined or under-planned. Just as the home-work trips, travelers will not easily change their destination according to the travel cost. At the same time, few experiments have been implemented to test this approach. By analyzing cellular-based location data, O-D flows can be obtained in a simple way as the trip generation data. We study two time intervals denoted by A and B for one weekday. Division of time intervals needs to be based on properties of the origin and destination zones. As a typical example, for the home-based trips, we assume that most travelers stay at home between 10:30 PM at night and 6:30 AM in the next morning (time interval A), and they will be at the workplaces or schools between 8:30 AM and 11:30 AM in the morning (time interval B). Because cellphones communicate with the MSC server at least once every two hours and the MSC server will save the pseudo-ID and updated cell ID information to the database, we can collect cellphone user information at each zone during time intervals A and B. By aggregating the cellular records in the MSC server, we generate a trip distribution table for cellphone users staying at zone i during time interval A and being at zone j during time interval B. Because of the high percentage of cellphone users in total travelers, we estimate the overall trip distribution table as follows: $P_{ij} = \frac{CT_{ij}}{\sum_{k \in J} CT_{ik}}$, $T_{ij} = T_i \times P_{ij}$, where P_{ij} is the probability of choosing zone j as his destination for a trip maker in zone i; CT_{ij} is the number of trips made by cellphone users from zone i to j; T_i is total number of trips produced in zone i, which we obtained from the trip generation; T_{ij} is the number of trips from zone i to j; J is the set of TAZ indices. Since the cellular-based data are easy to collect, we can accumulate historical data for a longer period, conduct statistical analysis and then obtain the trip distribution by $P_{ij} = \frac{E(CT_{ij})}{\sum_{k \in J} E(CT_{ik})}$, $T_{ij} = T_i \times P_{ij}$.

Modal Split

Modal split models predict the percentage of trips using each of the modes available to the given trip makers. We generally combine all the available modes into two dichotomous modes, transit and auto. Traditional methods, such as diversion curve, regression techniques, generalized costs and discriminant analysis, have been introduced to predict modal split. However, obvious drawbacks exist in current modal split models. For example, the characteristics of travelers who make the journey have been ignored. To apply the cellular-based location data to the modal split process, we need to pay attention to some characteristics associated with the cellular network system. First, if a cellphone is in use, the MSC server will save a record when the cellphone user travels from one cell to another cell. Second, the records include the pseudo ID and updated time information when saved to the MSC server.

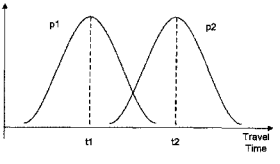


Figure 2. Distributions of travel time by two modes.

Large amount of cellphone data are required for analysis. Records passing through this preprocess can be used to derive travel time information. Let t_1 denote the timestamp when the cellphone leaves the origin cell, and t_2 the timestamp when the cellphone enters the destination cell. The travel time between the origin and the destination is $t = t_2 - t_1$. Assume there are mainly two travel modes between the origin and destination, transit and auto; and assume that the travel time by the same mode is a normally distributed random variable. As shown in Fig.2, the distribution on the left is the travel time by auto, and the distribution on the right is time by transit. The overall travel time distribution is a combination of the distributions by different modes. Given the data obtained from the cellphone records, we can use statistical decision theory to identify the existing modes and classify travelers into different modes.

Trip Assignment

Current trip assignment is based on various user equilibrium models, including static user equilibrium and dynamic user equilibrium. Cellular-based location data is not directly related to the output of the traffic assignment model, as we do not obtain arc flow information from the cellular data. However, as we have shown in the previous sections, cellular-based location data is a good source for OD flow information, which is the main input for traffic assignment? Therefore, a good use of cellular-based location data will contribute greatly to improve the reliability of trip assignment.

Experimental Results

In this section, an experiment has been implemented in a city with large populations. The city has a complicated transportation system; autos, buses, light rails and bicycles all have large groups of users. The experiment for this research is implemented in 2005. Cellular data are collected continuously for one month. Weekday and weekend data are processed separately. The tested area is about 20 square kilometers, controlled by one single MSC server. The area is divided into 12 TAZs and the division is based on the city roadway network, two city expressways and several major arterials taken as TAZ boundaries. The size of most TAZs ranges from 1.5 to 2.5 square kilometers, as shown in Table.1. Most TAZs are commercial areas.

No. of TAZ	Area(km^2)	Approximate Population	Land Use Description
1	1.96	35,650	Commercial/Residential
2	0.84	15,300	College/Recreational
...	...	...	...
12	1.67	21,550	Commercial/Residential

Table 1. Properties of TAZs.

Zone 1, zone 6, zone 10 and zone 11 were selected to test the new method. For trip generation, we collected four weeks of cellular-based location data and analyze the data based on two hours interval, using the method introduced before. The results we get are the trip production and attraction data for zone 1 and 6 between 6:00 AM and 6:00 PM, which are shown in Table 2. Fig. 3 shows the O-D flow obtained from zone 1 to zone 11 during weekdays over one month. The data reliability and stability is acceptable. Next, for the mode split part, we preprocessed the cellular data between zone 1 and zone 10, and obtained the histogram for the travel time between those two zones, as shown in

Fig.3. There appears to be two peaks in the histogram, one corresponding to the travel time around 30 minutes and the other around 60 minutes. This kind of shape is similar to the distribution of a combination of two normal distributed random variables. Therefore we would suppose there are two primary modes between zone 1 and zone 10, and one of them (probably the auto mode) has mean travel time about 30 minutes and the other with mean around 60 minutes might be by transit.

	Trip Generation in Zone 1		Trip Generation in Zone 6	
Time interval	Trip production	Trip attraction	Trip production	Trip attraction
6:00AM to 8:00AM	35197	48133	29400	29902
8:00AM to 10:00AM	48122	42053	28991	34333
10:00AM to 12:00AM	41306	26138	31693	24102
12:00AM to 14:00AM	28571	13609	25424	15023
14:00AM to 16:00AM	16882	5398	17922	13551

Table 2. Trip Generation in Zone 1 and zone 6.

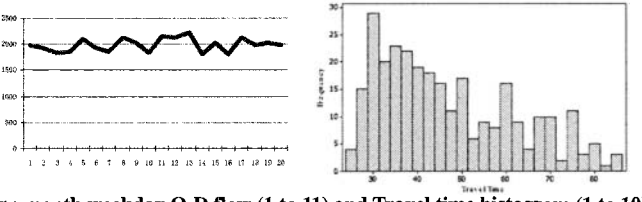


Figure 3. One month weekday O-D flow (1 to 11) and Travel time histogram (1 to 10).

Conclusion

This paper introduces the methodology of integrating the cellular-based location data in the transportation planning process. Cellular information is available directly from mobile carriers, and the collection requires less installation cost and labor as compared to the traditional information collection in transportation planning As we showed, using these data not only improves the model input data quality, but also make possible a great simplification of the current planning process.

References

M.D. Meyer, E.J. Miller. *Urban transportation planning (second edition).* McGraw-Hill Higher Education (2001)

M.B. Goncflalves, J. Cursi. *Parameter estimation in a trip distribution model by random perturbation of a descent method.* Transportation Research Part B 35, 137-161 (2001)

M. Patriksson. *The traffic assignment problem models and methods.* Utrecht, the Netherlands (1994)

R. Sankar, L.Civil. *Traffic monitoring and congestion prediction using handoffs in wireless cellular communications.* IEEE, (1997)

Y. Zhao. *Mobile phone location determination and its impact on intelligent transportation systems.* IEEE Transactions on Intelligent Transportation Systems Vol.1, No.1, (2000)

AVL DATA MINING FOR TRANSIT SERVICE PERFORMANCE

Jingtao Shan[1] and Fang Zhao[2]

[1]Lehman Center for Transportation Research, Department of Civil and Environmental Engineering, Florida International University, Miami, FL 33199; PH (305) 348-3822; Fax (305) 348-2802; email: jshan002@fiu.edu

[2]Lehman Center for Transportation Research, Department of Civil and Environmental Engineering, Florida International University, Miami, FL 33199; PH (305) 348-3821; Fax (305) 348-2802; email: zhaof@fiu.edu

Abstract

AVL system provides a rich and extensive database that can be analyzed to improve transit management and performance. This paper discusses the procedure of using archived AVL data to extract transit performance measurements. GIS representation of AVL data and spatial-temporal data aggregation level are developed. Samples of performance output, based on AVL data collected from Broward County Transit, are presented. The research had shown that extracting performance measures from AVL data is a cost-effective way for transit service analysis.

Key words: AVL, Transit Performance

Introduction

As an ATPS technology, AVL systems make tremendous amount of data available to transit planners and operators. Currently, the basic function of transit AVL system is real time vehicle tracking. The research efforts on using AVL data for bus operation were focused on vehicle arrival time prediction (Lin and Zheng, 1999). Information derived from AVL system also can be used to inform transit service planning and operation. However, the offline AVL data are underutilized because of the difficulties in storage and archival of huge amount of spatial-temporal records. Additionally, the lack of efficient tools and formal procedures to process and interpret the data hindered the usage of the valuable AVL data to the offline analysis (Furth, 2000; Furth *et al.* 2004). With widely deployment of AVL system, transit agencies begin to pay more and more attention to the utilization of these data to the transit service planning, scheduling, and performance improvement.

Because of the complexity of off-line analysis of AVL data, there is a definite requirement for a robust and effective approach to extract useful information. Researchers from Portland State University conducted a series of studies using data recovered by the Tri-Met Automated Bus Dispatch System (BDS) for offline performance analysis (Strathman *et al.* 2000 and 2002). Their studies indicated that the offline analysis of AVL data both enhanced service quality and improved efficiency.

The purpose of this study is to explore how data available from the AVL systems are useful for operation control and planning, and to develop procedures for processing these data in order to obtain valuable information. Before undertaking large-scale data analysis, a sample of data from Broward County Transit (BCT) bus system was acquired and analyzed to demonstrate the methods used to obtain information relevant to bus operations.

AVL System at BCT

The Broward County bus fleet is 100% AVL equipped. The AVL system provides dispatchers with real-time location information of each bus in the system with automatic updates occurring every 2 minutes. BCT AVL system records information at the level of individual block, which is a predetermined set of bus trips (a one-way segment of service) for a given vehicle. Vehicle numbers are used to identify each bus in the sequence of buses on a given route. The database was provided in dBase (.dbf) file format. The data archived contains information including date, time, route, block, vehicle ID, and vehicle location. Figure 1 shows a sample of original dataset archive from the real-time BCT AVL system.

IVLUID	ROUTE	BLOCK	LOCKTIME	NON _REPORT	LOCATION _INVALID	SPEED	XLOC	YLOC	OPERATOR_I	VHCL_DIREC
123	2	8	6/1/2004 4:26:51	0	0	0.00000	929312.85000	626914.95000	70640	WSW
123	2	8	6/1/2004 4:28:47	0	0	0.00000	929444.08000	626527.81000	70640	SW
123	2	8	6/1/2004 4:32:35	0	0	0.00000	926845.68000	629303.37000	70640	WSW
123	2	8	6/1/2004 4:34:35	0	0	0.00000	919027.54000	629395.23000	70640	W
123	2	8	6/1/2004 4:36:35	0	0	0.00000	916278.23000	629362.43000	70640	WSW
9720	2	11	6/1/2004 4:36:52	0	0	0.00000	929365.34000	702324.14000	59860	NNW
123	2	8	6/1/2004 4:38:35	0	0	0.00000	911048.63000	630205.59000	70640	WSW
9720	2	11	6/1/2004 4:38:45	0	0	0.00000	929365.34000	702324.14000	59860	NNW
123	2	8	6/1/2004 4:40:35	0	0	0.00000	904864.32000	630028.43000	70640	WSW
123	2	8	6/1/2004 4:42:35	0	0	0.00000	902174.07000	629949.69000	70640	WSW
9720	2	11	6/1/2004 4:42:40	0	0	0.00000	930073.99000	701454.72000	59860	SSE

Figure 1. Original BCT AVL Data Sample

GIS Integration of AVL Data

As shown in the Figure 1, AVL data were archived as a database rather than a map even though it full of location information. Integration of AVL data with GIS is necessary for advanced transit service analysis. In addition to AVL data, three main sources of data were essential to develop an integrated GIS map and to perform analysis: bus route map, bus schedule, and timepoint location. The schedule database contains information referencing route, direction, trip, and scheduled depart time. AVL data, recorded at variable locations by date, route and block, were linked with schedule data based upon time of occurrence and assigned to the nearest timepoint. The GIS representation process is done in four steps described as following:

1) Apply GIS Geocoding process to show AVL data as points on the map and originally continuous AVL data are distinguished by routes and blocks;
2) Assign moving bus to the correct trip based on the beginning and ending timepoints and the corresponding schedule departure time. Mapping AVL

data to an assigned piece of schedule is not straightforward. Because vehicles may traverse the same route several times, perhaps in different directions, in addition they may be delayed or ahead of schedule;

3) Calculate travel distance. The cumulative travel distance is calculated as the network distance measured from the reference point. The distance between two points was measured as the shortest distance along the route, so-called the network distance, which is distinguished from the Euclidean distance. The reference point was set as the beginning timepoint. As a bus drives away from the reference point, the distance increases. Conversely, as a bus gets close to the reference point, the distance decreases;
4) Construct bus trajectories on a time-distance diagram.

As an illustration, Figure 2 and 3 show the GIS integration process. A sequence of vehicle locations based on a schedule pattern is shown in Figure 2. Each data point in the figure is labeled by a number, indicating the sequence in which the data are received. One can follow the data sequence to trace the bus trajectory. Figure 4 shows a corresponding bus trajectories plot of time versus distance from the reference point to the bus location. Information needed to analyze the bus operation is covered in this graphical representation, including schedule adherence and bus speeds at every moment.

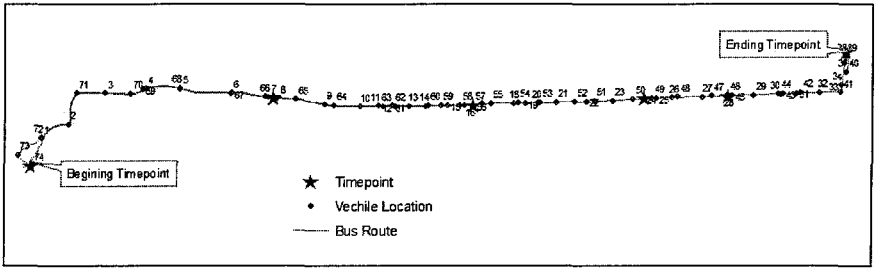


Figure 2. Vehicle Locations Identified by Schedule

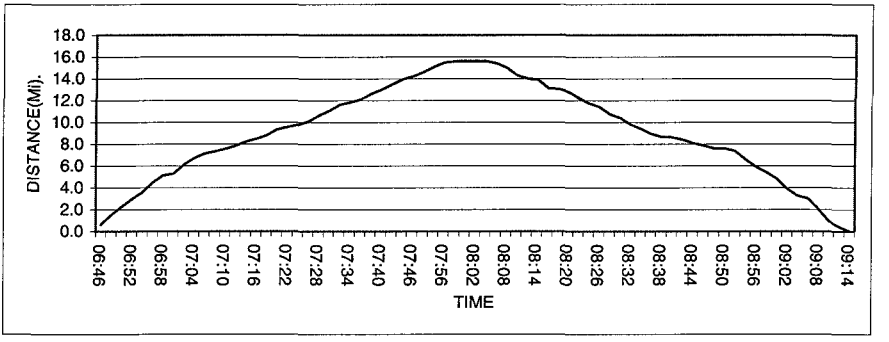


Figure 3. Bus Trajectories on a Time-Distance Diagram

Once the AVL data were integrated into GIS, transit performance information including travel time, travel speed, and schedule adherence can be derived by conducting schedule query program. Based on the archived AVL data, transit performance report can be provided at different spatial and temporal dimensions.

Data Aggregation and Analysis

One essential issue affecting the off-line AVL data analysis is to aggregate the captured raw data to appropriate spatial and temporal detail level. According to the characteristics of the data and the need of transit planners, properly aggregated data will not only save storage space, but also facilitate analysis of the data.

Spatial Level Aggregation

At spatial dimension, transit agencies usually need trip, timepoint, and route level performance analysis and report options. A sample of trip level summary table generated is shown in Table 1. The table covers a selection of morning peak trips for Route 72, an east-west route from downtown Fort-Lauderdale extending to the west of Broward County. While trip level analysis is useful in illustrating the correspondence between service delivery and schedule design, its level of aggregation is too great to be meaningful in schedule design.

Table 1. Trip Level Performance Summary

Date	Route	Block	Dir.	Start Time	End Time	Travel Time	Scheduled Travel Time
60104	72	1	E	6:45	7:52	68	70
60104	72	3	E	7:24	8:50	74	70
60104	72	4	E	7:45	9:03	78	70
60104	72	6	E	8:25	9:40	75	70
60104	72	8	E	8:43	9:52	69	68
60104	72	3	W	6:02	7:12	72	65
60104	72	4	W	6:19	7:28	59	65
60104	72	6	W	7:00	8:12	72	65
60104	72	7	W	7:39	8:53	74	70
60104	72	8	W	7:22	8:26	64	70
60104	72	1	W	8:04	9:16	72	70

Transit performance measurements are most commonly analyzed at timepoint level. Because scheduling practice is based on timepoints, timepoint data is all that is needed for basic running time and schedule adherence analysis. Table 2 shows a sample of output generated at timepoint level for the travel time analysis. Columns with multiple timing points (TP1_TP2, TP2_TP3 etc.) indicate the travel time between the timepoints. Table 3 provides descriptive statistics for travel times between timingpoints. Descriptive statistics is generated for travel times because these are more commonly used by transit planner for scheduling than the individual travel time shown in Table 2. The columns in Table 3 report indicators of operating performance including schedule, average and median travel time, standard deviation of observed travel time and the 85th percentile times. Travel time distributions

between timepoints or at route level may be developed over a designated service period as shown on Figure 4.

Table 2. Timepoint Level Travel Time

Date	Route	Block	Operator	Week Day	TP1_TP2	TP2_TP3	TP3_TP4	TP4_TP5	TP5_TP6
60104	72	1	9726	Tue	10	14	16	4	8
60104	72	1	9726	Tue	14	22	18	6	14
60104	72	1	9726	Tue	12	18	10	8	12
60104	72	1	9726	Tue	12	16	14	6	14
60104	72	2	9005	Tue	17	24	26	6	8
60104	72	2	9005	Tue	18	18	22	8	12
60104	72	2	9005	Tue	18	16	18	6	10
60104	72	2	9005	Tue	14	16	14	10	18
60104	72	2	9005	Tue	14	18	16	8	14

Table 3. Statistics of Timepoint Level Travel Time

Segment	Count	Schedule	Average	Median	Std. Dev.	Min	Max	85th Percentile
TP1_TP2	31	15	14.7	14	3.22	6	20	18.0
TP2_TP3	32	15	17.0	16	5.59	8	33	22.5
TP3_TP4	32	20	14.4	14	4.55	3	26	18.0
TP4_TP5	30	10	7.3	8	2.76	3	16	8.5
TP5_TP6	30	10	11.4	12	3.23	6	18	14.5

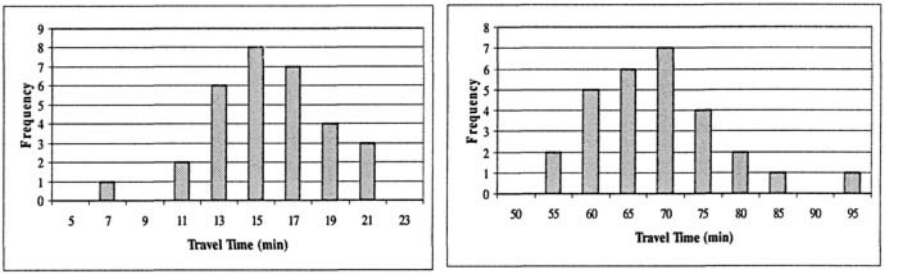


Figure 4. Travel Time Distribution for Route 72 (Left: TP1_TP2, Right: Route Level)

Temporal Level Aggregation

It has been observed that traffic conditions at a particular time of day and day of week are related to like times and days. One of the key advantages of using AVL system to collect a large sample of data is that it provides information on the variation of service across different times and days of week. For example, Figure 5 presents average travel speed of route 72 for each day of week during different time periods. It indicates that the travel speed patterns for weekday and weekend have obvious discrepancy, as well as different hour during a day.

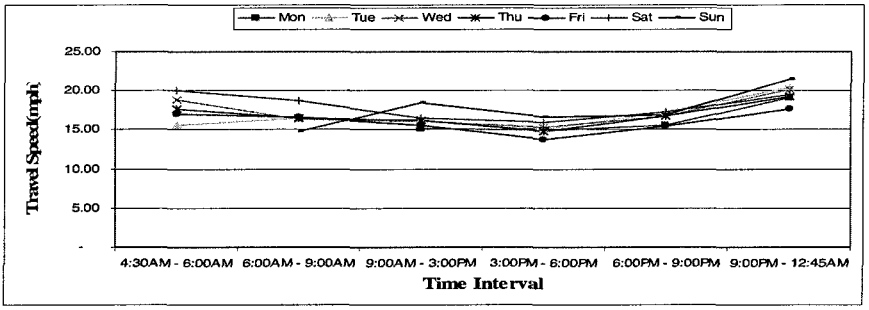


Figure 5. Route 72 Travel Speed for Day of Week

The performance reports generate by AVL data may serve a variety of functions. Service planners can identify service insufficient. Schedulers can evaluate observed running times to determine whether adjustments in the schedule are needed. The operational staffs can assess schedule and headway adherence to identify routes and locations where better control needed.

Conclusion

Archived AVL data are a rich source of transit operation information and, if properly utilized, could shed light on the appropriateness of bus schedules, conditions of operations, service quality, and system performance. This paper described the basic procedure of using archived AVL data to extract transit performance measurement at different spatial and temporal detail level. The research had shown that it is a more efficient and feasible way to obtain performance information with low-cost. It is hope that the research results will bring benefits to both transit providers and transit users.

References

Furth, P. (2000). "Data analysis for bus planning and monitoring." Transit Cooperative Research Program Synthesis 34, Transportation Research Board, Washington, D.C.

Furth, P., Muller, T. J., Strathman, J. G. and Hemily, B. (2004). "Designing automated vehicle location systems for archived data analysis." *Transportation Research Record* 1887, 62-70.

Lin, W. and Zheng, J. (1999). *An experimental study on real time bus arrival time prediction with GPS data. Transportation Research Record* 1666, Transportation Research Board, 101-109.

Strathman, J. G., Dueker, K. J., Kimpel, T., Gerhart, R. L. Turner, K., Taylor, P., Callas, S., and Griffin, D. (2000). "Service reliability impacts of computer-aided dispatching and automatic vehicle location technology: A Tri-Met case study." *Transportation Quarterly*, Vol. 54, 85-102.

Strathman, J. G., Kimpel, T. J., Dueker, K. J., Gerhart, R., and Callas, S. (2002). "Evaluation of transit operations: data applications of Tri-Met's automated bus dispatch system." *Transportation*, Vol. 29, 321-345.

A New Approach for Reliability Assessment of Urban Transportation Networks

A. Shariat Mohaymany[1] and M. Mesbah[2]

[1]Assistant professor, College of Civil Engineering, Iran University of Science and Technology; PH (9821) 73913143; FAX (9821) 77240398; email: Shariat@iust.ac.ir
[2]Postgraduate Student, College of Civil Engineering, Iran University of Science and Technology; PH (9821) 88000527; email: mesbah@iust.ac.ir

Abstract

Transport systems are subject to various types of short term incidents which affect their ordinary functions. Unlike the natural disasters in which the elements of the network would usually fail for long periods, short term incidents temporarily degrade the link capacities and often the network would not reach a new equilibrium before the normal functions return. This paper presents a method for a system level network reliability evaluation with a special attention to the role of network topology. The study is aimed at representing a framework for analyzing vulnerable networks first by introducing a reliability measure and secondly by applying a mathematical analysis method that efficiently executes the algorithm for a large network. The results of the proposed evaluation procedure for transportation system are promising. The study demonstrates that the reliability measure and network topology concepts can be included in the transportation network modeling process at a reasonable effort.

Introduction

Many stochastic events affect the function of a transportation network which in turn, can significantly influence the economy of the region. Therefore, in recent years, measuring the ability of the network to reasonably function after such stochastic events (i.e. reliability) has become an important research topic. A complete review of these studies are given in (Berdica, 2002) and (Watling, 2004).

Events are classified into stochastic and recurrent causes (Watling, 2004). Stochastic causes such as earthquakes and hurricanes are events with low probabilities and high consequences. Right after the disaster takes place, during the confusion state (defined by Iida, 2000), in which the main concern is rescue, the system's reliability on connectivity (Bell, 1997) should be used as the performance measure. In this period,

system equilibrium is not reached. In subsequent periods, consequences usually linger for such a long time that the network is expected to reach a new equilibrium. In such cases, travel time (Asakura, 1996), capacity (Chen, 2002), and consumer surplus (Poorzahedy, 2005) are used as performance measures. It should be noted that although different measures of performance are introduced, an extensive use of traditional user equilibrium (UE) or stochastic user equilibrium concepts (Du, 1997) is observed in such analysis. Clearly, this level of analysis takes into account the network topology in evaluating the performance measure.

The second includes those who cause low consequences which takes place daily in the network like traffic accidents, day to day travel time variations, and traffic signal failures. The nature of such events does not easily satisfy Wardrop principles; however, modeling user perception and response to risk was tried; a comprehensive overview of which is available in (Noland, 2002). Similar to the confusion state in stochastic events, each traffic incident has a transition state in which no equilibrium is reached. The reliability of the network in this period is the focus of this paper. There are two approaches for analyzing the network (e.g. the calculation of total travel time) in the transition state: i) dynamic assignment and ii) micro simulation. Generally, dynamic assignment (DA) is an option for a dynamic case. Such an approach, however, is to ensure equilibrium in the network, as expressed above, and is not a justified assumption. Moreover, implementation of DA requires much larger amount of computational overhead than a UE. The second approach is to apply microscopic traffic simulation software and model details of the system. Today, traffic simulation software programs enable the analyst to simulate disturbances like traffic accidents in detail. The major disadvantage of this approach is the need for collecting large amount of information and calculations to perform a single run. Hence, simulating a large real network is often infeasible.

In this study the ordinary UE approach was employed, but a series of assumptions were made to circumvent the problem of studying a transition state by a static approach. A concept called "partial assignment" was also used to present flow diversions during transition states.

Methodology

The assumptions made for estimating the reliability against traffic incidents are:
1. There would be one accident in the network at each analysis (named "state"), but the effect of the accident in one link on upstream links is of importance.
2. Without loss of generality, it is assumed that an accident always occurs at the downstream end of the link and would completely blocks it for a known period called the incident duration (ID). No accident occurs in nodes.
3. Drivers become aware of the accident no earlier than when they encounter the traffic jam (queue).
4. Accidents are so undesirable that the drivers try to avoid them as much as possible.
5. Stop delay and traffic diversion delay are considered as the two major sources of delay and the congestion effect is considered in route choice.

Assume that a detailed network N(V,A) is given where V is the set of nodes and A is the set of links, with M=|A|. Let T be the set of origins and S be the set of destinations, S and T $\subseteq$ V. Let t_{ij} be the travel time on link (i,j). Let also, x^s_{ij} be the flow over link (i, j) destined to s, and x_{ij} denote the flow over link (i, j),

$$x_{ij} = \sum_{s \in S} x_{ij}^s \quad \forall (i,j) \in A, s \in S$$

Finally, let $D = \{d_t^s \mid s \in S, t \in T\}$ be the origin-destination (O/D) trip matrix.

The procedure for reliability assessment is outlined in the following 4 main stages:
Stage 0: Carry out an assignment for finding the flow pattern. Find also, x^s_{ij} for all links. set m, the link counter, to 1.
Stage 1: Generate an incident in (i, j) with the respective ID_{ij}.
Stage 2: Propagate incident induced queue in the network and calculate the delay.
Stage 3: Compute the total travel time (TTT) for the range of disturbance.
Stage 4: If m<M then increment m by 1 and go to stage 1; otherwise, calculate reliability and stop.

Note that procedure continues until an incident in every link of the network is generated and computation of the respective TTT is done. Reliability is the ratio of number of states in which TTT_{ij} remains suitable in comparison to a reference TTT, to total number of the states (M), i.e. $R(\theta)=Prob\{(TTT_{ij} /TTT^0)< \theta \}$, where TTT^0 is the network travel time in base state, and θ an acceptance threshold.

It should be noted that according to assumptions 3 and 4, drivers enter a blocked link and when they encounter the queue, there would be no way but joining it. As a result that link would be filled by cars. At the immediate upstream intersection drivers may have an alternative route which they would choose (Assumption 4). Therefore, for each incident generated in stage 1, at least one link would be filled. Stage 2 of the above procedure needs to use the concept of partial assignment which is explained in part 3 of this paper. The detailed formulation for stage 3 is discussed in part 4.

Flow Diversion Algorithm

Queue propagation and dissipation in the network is a major challenge for reliability assessment explained in previous section. This phenomenon is directly related to the existing network topology. The more the topology offers alternative routes, the higher the network reliability. Even if such routes are not chosen in the base state of the user equilibrium. If the aforementioned queue propagation beyond one link happens after an incident, drivers at the upstream node would seek their chance to divert to a new route. According to the fact that commuters are fully familiar with the network, they would analyze and find the alternative routes at the moment of decision making. Finding the new route from the upstream node based on user response is called partial assignment (PA) in this study. If all users that used to choose the blocked link can find an alternative, the queue would not spread. If not, queue would propagate through the network.

For formulation purposes, it is assumed that the incident initially is generated in link (i,j) (stage 1 in part 2) and it fills the link. Set t_{ij} to infinity and set link (i,j) as inactive. Nonzero volumes, including future x^s_{ij}, that already are flowing through upstream links toward the blockage (point i) are to be diverted. The starting node, i, is assumed a sink and source node for the flows on link (i,j) (x^s_{ij}). This means they continue their paths assigned prior to incident to node i since it was the shortest path and the drivers are not aware of the accident according to the aforementioned assumption 3. At point i, a partial assignment (PA) is carried out by $d^s_i = x^s_{ij}$ as a single line O/D trip matrix to all $s \in S$. The base state equilibrium travel times are considered as initial travel time in the links' cost function. Flows resulting from this step (y_{ij}) should be stored separately (see the first PA in Figure 1). Whether $y^s_{ij}=0$ for all $s \in S$ where (i,j) is the blocked link, the users have found an alternative path; otherwise, there is an OD pair for which the network prepares with a single path from node i, whose path is now blocked (see the first condition in Figure 1). In the former case (negative answer in Figure 1) no new link would be blocked. On the other hand for the latter case, the links pointing to i (i.e. links (p, i) for all p) are examined for link activation.

Let S_{ij} be the set of destinations in which x^s_{ij} is nonzero, $S_{ij} = \{s \in S \mid x^s_{ij} \neq 0\}$. Link (p,i) is blocked if S_{pi} (provided in stage 0) and S_{ij} (Calculated in the previous step) have any member in common. A common member (say $s=s_1$) indicates that for the flows in link (p,i) destined to s_1, the shortest path includes the blocked link. This portion of flow on link (p, i) would activate the link for blockage. Furthermore, users destined to s_1 do not allow the rest of the flow on link (p, i) to reach point i. That is why flow y_{pi} should be excluded from partial assignment OD matrix (subtract demand in Figure 1). Let B be the set of activated links arranged in order of earlier time that queue fills them and B' be the set of activated and then inactivated links.

Next step is to run another PA with the updated demand. This time, resulting flows (y_{pq}) are added to the base state flows (x_{pq}):

$$x_{pq} = x_{pq} + y_{pq} \quad (p,q) \in [A - B'] \tag{1}$$

In the last step, the first member of B (say link (p,q)) is selected for another cycle of calculations, and this continue until B is empty.

It should be noted that the ID for the upstream link is always less than the downstream one; whereas, a link is filled at a speed proportional to its flow, while the queue is dissipated, after the accident, proportional to the saturated flow. So, this cycle would continue until ID_{pq} gets less than the time required filling the link (p,q) or as mentioned above, until there would be an alternative path for all flows.

Besides, interestingly, queue dissipation could be considered parallel to activation procedure by only conveying the factor ID to the following upstream links. Having ID computed for each link, dissipation calculations could be performed for that link.

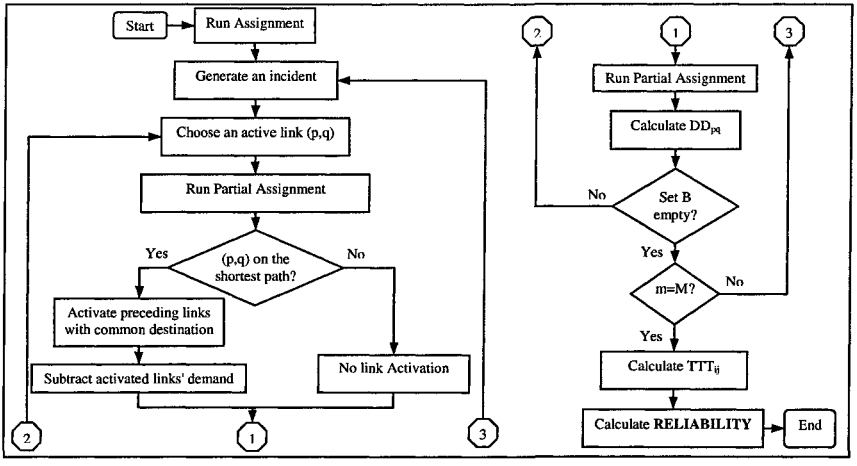


Figure 1. Reliability Assessment Flowchart

Network Delay

After each cycle of algorithm in part 3, traffic is diverted and the flow pattern is valid for short period from the moment of running the partial assignment (PA) in one link until the next PA in another one (or the time of the last queue dissipation for the case of the last activated link). Let this valid period be VP_{pq} for PA in link (p,q) and incident initially generated in link (i,j).

Having the adjusted flow (x) from equ.1 and the respective travel time (t) from cost function for all links in each VP_{pq}, flow diversion delay in that cycle (DD_{pq}) is calculated. This delay is the difference of adjusted total travel time (per hour) and base state travel time (rate again) multiplied by validation period in one hour.

$$DD_{pq} = \left(\sum_{(k,l)\varepsilon A-B'} x_{kl} t_{kl} - \sum_{(k,l)\varepsilon A-B'} x_{kl}^{0} t_{kl}^{0} \right) \times VP_{pq} / 1hr \qquad (2)$$

where x_{kl}^{0} and t_{kl}^{0} is the base state flow and travel time, respectively.

When B gets empty in the last cycle, diversion delays in all cycles (equ.2) are summed to derive the total diversion delay resulted from incident in link (i,j), (DD_{ij}):

$$DD_{ij} = \sum_{(p,q)\varepsilon B'} DD_{pq} \qquad (3)$$

For the inactivated set of links (a member of set B') stop delay (SD) is equal to number of cars in the queue multiplied to their average experienced delay. An approximation of these factors could be derived from link properties like flow, width, and ID which are easily calibrated.

Summation of delays (flow diversion delay and stop delay) with the base state TTT^0 results in the network total travel time (TTT_{ij}) for the incident in link (i, j). This TTT_{ij} is taken as the performance measure as described in part 2 of this paper:

$$TTT_{ij} = (DD_{ij} + SD_{ij}) + \sum_{(k,l)\in A} x_{kl}^0 t_{kl}^0$$

Conclusion and Further Work

Studying reliability of a transportation network in transition state after short term incident is highlighted in this paper. The procedure is based on a concept called "partial assignment" which is applied to analyze queue propagation in the network by the traditional UE approach. The proposed framework is mathematically efficient because network states are a linear function of the number of links. Also, this method considers queue propagation and dissipation in the network at the same time. Dissipation effects are inserted in the incident duration (ID) for each link. All above compressions enable the procedure to consider all states and compute total travel time during disturbance (the performance measure).

The presented procedure is coded and implemented for small test networks. However, using different functions for the rate of filling and emptying of a link with and from a queue is under development. The performance of this framework on a large network should be examined although theoretically promising results are predicted.

References

Asakura, Y. (1996). "Reliability measures of an origin and destination pair in a deteriorated network with variable flow." in Bell, M.G.H. (Eds) Transportation Network: Recent Methodological Advances, Pergamon Press, Oxford, 273-288.

Bell, M. G. H., and Iida, Y. (1997). Transportation Network Analysis, Wiley, UK.

Berdica, K. (2002). "An introduction to road vulnerability: what has been done, is done and should be done." Transport Policy 9(2), 117-127.

Chen A., Yang H., Lo, H. K., Tang, W. H. (2002). "Capacity reliability of a road network: an assessment methodology and numerical results", Transportation Research 36B(3), 225-252.

Du, Z. P., and Nicholson, A. (1997). "Degradable transportation systems: sensitivity and reliability analysis." Transportation Research 31B(3), 225-237.

Iida, Y., Kurauchi, F., Shimada, H. (2000). "Traffic Management System Against Major Earthquakes." IATSS Research, 24(2).

Noland, R., and Polak, J. (2000). "Travel Time Variability: A Review of Theoretical and Empirical Issues." Transport Reviews 22(1), 39-54.

Poorzahedy, H., and Shetab, S. N. (2005). "Network performance improvement under stochastic events with long term effects." Transportation 32, 65-58.

Watling, D., Sumali, A., Connors, R., Balijepalli, Ch. (2004). Advancing Methods for Evaluating Network Reliability, Final Report, University of Leeds, UK.

Model and Algorithm for Iterative Design of Bus Network

Bin YU, Zhongzhen YANG
College of Transportation and Logistics, Dalian Maritime University, Dalian, China
E-Mail: minlfish@yahoo.com.cn

Abstract

An iterative approach for bus network design is developed, which optimizes bus network and transit assignment simultaneously. The approach takes the changed stops and OD traffics into account during designing bus network. An improved ant colony optimization (IACO) and label-marking method are designed to get a Pareto optimal solution. The model and the algorithms are tested with the bus network in the city of Dalian and some conclusions are drawn.

1. INTRODUCTION

Existing literatures for bus network optimization mainly focus on two types of models: one combines the network design and service frequency (Dubois *et al*, 1979; Baaj and Mahmassani, 1995); and the other solely concentrates on network optimization (Wang *et al*, 2001; Sonntag, 1979). Most of them are based on the given stops or zones and OD traffics, while the being changed OD traffics due to the new network are hardly considered. This paper presents an iterative approach, which optimizes bus network and transit assignment simultaneously, and takes the changed stops and OD traffics into account during designing bus network. To realize the goal, an iterative approach, which deals with the interactions between the lines and the stops, is developed. Due to the complex structure of the model and in order to get a Pareto optimal solution, an improved ant colony optimization and label-marking method are used for bus network design and transit assignment, respectively.

2. OPTIMIZATION MODEL

In this paper, first, a large initial stop collection, which consists of current stops and potential stops, and a vacant network are established. Then a new bus network is designed based on the initial stop collection and OD traffics. During the process, some infeasible stops in the stop collection are abandoned and a new stop collection and a bus network are obtained. Thus, the traffic demand will change to generate a new OD matrix. To estimate the new OD traffic, transit assignment model based on optimal strategies (Spiess and Florian, 1989) are used. Then, based on the new stop collection and the new OD traffic, the bus network will be redesigned. As Fig.1 depicted this process does not stop until some criteria are satisfied. Two components of this approach are discussed here and then in next section IACO and label-marking method are designed to solve them, respectively.

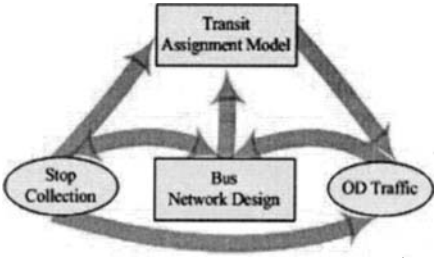


Fig.1 The process of iterative approach

2.1 Bus Network Design Model (BNDM)

Our bus network design aims to maximize direct traveler density (YU *et al*, 2005), i.e. to maximize the number of direct travelers per unit length, which considers the direct travelers and the bus line length simultaneously. This method differs from previous researches (Wang *et al*, 2001; Simonis, 1981; Michael *et al*, 1997) in which most of them firstly identified the shortest path between the origin station and the terminal, and then sought the route bearing the most direct travelers among these shortest routes. We have freed from the shortest path constrain, but seek the route with most direct travelers per unit length in all possible paths between the origin station and the terminal.

This can be explained by two reasons. First, the direct travelers are not always the largest on the shortest paths, which means it is not reasonable to lay all routes along the shortest path between the origin station and the terminal. On the other hand, when the objective is set to be maximizing direct travelers, due to the increase of the direct travelers in accordance with the increase of the length of the transit line, a certain route may be abandoned because it is comparatively shorter, even if this short route abounds in direct travelers. This can result in a deviation of the routes from the demands of the travelers. In addition, the overall length of the entire network increases where longer line is laid, which consequently increases the operating costs. Therefore, this study enhances the network utilization rate as high as possible. The routes with high demand are laid first to facilitate passenger trips and to benefit the transit enterprises.

The bus line optimization problem is equivalent to find the corresponding stops and sections. Here, N denotes the stop collection and A denotes the collection of road sections. All of the alternative bus lines, the sub-collection of A, are denoted as S_{OD}. A binary decision variable x_{ij} represents whether line S_{OD} is included in the line plan. The incident matrix, Δ_{ij}, indicates whether line S_{OD} passes through the adjacent stops i and j. Then the bus network optimization model is as follows:

$$\max: D_{od} = \frac{\sum_{i\in N}\sum_{j\in N} SP_{ij}x_{ij}}{\sum_{g\in N}\sum_{h\in N}\Delta_{gh}l_{gh}x_{gh}} \quad s.t. \begin{cases} L_{min} \le L_{od} \le L_{max} \\ Q_{od}^{sum} > Q_{min} \\ Q_{od}^{kl} < Q_{max}^{kl} \\ \forall l_{gh} > 0.5km \\ od \in F \\ m \ne n \quad \forall m,n \in S_{od} \\ NTR > 50\% \end{cases} \tag{1}$$

where D_{od} = direct traveler density;

F =collection of the origin and destination stops;

o, d= origin and destination stops;

SP_{ij}= the number of direct travelers between stop i and j;

$L_{od} = \sum\sum \Delta_{gh} l_{gh} x_{gh}$ = length of bus line;

L_{min} / L_{max} = minimum/maximum length of the bust line;

l_{gh} = length of the road section between stop g and h;

Q_{od}^{kl} = direct travelers of section k, l;

Q_{max}^{kl} = maximum direct travelers of section k, l;

$Q_{od}^{sum} = \sum\sum SP_{ij}x_{ij}$ = total direct travelers of the line;

Q_{min} = the threshold of direct travelers of the bus line;

NTR = non-transfer ratio of the entire network.

It is a non-linear integer programming problem, with the constraints as follows: *Length constraint* is 30-40min per single-trip time; *Minimum line-setting direct travelers constraint* means the least direct travelers for a line; *Section direct travelers constraint* means the section direct travelers must be less than the capacity; *Stop spacing constraint* is the average stop space l_{ij}=0.5~0.6km; *Trend constraint* means that a bus line should not contain a loop; *non-transfer ratio constraint* is used for the minimum direct-through ratio. It can be calculated by dividing the total direct travelers by the total O-D passenger. If there is a comparatively big discrepancy between the calculated NTR and the adopted one, the network should be redesigned with the calculated one, until the adopted NTR approximates to the calculated one.

2.2 Transit Assignment Model

The transit network is assumed to be represented by a standard node/link type network, where a set of nodes $i \in I$ is connected by a set of links $a = (i, j) \in A$. The set of links going out of node i (forward star) is denoted A_i^+, and the set of incoming links (backward star) is denoted A_i^-. Travel time (or cost) c_a, line section flow v_a and service frequency f_a are associated with each link a. The demand between the nodes i and j is given by g_{ij}.

The assigning of the trips from all origins to destination r according to the optimal strategy corresponds to solving the following linear optimization problem:

$$Min \sum_{a \in A} c_a v_a + \sum_{i \in I} \omega_i \tag{2}$$

subject to

$$\sum_{A_i^+} v_a - \sum_{A_i^-} v_a = g_{ir}, \quad i \in I,$$

$$v_a \le f_a \omega_i, \quad a \in A_i^+, i \in I,$$

$$v_a \ge 0, \quad a \in A.$$

Note that the variables ω_i represent the total waiting time (in person minutes) at node i.

The waiting time at a node depends on the set of attractive links $\bar{A}_i^+ \subseteq A_i^+$, i.e. the set of outgoing links which are considered for travel by the travelers by boarding the first vehicle leaving on any of these links. For any given set of attractive links $\bar{A}_i^+$ at node i, the combined waiting time is proportional to the combined total frequency of these links

$$W(\bar{A}_i^+) = a / \sum_{a \in \bar{A}_i^+} f_a, \quad a > 0 \tag{3}$$

and the probability of leaving node i on link a is

$$P_a(\bar{A}_i^+) = f_a / \sum_{a' \in \bar{A}_i^+} f_{a'}, \quad a \in \bar{A}_i^+ \tag{4}$$

Given the above relations, any strategy for reaching destination r is completely defined by the corresponding subset of attractive links $\bar{A} \subseteq A$.

The optimal strategy for reaching a destination is the one that minimizes the total expected cost. Note that the cost of a strategy is the sum of link travel times c_a weighted by the probability of traveling on links a, and the waiting time at nodes i weighted by the

probability of traveling through node i. It has been shown that for fixed link travel time c_a,

3. SOLUTION ALGORITHM

Due to the complex structure of the approach, IACO and label-marking method are designed to solve them, respectively.

3.1 IACO for BNDM

If we take the origin as the *nest* and the terminal as the *food*, the optimal problem can be simplified to a process of the ant colony searching for *food*.

3.1.1 Generation of Solutions

Ant colony optimization is a meta-heuristic technique that uses artificial ants to find solutions to combinatorial optimization problems. Using ACO, one ant's route is constructed by incrementally selecting stops until terminal. The decision making about selecting stops is based on a probabilistic rule taking into account both the visibility and the pheromone information. Thus, to select the next customer j, the ant uses the following probabilistic formula.

$$p_{ij}^{\ k} = \frac{\tau_{ij}^{\alpha} \times \eta_{ij}^{\beta}}{\sum_{h \notin tabu_k} \tau_{ih}^{\alpha} \times \eta_{ih}^{\beta}} \tag{5}$$

where $p_{ij}^{\ k}$ -the probability of choosing the j_{th} stop for the k_{th} ant.

τ_{ij} and η_{ij} - the pheromone concentration and the visibility on the arc(i, j) , respectively.

α and β- the relative influence of the pheromone trails and the visibility values, respectively.

$tabu_k$ - the set of the infeasible nodes for the k_{th} ant.

3.1.2 Local Search

In our algorithms we apply the 2-opt algorithm (Croes, 1958) within routes to improve the routing. More precisely, in our case, at each step of the algorithm all edges are investigated according to a randomly chosen order and the first exchange encountered that leads to a non-dominated new solution is applied. If all pairs of edges were considered and no solution is accepted, a Pareto local optimum solution is reached.

3.1.3 Update of Pheromone Information

First, Pheromone updating is conducted by reducing the amount of pheromone on all visited arcs in order to simulate the natural evaporation of pheromone and to ensure that no one path becomes too dominant. This is done with the following pheromone updating equation,

$$\tau_{ij} = \rho \times \tau_{ij} + \sum_{k} \Delta \tau_{ij}^{k} \qquad \rho \in (0,1) \tag{6}$$

where $\Delta \tau_{ij}^{k}$ = the pheromone increments of the k_{th} ant on edge (i, j).

ρ = the parameter that controls the speed of evaporation.

k= the No. of the ant.

In this paper, Ant-Weight (YU *et al,* 2005), is adopted. Specifically, the pheromone increment update strategy is written as:

$$\Delta \tau_{ij}^{k} = \frac{Q}{f^{k}} \times \frac{f^{k} - f_{ij}}{(n^{k} - 2) f^{k}} \tag{7}$$

where Q remains a constant.

f^k = the sum of object value of the k_{th} route.

f_{ij} = the object value of edge (i, j).

n^k = the number of stops of the k_{th} route and $n>2$.

This update strategy consists of two components: the first one is global pheromone increment Q/f^k that is related to the quality of the k_{th} optimization route; the second one is local pheromone increment $(f^k - f_{ij})/(n^k - 2)f^k$ that is according to the contribution of edge (i, j) to the optimization route.

3.2 Label-Marking Method for Transit Assignment Model

The transit assignment model can be solved efficiently with the following label-marking method (Spiess and Florian, 1989):

Step0: (Initialization)

$$u_i \leftarrow \infty, i \in I - \{r\};\ u_r \leftarrow 0;\ f_i \leftarrow 0, V_i \leftarrow 0, i \in I;\ S \leftarrow A;\ \bar{A} \leftarrow \phi$$

Step 1: (Get next link) if $S = \phi$ then got to step 3,

Otherwise find $a = (i, j) \in S$ with minimum $u_j + c_a$;

$S \leftarrow S - \{a\}$.

Step 2: (Update *i*-node) If $u_i \geq u_j + c_a$ then

$$u_i \leftarrow \frac{f_i u_i + f_a(u_j + c_a)}{f_i + f_a}, \quad f_i \leftarrow f_i + f_a, \bar{A} \leftarrow \bar{A} + a;$$

go to step 1.

Step 3: (Loading) Do for every link $a \in A$, in decreasing order of $(u_j + c_a)$:

$$\text{If } a \in \bar{A} \text{ then } v_a \leftarrow \frac{f_a}{f_i} V_i,\ V_j \leftarrow V_j + v_a, \text{ otherwise } v_a \leftarrow 0.$$

In the label-marking method, V_i is the total flow entering node *i*. u_i is the expected travel time from *i* to *r*. *S* represents the examined links.

4. NUMERICAL TEST

The data in Dalian city is used for numerical test. The road network consists of 3,200 links and 2,300 nodes. At present there are 89 bus lines, which extend 1,130km, and 1,500 bus stops. We get bus OD traffic matrix through on-board survey of all of the 89 line. The improved ACO is coded in Visual C++.Net 2003 and executed on a PC equipped with 512 MB of RAM and a Pentium processor running at 3000 MHz.

Totally 81 bus lines are designed with our method, which extend 925km and 1,220 bus stops. Although the number of optimized bus lines and stops are less than the existing one, the service area of the transit system does not decrease. By comparing the optimized bus network with the existing one, it is known that the direct traveler density of current bus network is 28.8persons/km much lower than the optimized 35.2persons/km (Fig 2). This is mainly because that current bus lines overlap each other to an extend that can disperse the trip flow, thus lowers the efficiency of the whole network. Furthermore, to estimate the optimized bus network, we compare the NTRs of the existing network and the optimized network. In optimized case, direct trips shares 57%, while the existing one is about 41%. Although the number and the total length of the optimized bus lines are less than the existing one, the NTR of the optimized network is improved. This indicates that the bus network deployment of Dalian is unreasonable and also the opposed model is effective.

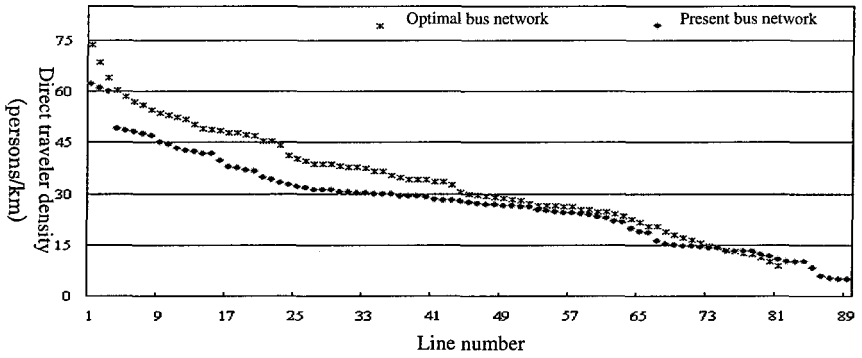


Fig.2 Comparisons of direct traveler densities

5. CONCLUSIONS

A new dynamic iterative approach for urban bus network is developed, which aims to optimize the bus lines and stops simultaneously and takes the effects of the changed demand on stops and OD traffics into account. This approach consists of two components: the first one is bus network design and the second one is transit assignment. The two components are iterated till the convergence. Since bus network design is an NP-hard problem, an improved ACO is adopted to get a Pareto optimal solution in this paper. Then, with data of Dalian city we test our model and algorithm, and compare some indices between the optimized and existing bus networks. Results show that the model and the algorithm are effective.

Reference

Baaj, M. H. & Mahmassani, H.S. (1995): Hybrid Route Generation Heuristic Algorithm for the Design of Transit Networks. *Transp. Res., Part C* 3(1),31-50.

Bin Yu, Zhongzhen Yang, Chuntian Cheng, Chong Liu, (2005): Optimizing Bus Transit Network With Parallel Ant Colony Algorithm, *Proceedings of the Eastern Asia Society for Transportation Studies*, Vol.5, pp.374-389.

Croes, G. A.(1958): "A method for solving Traveling Salesman Problems". *Operations Research* **6** , pp. 791–801.

Michael R. Bussieck, Peter Kreuzer, Uwe T. Zimmermann, (1997): Optimal Lines for Railway Systems. *European Journal of Operation Research*, 96(1) 54-63.

Dubois, D., Bel, G. & Libre, M. (1979), A Set of Methods in Transportation Network Synthesis and Analysis, *Journal of Operational Research Society*, 30(9), 797–808.

Sonntag, H., (1979): Ein heuristisches verfahren zum entwurf nachfrageorientierter inienführung imöffentlichen personennahverkehr. *Z. Oper. Res. Ser. A-B*, 23: B15–B31.

Spiess H., Florian M. (1989): Optimal Strategies: A New Assignment Model for Transit Networks. *Trans. Res. B* 23(2), 83–102.

Wang Wei, Yang Xinmiao, *et al* (2001): Planning and Management Technology for Urban Transit System, *Science Press*, Beijing.

A Genetic Algorithm for Bus Schedule Synchronization

Fabian Cevallos[1] and Fang Zhao[2]

[1]Center for Urban Transportation Research, University of South Florida, 4202 E. Fowler Ave., CUT 100, Tampa, FL 33620-5375; PH: (954) 234-4183; Fax: (813) 974-5168; email: cevallos@cutr.usf.edu

[2]Lehman Center for Transportation Research, Department of Civil and Environmental Engineering, Florida International University, Miami, FL 33199; PH: (305) 348-3821; Fax: (305) 348-2802; email: zhaof@fiu.edu

Abstract

The bus transfers synchronization problem is to find an optimal bus schedule for the entire transit system that minimizes riders' transfer time. This is a complicated problem because of the large set of binary and discrete values involved. The combinatorial nature of the problem poses a computational burden and makes it difficult to solve using classical methods. This paper presents a Genetic Algorithm (GA) approach to synchronizing bus schedules to minimize transfer times. The algorithm takes existing scheduled timetables and ridership data at transfer locations as input and addresses randomness in bus arrivals. The algorithm attempts to find an optimum solution for the bus schedule synchronization problem by shifting existing timetables. Using scheduling data from Broward County Transit, total passenger transfer times were calculated for the existing and the proposed systems. The algorithm was able to reduce system-wide transfer time by 13.4%. Results from this case study and recommendations for future research are presented.

Introduction

This paper focuses on transfer time optimization under the no-holding policy, which involves shifting the times of existing timetables for all or selected routes such that the overall time for all passengers that transfer from one vehicle to another is minimized. Optimization of transfer times is a subject that has been studied by many in the past. However, the classical optimization models such as Mixed-Integer Programming (MIP) and Mathematical Programming are not suitable for this type of problems due to binary and discrete variables associated with large sets of combinatorial values. In recent times, there have been some interests in solving this type of scheduling problems using evolutionary methods. For instance, Deb and Chakroborty (1998) presented a genetic algorithm based theoretical framework for the formulation of transfer optimization. However, further research is still needed to develop algorithms to solve realistic transfer optimization problems.

Problem Statement

Providing a coordinated service in which transfer times for the whole system are minimized is not easily achieved. Kikuchi and Parameswaran (1993) state that transit schedule optimization is an extremely difficult task even for a small transit network. This is mainly due to the complexity and combinatorial nature of scheduling problems.

To illustrate the complexity and computational intensity of the problem, assume a relatively small transit system with 40 directional routes, 10 trips per route, and 30-minute headways. The timetables will be shifted to optimize transfer times. For each route, there will be 30 possible time shifts per each trip if one-minute shifts are allowed. For 40 routes, the number of combinations of all possible timetable shifts will be (30 x 10)^40 $\approx$ 1.216E+99. This is a number of astronomical magnitude and it is impossible to find a global optimal solution based on exhaustive search or mathematical programming techniques in a reasonable amount of time. The computational complexity of the problem increases when other variables such as all the possible transfer locations, ridership demand, random arrivals, etc., are to be considered.

From a practical point of view, transit agencies do not usually have efficient ways to create timetables that provide an optimal coordinated service. In most cases, coordination is only provided at selected transfer locations, usually major terminals. This is mainly due to a lack of tools and practical methodologies to tackle these complex optimization models.

Research Objectives

This research attempts to develop an optimization model based on a genetic algorithm to solve the transfer time optimization problem of a fixed-route system. Optimization of transfer times is examined from the scheduling point of view by considering all trip combinations for all the timetables without modifying existing headways. The goal is to coordinate schedules to reduce the time spent during transfers by searching for the optimal combination of bus departures and arrivals.

This methodology is to be used as part of the timetable development during which transfer times are minimized without a major impact on the amount of service provided. While approaches that involve headway modifications (e.g., Deb and Chakroborty, 1998) have practical applications, they may not be suitable without being sensitive to other design, operational, and local variables. Transit agencies in the U.S. cannot easily modify headways without impacting contract labor rules, analyzing resource reallocation, and addressing political jurisdiction equity issues. Because of these issues, there is the risk of headway modifications not being accepted by the public or politicians, defeating the original purpose of improving transit services.

Optimization Model

The model formulation is as follows, which minimizes the sum of all transfer times occurring between a pair of transit routes i and j at transfer points k and l and from bus p to bus q:

Minimize

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{D_i}R_k\sum_{l=1}^{D_j}C_{ijkl}\sum_{p=1}^{B_i}\sum_{q=1}^{B_j}\{[d_{jlq}+(\Delta_{jlq}+v\sigma_{jlq})+S_j]-[a_{ikp}+(\Delta_{ikp}+v\sigma_{ikp})+S_i+w_{ijkl}]\}$$

Subject to:

$$-b_i \le S_i \le b_i,\quad b_i \le \tfrac{1}{2}\, h_{\min}^{(i)},\ i = 1, 2, \ldots, N$$

$$-b_j \le S_j \le b_j,\quad b_j \le \tfrac{1}{2}\, h_{\min}^{(j)},\ j = 1, 2, \ldots, N$$

$$d_{jlq} >= (a_{ikp} + w_{ijkl})$$

$$C_{ijkl} = \begin{cases} 0, & \text{if there is no connectivity at transfer points from route } i \text{ to route } j \\ 1, & \text{if there is connectivity between routes } i \text{ and } j \text{ at transfer points } k \text{ and } l \end{cases}$$

where

a_{ikp}	=	arrival time of bus p for Route i at transfer point(s) $k \forall r_n$
d_{jlq}	=	departure time of bus q for Route j at transfer point(s) $l \ \forall r_n$
$\Delta_{ikp}, \Delta_{jlq}$	=	deviation from schedule due to random bus arrivals
v	=	schedule variability coefficient
$\sigma_{jlq}, \sigma_{ikp}$	=	standard deviation of bus arrivals and departures
r_n	=	set of timetables for all N routes
R_k	=	ridership demand at transfer point k
S_i, S_j	=	time shifting for Routes i, j, respectively
b_i, b_j	=	time interval for time shifts
$h_{\min}^{(i)}$	=	minimum headway on Route i
$h_{\min}^{(j)}$	=	minimum headway on Route j
w_{ijkl}	=	walk time between bus stops at transfer points k and l from Route i to Route j
C_{ijkl}	=	connectivity Matrix from Route i to Route j and from transfer point k to l
N	=	number of bus routes in the system
D_i	=	number of transfer points on route i
D_j	=	number of transfer points on route j

B_i = number of buses in service on route i
B_j = number of buses in service on route j

In the formulation, S_i and S_j are the decision variables, which assume a minimum value of one minute. A solution is represented by a vector **S**, which stores all the S_i, that is $\mathbf{S} = \{S_i\}$.

Implementation

The data flow process starts with the collection of the timetables from scheduling software from which scheduled arrival and departure times can be obtained. Real-time arrivals and departures may be obtained from AVL or APC systems. Transfer ridership data may be obtained from an electronic fare collection system. The model formulation is implemented using a GA-based algorithm. A solution is coded in GA using integers that represent the time shifts to be applied to all routes. In other words, the string that represents a solution contains N integers, where N is the number of routes. The crossover operator is the standard uniform crossover as used in classical GAs. The operator swaps genes from two chromosomes based on the crossover rate. The mutation operator replaces the existing gene value with another value from a list of all possible time shifts $(-h_{\min}/2,\ ...,\ 0,\ ...,\ +h_{\min}/2)$ based on a predetermined mutation rate.

Computer Program

In order to test the methodology, a computer program has been developed. The development platform is dBASE Plus and a database that consisted of several dbf files has been created. The database stores the necessary data to perform the calculations in the optimization program. Figure 1 depicts the Graphical User Interface (GUI) for the program.

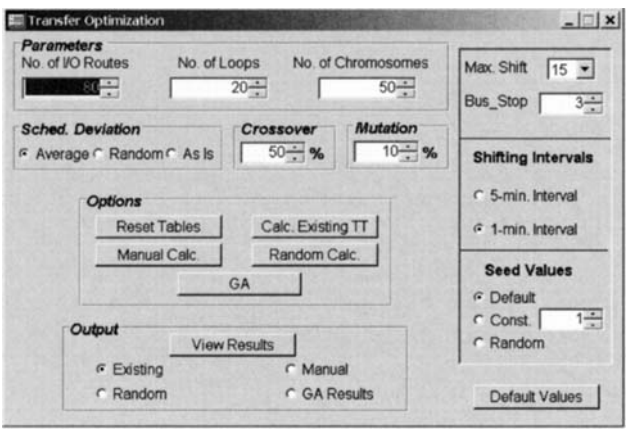


Figure 1. Transfer Optimization Software GUI

The user interface allows for entering parameters such as the number of directional routes, desired number of iterations or generations in the genetic algorithm, the number of chromosomes, as well as crossover and mutation rates. Selections for random arrivals/departures, average schedule deviation values, or scheduled arrivals/departure values are also included in this form, as well as maximum shifting values, walking times between transfer stops, and shifting intervals. Lastly, the GUI allows for the selection of seed values, which are used in the generation of random numbers.

Case Study

The input data include definition of 40 routes (80 directional routes), 255 transfer stops, and scheduling data from Broward County Transit (BCT). To test the genetic algorithm, a dataset for ridership and schedule deviations is constructed based on data obtained from BCT.

With a Dell Precision M60 with a Pentium M processor 1.7 Ghz, 1GB RAM, and a 60GB (7200 rpm) hard drive the transfer optimization application was run several times. Different crossover and mutation rates as well a combination of generations and chromosomes were tested and the median values of transfer times were used.

Using this methodology, the total transfer times were reduced from 3,150.37 to 2,728.32 hours per day, representing a significant 13.4% improvement, after 12.36 hours of CPU time. Figure 2 shows the results from these iterations by plotting transfer hours saved against CPU time. The efficiency of the genetic algorithm depends on the performance of its operators. Cevallos Fabian and Fang Zhao (2006) reported that a crossover rate of 50% and a mutation rate of 10% produced best results. The best transfer time in this case study was obtained using a combination of 40 generations and 60 chromosomes.

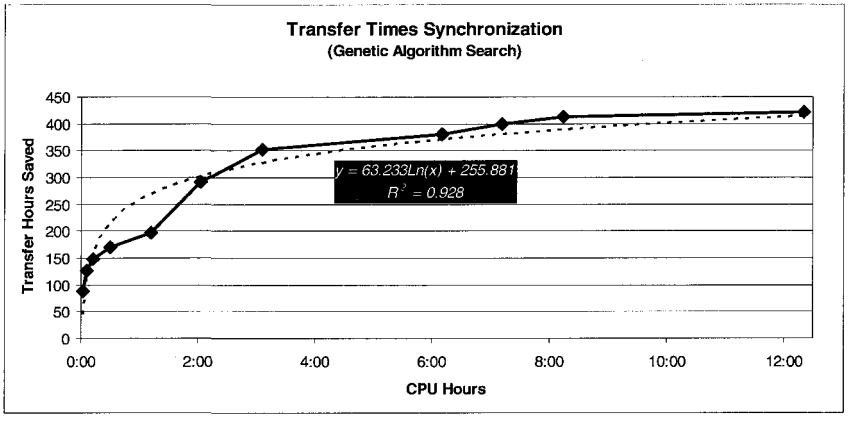


Figure 2. Computational Results

Summary and Conclusion

This research introduces a systematic approach to synchronization of bus arrival times in a transit system of multiple bus routes and numerous time points to minimize total user transfer time. It utilizes a procedure based on business knowledge and makes use of the genetic algorithm to find the optimal solution. The proposed approach is demonstrated to be capable of solving the schedule synchronization problem.

The approach is computationally feasible and is relatively easy to implement (Cevallos and Zhao 2006). The results show that significant transfer time savings are obtained after a reasonable amount of computation time. The methodology is flexible in terms of the type of optimization model used and may be easily improved or modified. The approach is easy to understand and implement, and can use readily available scheduling and transit technology data from most transit agencies. And the methodology can be easily incorporated into the scheduling process, making it likely to be accepted by transit schedulers.

Further research is needed to increase the algorithm efficiency with parallel computing and to determine the best genetic algorithm parameters. Innovative algorithms like simulated annealing or other evolutionary strategies are also worth pursuing. Finally, there is a need for improved techniques to capture bus arrivals and ridership data from existing transit technologies to facilitate the data gathering, formatting, and manipulation processes, so the data may be easily available to be used in the transfer time optimization algorithm.

References

Cevallos Fabian and Fang Zhao (2006). Minimizing Transfer Times in a Public Transit Network with a Genetic Algorithm. Compendium of Papers of the *85th Annual Meeting of the Transportation Research Board,* Washington, D.C.

Deb, K. and P. Chakroborty (1998). Time Scheduling of Transit Systems with Transfer Considerations Using Genetic Algorithms. In David Montana (Ed.) *Special Issue on Evolutionary Algorithms for Scheduling , Evolutionary Computation*, 6(1), pp. 1-24.

Kikuchi, S. and J. Parameswaran (1993). "Solving a Schedule Coordination Problem Using Fuzzy Control Technique." *Proceedings of the Intelligent Scheduling Systems Symposium*, ORSA-TIMS, San Francisco, CA.

Chinese Smart Bus Demonstration Project and Its Implementation

Wu Sufeng[1], Yang Xiaoguang[2], Lin Zhongshuai[3], Huang Canbin[4]

ABSTRACT

This paper describes the results of a research project to better understand bus riders' satisfaction with and the change of the operating index of Zhongshan's Smart Bus, a national ITS demonstration project in China. The objective of the research is to let the demonstration project play an important reference role to the development of national ITS-Transit in China by collecting and analyzing data associated with bus rider' perceptions related to the use of the Smart Bus and the operating index of Smart Bus. The research methods involved a questionnaire investigation of system users and contrasting the bus' operating indexes before and after the demonstration project implemented. The overwhelming majority of the Smart Bus users survey reportes that Smart Bus could improve punctuality, safety and so on. Through contrasting the bus' operating indexes we have found that Smart Bus reduced average fuel consumption and maintenance expense. But it also shows that there are some disadvantages of the demonstration project. For example, the intelligent dispatching subsystem was not used fully.

1. INTRODUCTION

This paper studies the effect of the Smart Bus demonstration project in Zhongshan, China. Firstly we survey the citizens' response and satisfaction degree with the demonstration project. Then we compare the operating indexes of the bus after the project with those before the project. Finally the advantages and disadvantages of the demonstration project have been brought forward after analyzing. The purpose of this paper is to help other cities understand the implementing effect of the Smart Bus demonstration project and let the demonstration project play an important reference role to the development of APTS(Advanced Public Transportation System).

1.1 Background

In order to promote Chinese Intelligent Transportation System(ITS)' development, nine cities had been chosen as the first group of ITS demonstration cities by the Science and Technology Department of China in 2002. Zhongshan was selected as the Advanced Bus System demonstration city and the project will be carried out in the Zhongshan Bus Company. Demonstration Project is classified three stages: near-term, medium-term and long-term stage.

---

1 PH.D. Candidate, Department of Transportation Engineering ,Tongji University, 1239 Siping Road, Shanghai, P.R.China, 200092,Email: wusufeng@126.com

2 Professor, Tongji University

3 Master's degree candidate , Tongji University

4 Master's degree candidate , Tongji University

The near-term project has been put in practice *(1)*. The project was finished on October 2004. In order to make the project applicable all around the country, Zhongshan ITS Led Subgroup commissioned Tongji University to evaluate the effect of this project. The main data of this paper are obtained from the evaluation.

1.2 Zhongshan's Smart Bus Demonstration Project

The demonstration project is based on the Zhongshan Bus No.1.There were 14 electronic stop boards being installed along the route. Passenger counter and vehicle logic unit were also installed on the 36 buses operating on the route of No.1 bus. Dispatching and operating management data centre and monitoring scheduling system were constructed at Zhongshan Bus Company. Through the construction, the Bus Company has mostly achieved these functions, such as automatic passenger counter, automatic bus-stop-reporter, real-time dynamic travel time prediction, vehicle running status real-time monitor and vehicle intelligent scheduling system, etc.

1) Electronic stop boards (2)

14 electronic stop boards have been installed along the route of No.1 bus. Main functions of the electronic stop boards are as follows:

- Show the operating time and the distance to the stop of the next bus;
- Display the crowd condition on the arriving bus;
- Release other information, such as road, weather, traffic information, etc.

2) Car-mounted equipments (2)

Vehicle logic unit (GPS, GPRS, bus-stop-reporting automatically integrative machine) and passenger counters were installed on the 36 buses of the No.1 bus. Vehicle logic unit primary functions include:

- Reporting the stops manually and automatically;
- GPS Satellite Positioning System;
- Bidirectional alternation function of GPRS information.

3) Data Center (2)

The data centre of dispatching and operating was built in the Zhongshan Bus Company, Primary functions are as follows:

- GPRS communication service;
- Service of the stop board;
- Data storage service.

4) Monitoring and controlling center

Monitoring and Controlling Center includes controlling subsystem, public transit dispatching subsystem in real-time, intelligent scheduling of vehicle subsystem.

2 PREVIOUS STUDIES

With the development of ITS technology, APTS has been experiencing huge development in

theory and application. Similar systems have been developed in USA, Japan, Europe, etc respectively *(3)*. With the continual development of APTS, the result evaluation of APTS is being deepened constantly too. American Ministry of Transportation has announced the guidance files for appraising APTS, such as Public Transportation Agencies Performance an Evaluation Approach Based on Data Envelope Analysis*(4)*; Evaluation ITS Technology for Bus Transit Systems*(5)*,and so on.

European Union has also gotten great development in the research of APTS evaluation *(3)*. Japan has also carried on deeper exploration in APTS evaluation *(3)*. China carried on deep research to APTS evaluation in recent years too, but limited by the restriction of our own development of APTS, the research of appraising the method is in the theory aspect basically.

There are some scholars studied the impact of real-time information on transit passenger behavior. Kelley Klaver Pecheux, etc. have studied customer use of and satisfaction with real-time bus arrival information in Portland *(6)*.Mishalani,McCord and Wirtz studied the relationship between perceived and actual waiting times experienced by passengers awaiting the arrival of a bus at a bus stop*(8)* .

3 RESEARCH APPROACH

The Smart Bus demonstration project in Zhongshan is an important component of national ITS demonstration projects, which will play an important reference role to the development of national APTS. In order to be just, the effect evaluation of the project mainly adopts the way to investigate and analyzes the relevant data in the study, including contrasting the data before and after the demonstration project, and the data of demonstration route with normal route (this paper compares the data of No.1 bus with corresponding data of No.9 bus and No.12 Bus because the operating condition of these three routes are similar). The appraisal includes two aspects mainly:

- Citizens' satisfaction degree with the demonstration project.
- The change of the operating index *(3)*.

4. PUBLIC SATISFACTION ANALYSIS

In order to obtain the public's (including general citizens and passengers who take No.1 bus) satisfaction with the system, questionnaire investigation was the main method. The questionnaire investigation of the passengers of No.1 bus was held twice, the first date was on March 6 , 2005 (Sunday), the second date was on March 8 , 2005 (Tuesday), both of the investigate time was from 7:00 to 19:00. The time arrangement like that not only can reflect characteristics and feeling of passengers on the work end and the working day, but also can compare these data with relevant data investigated on March 2 , 2003 (Sunday) and on March 4 , 2003 (Tuesday) . The questionnaire investigation of general citizens in Zhongshan was launched from March 2, 2005 to March 7, 2005, mainly to understand the satisfaction degree of general citizens with the system in Zhongshan, including persons who take public transit and persons who do not. It can be used for contrasting with the investigation data of passengers on No.1 bus.

1) Cognitive Degree to the Information System

The investigation shows 63.4% of the passengers of No.1 bus and 61.4% of the general citizens noticed the electronic stop boards installed This indicates public information subsystem of the demonstration project has been paid more attention to by passengers and general citizens.

2) Validity Evaluation of Information System

Twenty-third percent of the passengers and general citizens considered the information that the electronic stop board released was very useful. Fifty-six percent of the passengers and general citizens thought the information that the electronic stop board released was useful, only 3% of the passengers and 2% of the general citizens thought the information system was useless.

3) Accuracy Appraisal of Information System

The survey of passengers of No.1 bus showed that 61% of the passengers thought the stop board provided accurate information.13% of the passengers thought the information were very accurate, only 5% of the passengers thought the information was inaccurate, 21% of the passengers expressed that they had not paid attention to that.

4) Impact Analysis of People Experience Waiting for Bus

After the electronic stop board was installed, 42% of the passengers said that they feel waiting time was shorter than before, but 30% of the passengers thought that they feel waiting time had no change, only 2% of the passengers thought the time was longer than before. Investigation showed resident's average feeling waiting time was high up to 13.1 minute, which was far greater than the waiting time in point of fact that was about 2.44 minutes. Therefore the passengers' and general citizens' waiting bus feeling are greater than the time of waiting actually. Thus offering real information helps to change people's traditional understanding of the bus, and will increase the appeal of the bus.

5) Will of Taking Bus More Analysis

Thirty-six percent of passengers of No.1 bus and 24% of the citizens said they were ready to take buses more if other public transit lines were improved like No.1 bus; Fifth-second percent of the passengers of No.1 bus and 56% of the citizens would take more buses probably; while 12% of the passengers of No.1 bus and 20% of the citizens said that their will of taking bus would not change.

5. ANALYSIS OF SMART BUS OPERATING INDEX

Several operating indexes of the Smart Bus were studied before and after the demonstration project implemented, such as safety, punctuality, speed, turn-round time, average fuel consumption, etc.

1) Safety

Since the project came into operating, the buses of No.1 bus operated 730,000km altogether without any traffic accident happened. But before the system came into operating, the buses operated 3,649,000 km altogether from January of 2004 to October with 6 traffic

accidents happened. It is obvious that the system's security has gotten greater improvement after the demonstration project implemented than before.

According to the analysis, the improvement of the security mainly comes from the following aspects:

- Dispatching Centre monitors the drivers' behaviors in real-time, and monitoring system shows and notes the driving speed in real-time, and gives a warning information when the vehicles are over speed. Because of restricting the drivers ' unsafe driving effectively, the traffic accidents have been reduced.
- The drivers do not need report the stop manually, that can reduce drivers ' operation processes while arriving in or leaving out the station, and let the driver drive attentively, then improved the safety.

2) Punctuality

Researchers compares the interval reaching time of No.1 bus with that of No.12 bus which operating condition is very similar to No.1 bus at the Jinhua Hotel Station on March 6, 2005. Through statistical analyzing, the interval reaching time of No.1 bus at that station is 5.67 minutes. The departing interval is 5 minutes in this period. On the contrary, the interval reaching time of No.12 bus at the same station is 6.6 minutes on average while the departing interval is 5 minutes too. It is obviously that punctuality of No.1 bus is better than No.12 bus.

3) Overall Trip Speed

After the project, the vehicle's average operating speed is 20. 68km/h, while before the project in 2003 it is 22.21km/h [7].

According to the investigation and analysis, the mainly reasons that caused decrease of the speed are:

- the number of motor vehicle on the road in Zhongshan increased quickly;
- More stops had been set up along the route of No.1 bus that influenced the speed of the public transit.

In order to improve the speed and attraction of the public transit vehicle, it need to give priority to the public transit vehicle in time and space, such as setting up the public transportation lane and the public transit priority signal, etc.

4) Vehicle Turn-Round Time

In theory, Dispatching Centre controls the vehicle in real time, can prevent the driver stop at will, which can improve the speed of the bus, thus reduces the operating time. But at present, because the intelligent system did not fully function, the turn-round time had no change.

5) Average Fuel Consumption

The fuel consumption was 0.305l/km/veh after the project from November to December on average, while it was 0.322 l/km/veh before the project from January to October, 2004, had reduced by 5.3%. The fuel consumption on average was 0.369 l/km/veh from November to December 2002 before implemented. The fuel consumption was 0.321 l/km/veh at the same period in 2003.It was obvious that the fuel consumption reduced after the project. The reasons are mainly as follows:

- Dispatching centre supervised the operating state of the vehicle constantly.

- The driver had to drive the vehicle more smoothly, avoid accelerating, moderating frequently. So the consumption of the fuel reduced.
- The Managing Centre could detect the location of bus in real-time, which can avoid vehicles being used privately by the drivers, reduced the oil consumption also.

6. CONCLUSION

Through the survey of passengers of No.1 bus and general citizens, 84% of the passengers and general citizens thought the information that the electronic stop board released was useful. Sixty-first percent of the passengers thought the stop board giving accurate information, while 13% of the passengers thought it gave very accurate information. 42% of the passengers expressed feeling the waiting time was shorter after the electronic stop board was set up .We can draw a conclusion from analysis mentioned above that the implementation of the demonstration project received well appraise from citizens and passengers.

Through the operating data analyzing of the Advanced Bus System after the project implemented, the safety of the bus has improvement. The interval reaching time of the vehicle is more equality. The passengers ' waiting time reduces effectively. The average fuel consumption and repair frequency are also decreased, and the operating incomes of No.1 bus increase by 1.5% during the two months after the project implemented compared with that of the same period in 2003. It is obvious that the Smart Bus does well "reducing expenditure" and "increasing income ".

REFERENCES

1. Zhongshan Construction Bureau, ITS Research Center of Tongji University. The Study of Advanced Bus System Demonstration Project Implementing Scheme in Zhongshan.2003.
2. Zhongshan Construction Bureau, Hisense Network Technology Co.,LTD. Guangzhou Kanglejin Information System Co., LTD, Zhongshan Bus Company, Material of Final Acceptance of Construction of Advanced Bus System Demonstration Project.2005.
3. Tongji University. The Study of Evaluation Method of Intelligent Transportation System (ITS) Project, 2004.
4. Pasquale Carotenuto, Paolo Mancuso and Luca Tagliente. Public Transportation Agencies Performance: an Evaluation Approach Based on Data Envelope Analysis,2001.
5. Randolph Hall, Maged Dessouky, Lei Zhang, Ajay Singh, Vishal Patel. Evaluation of ITS technology for bus transit systems,1999.
6. Kelley Klaver Pecheux, Pamela J. Vandergriff.Customer Use of and Satisfaction with Real-Time Bus Arrival Information in Portland, Oregon.84th Annual Meeting of the Transportation Research Board, 2005.
7. Zhongshan Construction Bureau, Tongji University. Report on Analysis of Zhongshan's public transit system.2003.
8. Rabi G.Mishalani,Mark M.McCord and Hohn Wirtz,Passenger Waiting Time Perceptions at Bus Stops. 84th Annual Meeting of the Transportation Research Board, 2005.

Urban Development with Sustainable Public Transit Services

Z. W. Wang and Hong K. Lo[1]

[1]Department of Civil Engineering, Hong Kong University of Science and Technology, Hong Kong, China; PH: (852) 2358-8472; Fax: (852) 2358-1534; email: cehklo@ust.hk

Abstract

It is known that urban density is positively correlated with transit service usage. Previous studies, however, analyzed the problem descriptively based on statistical approaches. This study seeks to derive prescriptive results of the relationship between urban density and rail service sustainability. As a start, this study considers an idealized metropolitan region with a uniform urban development density and a central business district (CBD) at its center. Trips generated from the region to the CBD are either served by the rail service or private cars, as described by a discrete choice modal split model. We study the sensitivity of urban development density on the financial sustainability of the rail service by examining the supply and demand patterns. Through the analysis, the result sheds light on the threshold urban density required, below which the service cannot be sustained without subsidy.

1. Introduction

Public transit services (PTS) are important assets to any major city. It is ideal if the PTS are financially sustainable, with affordable fares and expedient quality. The key lies in how the urban form and density are planned to ensure their sustainability. The relationship between urban density and transport characteristics has been studied. It has been shown that urban density is positively correlated with PTS usage. Previous studies, however, analyzed the problem descriptively based on statistical approaches. Newman and Kenworthy (1999) found that higher urban density is positively correlated with lower levels of car ownership and car use, and higher levels of transit use. Sinha (2003) pointed to a similar conclusion, that the fundamental variable to increased transit use and sustainability is urban population density or activity density. From the literature, there is no doubt that high density is positively correlated with high transit usage. These descriptive results are important and stimulating. Yet they are not sufficiently refined to be adopted as policy guidelines. Basically, how high should the urban density be for the provision of sustainable PTS? A prescriptive answer provides important guidelines for integrating the planning of urban development with sustainable PTS. Granted high density developments are not for every region, the answer at least allows planners to understand the implications of

PTS quality and sustainability in relation to a certain density development. The answer to this prescriptive question is specific to the type of PTS under consideration. In this study, we consider the case of rail service in an idealized city.

2. Model Formulation

Consider an idealized metropolitan region (Figure 1) with a CBD located around terminal O on which two highways and a rail intersect. This region is symmetric with respect to the rail line. The demand for travel generated from any location in the metropolitan region to the CBD is served by either the rail or private car (or taxi). Individual commuters choose one of the two modes according to their generalized costs. If one chooses the rail service, he or she will have to walk to the nearby station; in this first study, we assume that connecting access modes (such as minibus) are not available. On the other hand, if one decides to use private car, he or she will also have to drive to nearby highway entrance to get on the highway. So, competition exists between these two available modes. The rail service can improve its service quality to attract more passengers so as to sustain itself financially. The issue of financial sustainability is also related strongly to urban density. For an area with a low urban density, it could be that no matter how much the rail service improves its service quality, there just is not enough demand to sustain its service, not to mention the cost spent for the service improvements. In other words, there exists a threshold urban density below which the rail service simply cannot sustain.

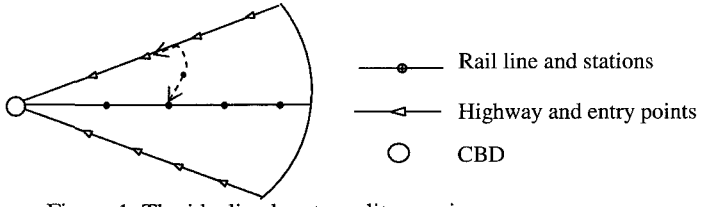


Figure 1. The idealized metropolitan region

Assumptions. To simplify the analysis, as a first study of this kind, we make several assumptions as described below, some of which can be removed in future studies:

a. Uniform urban density: It is assumed that the urban density is homogenous. This assumption can be removed readily with a more refined numerical model.
b. Trip generation: It is assumed that the population covered by the rail line and the highways make similar number of trips on a daily basis.
c. Logit modal split: The logit model is used to describe passengers' combined mode and station (interchange point) choices, in which their total costs, including the travel time and monetary cost, are considered.
d. Fixed inter-station spacing: The inter-station spacing is considered as fixed and given. This deserves more detailed analysis in the future.

e. Rail operation strategy: The operations characteristics, including fare and train cruising speed, are regarded as fixed. On the other hand, we allow the headway to vary in order to study its effects on cost and revenue.
f. Sufficiently high rail capacity: The capacity of the rail service is sufficient to carry all the demand and the crowdedness effect is ignored.
g. Highway congestion: The travel time of highway traffic is determined by the BPR function to take into account the delay due to traffic congestion.
h. Vehicle occupancy: It is assumed each private car carries one passenger.

Parameters. The variable and parameter definitions are as follows:

d urban density
t trip generation rate
r_i distance between the ith railway transit station and the CBD
l_i distance between the ith interchange point and CBD
v_w walking speed
v_c average private car speed on minor street
v_t average cruising train speed
v_i average private car speed on the ith highway link counted from the CBD
f_i traffic flow on the ith highway link counted from the CBD
C_i road capacity for the ith highway link
h_t average transit service headway
γ_c cost of operating a car per unit time
γ_t average value of time
D_i^c traffic demand for the ith interchange point on highway
D_i^c passenger demand for the ith railway station
F_i transit fare for service from ith station to CBD
γ_e average daily cost for owning or hiring one vehicle
γ_p average operating cost per vehicle-hour
κ average construction cost per day
K total service time
N total number of interchange points on highway

Disutility Functions. The disutility function for the rail service includes passengers' access time to the service, station waiting time, travel time, and fares. Walking is taken to be the access mode to the rail service and the walking distance is simply the Euclidean distance between the origin and the intended station.

Consider a trip maker from a location with polar coordinate (ρ,θ), where θ is defined to be 0 along the rail, and $\pm\theta_0$ along the highways. We define $u_i^t(\rho,\theta)$ at

(ρ,θ) to be the disutility of a travelers walking to the i th railway station and taking train to the CBD.

$$u_i^t(\rho,\theta)=\gamma_t\left[\left(\rho^2+r_i^2-2\rho r_i\cos\theta\right)/v_w+\sum_i r_i/v_t+h_t/2\right]+F_i \tag{1}$$

As for the disutility function of private car usage, it is assumed that dense circumferential and radial streets are available to connect the metropolitan region to the highways. Private car users first drive on the street network so as to get access to the highway to CBD. The highway is accessible only through interchanges. The route of a private car to CBD is schematically shown by the dashed line in Figure 1. The travel time on the highway is determined by the BPR function. Similarly, by defining $u_i^c(\rho,\theta)$ as the total cost of the trip maker using private car from his origin at (ρ,θ) to the CBD through the ith interchange point in highway, it can be expressed as:

$$u_i^c(\rho,\theta)=(\gamma_t+\gamma_c)\left[\left(\rho-r_i+\rho(\theta_0-\theta)\right)/v_c+\sum_i(r_i/v_i)\left(1+\alpha\left(f_i/C_i\right)\right)^\beta\right] \tag{2}$$

Patronage Estimation. Suppose the trip generation follows a linear relationship with the population residing in an area, the expected number of trips for a given region Ω can be written as:

$$T=t\cdot\left(d\cdot Area(\Omega)\right) \tag{3}$$

With the disutility functions defined, the passenger demands for different modes and different stations (interchange points) can be estimated by the logit split model. Traffic congestion on the highways makes the travel costs of private car users vary with the traffic flows. Thus, to solve for the demand split, one must solve the system of nonlinear equations described above. Firstly, trip makers attracted to each interchange point on the highway is determined according to the logit split model. For the i th interchange point, the demand is:

$$D_i^c=\int_\rho\int_\theta t\cdot d\cdot\exp\left(u_i^c(\rho,\theta)\right)/\left[\sum_j\exp\left(u_i^c(\rho,\theta)\right)+\sum_j\exp\left(u_i^t(\rho,\theta)\right)\right]d_\rho d_\theta \tag{4}$$

Therefore, for the i th link from the CBD, the traffic flow is expressed as:

$$f_i=\sum_i^N D_i^c \tag{5}$$

Solving the system of equations formed by (4) and (5), we get the equilibrium traffic flow on each highway link. With the obtained traffic flows, the patronage for the rail service can be derived:

$$D_i^t=\int_\rho\int_\theta t\cdot d\cdot\exp\left(u_i^t(\rho,\theta)\right)/\left[\sum_j\exp\left(u_i^c(\rho,\theta)\right)+\sum_j\exp\left(u_i^t(\rho,\theta)\right)\right]d_\rho d_\theta \tag{6}$$

Total Costs of Rail Service. Total costs of the transit service include both the fixed construction cost, and the operation cost that mainly depends on its headway h_t. The revenue is expressed as a negative term to offset the cost:

$$TC = \kappa + \gamma_e \left(\left(\sum_i r_i / v_t \right) / h_t \right) + \gamma_p \left(K / h_t \right) - \sum_i F_i D_i^t \tag{7}$$

The first term of (7) represents the construction cost. The second and third term depict the equipment and operation costs, respectively. The last one describes the collected revenue (expressed as a negative cost term).

3. Solution Process

In this model, urban density and headway of the rail service are two independent variables. Given a particular urban density, the rail service varies its headway in order to minimize its total costs. By determining the optimal headway and corresponding total cost for a particular urban density, we can study the sensitivity of the total cost to the urban density. Or, one can estimate, for a given urban density, the minimum total cost achievable by optimally setting the operation strategy.

The solution process of this model is straightforward. If both the urban density and headway are fixed, the total costs of the rail service as shown in (7) can be determined by solving the nonlinear equation system (4)-(6) through the Classical Newton method. Then, we relax the headway variable and obtain the total costs for different headways. Of course, the headway should be restricted within the feasible range. Finally, the relationship between the total cost and headway is derived through regression analysis. The result shows that a linear relationship is obtained between the total cost and headway based on the limited computational results obtained in this study. This result can be interpreted that headway changes in the feasible range do not affect passengers' choice substantially. This is possibly due to the fact that in this numerical example, waiting time constitutes a relatively small portion of the total journey time. Then, we can get the optimal total costs for any given urban density level. After that, we vary the urban density and repeat the above process to study the sensitivity of urban density on total costs of the rail service.

4. Case Study

To illustrate the problem, an example involving two railway stations and two interchanges on the highway is implemented. The parameters used are as follows:

$v_w = 4$ km/h $v_c = 40$ km/h $v_t = 100$ km/h $v_1 = v_2 = 60$ km/h $C_1 = C_2 = 5000$ *vph*

$\gamma_c = \gamma_t = \$40$ per hour $\gamma_p = \$8$ per veh·h $r_1 = l_1 = 6$ km $r_2 = l_2 = 10$ km $K = 10$ hr

$\gamma_e = \$10000$ per train per day $t = 0.1$ trip/population $\kappa = \$140000$ per day

The total costs of running the rail service under different urban densities and headways are presented in Figure 2. For simplicity, only 3 cases are plotted to illustrate the results. From this figure, again the total costs depend substantially on

the urban density. When the density is high (13500 persons/sq. km), there is profit (expressed as negative total cost) to be gained. If one is aiming at achieving cost-revenue neutral, from this result, it appears that an urban density of 13,000 persons/sq. km is required. We, however, caution that this number is based on the limited result of this study. But it illustrates what possible questions one can answer with this modeling framework.

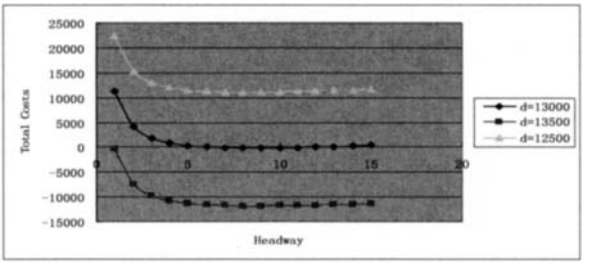


Figure 2. Total costs for rail transit service

5. Concluding Remarks

This study developed a model to analyze the relationship between urban density and sustainable rail transit service. The results provide some crude guidelines for integrating the planning of urban development with sustainable rail service. However, we emphasize that in this preliminary study, to simplify the model, we have adopted a number of assumptions, which limit the generality of this study. Currently, we are extending this model for realistic scenarios in Hong Kong by relaxing these assumptions, as well as considering the effect of feeder bus services. We anticipate that this model, once developed and calibrated, would be useful for providing guidance for the development of transit-oriented cities with sustainable transit services. In particular, the relationship between density of urban development and sustainable rail services, despite to a first order accuracy, would be instrumental for land-use planning.

Acknowledgement
The study is partially supported by the Competitive Earmarked Research Grant #617105 from the Research Grants Council of the Hong Kong Special Administrative Region.

References

Newman, P. and Kenworthy. J. (1999). "Costs of automobile dependence: Global survey of cities." *Transportation Research Record*, 1607.

Sinha, K. (2003). "Sustainability and urban public transportation." *ASCE Journal of Transportation Engineering,* 129, 331-341.

THE 1st TRUE HIGH SPEED RAIL SYSTEM/"BULLET TRAIN"

FOR THE USA

Mike Lehman

4600 N Clarendon, #1211, Chicago, IL 60640 usbullettrain@gmail.com, 773-334-6080

Abstract

Advanced countries are implementing "true" High Speed Rail/HSR systems and the US is earnestly trying to also; of the many concepts proposed, the Great Lakes HSR-GL/TGV system should be the **one built**. Many millions of people would be able to use the system and even more benefit from it's numerous advantages.

I've received positive reviews relative to this concept from academics, consultants, the transport industries and others. This is **not** the Midwest HSR initiative, rather, another transportation choice/mode, a separate dedicated "true" HSR / "bullet train" system. The Great Lakes to North East US regions=25% of all US inter-city travel by road and air!

The benefits of the brilliant safety records(no deaths on similar decades old Shinkansen or TGV HSR systems), non-reliance on oil(electric powered), less pollution(air and noise), and less road congestion true HSR/TGV systems offers outweigh the initial startup costs and land expropriations necessary for this new transport system.

Commercial jets expel thousands of gallons of petroleum exhaust into the atmosphere and create dreadful amounts of noise(TGVs uses domestic coal and other alternative electric power and is much quieter). Ohare airport generates thousands of additional traffic congesting and polluting vehicles daily-not a concern with the Great Lakes/GL/TGV central business district/CBD or current Northeast HSR corridor/NEC CBD destinations.

Astoundingly!, estimates of life expectancy of people that live within several miles of a major airport is reduced by 6 or more years due to toxic airplane emissions. In Illinois, it's also reported that the air pollution created by Ohare airport alone is greater than all electric power plants in the state combined! TGV, an alternative to more airplanes/autos.

This TGV system would conserve over 2 billion gallons of fuel a year(500,000 flights), relying on alternative energies. In addition, a new airport consumes double the land that the entire GL/TGV system concept would, 15,000 vs. 7,000 acres. Lastly, discount airlines with multiple airplane/airport transfers per route have as long or longer travel times in the Northeast quarter of the US than most GL/NEC HSR route travel times.

The Great Lakes TGV corridor would connect **45 major US city pairs** and hence, many intercity passengers while other proposed US HSR system concepts connect only about a **dozen** or so major city pairs. In the Northeast and Great Lakes corridors there are about 1-2 billion individual intercity trips annually, consequently, the 40 million trips a year or more estimated for the GL/TGV system seems very attainable. There is existing infrastructure throughout Pennsylvania to facilitate HSR travel amid the mountains there- the major concern in adaptation of this TGV concept. The time is now to build **true** HSR.

Introduction

Justification of dedicated TGV-High Speed Rail/HSR between Chicago and Philadelphia
The Great Lakes(GL/TGV) on to the NEC

This concept is for an exciting, strategic and practical HSR "bullet train"/TGV type project. The TGV is the French electric HSR design-system applied that uses both "dedicated", and existing infrastructures and rail track-ROW. The energy/economic, security, and transportation/health rationales for this **new, dedicated** mode would be mostly for servicing the states of Illinois through to New York/ Jersey, connecting the cities of Chicago, Cleveland, Pittsburgh, Philadelphia and New York. Detroit and Cincinnati/Ohio can be individual HSR/TGV line origin-destination points(total linked US HSR systems' state populations are over 140 million, about 1/2 of the US population).

Extra states would benefit by their HSR link to **Acela/Northeast corridor**/(NEC) service or by other modes to any city HSR station, including ones connected radialy to Chicago by conventional trains. The overall population reach serviced by both the GL/TGV and NEC HSR systems combined is well over 170 million people in 18 states- **3 times the French TGV population source!** Philadelphia would be the logistic hub where Great Lakes TGV corridor trains would meet the Northeast HSR corridor and either terminate there or continue on, alternating either northbound to NYC/Boston or southbound to Baltimore/Washington DC, or, even perhaps east to Atlantic City/the Atlantic Ocean.

The Great Lakes(GL/TGV) mode could carry in excess of 40 million passengers a year, drawing travelers from air and bus but mostly automobile modes in addition to acquiring induced new travelers. Over the very long functional life, the properties of the GL/TGV and large initial capital investments would prove to be very productive/efficient. Relative past costs and subsidies of the above mentioned cities' air transport, interstates and highways were far more expensive than what this new TGV route's would be in contrast.

40 million GL/TGV passengers a year is equivalent to about 1/3 of commercial aviation enplanements in the Great Lakes/Northeast corridor cities of the over 600 million annual

domestic enplanements in the US. In Japan(pop. 120 million) HSR usage is over 130 million trips a year; in France(pop. 55 million) HSR usage is over 20 million trips a year.

This proposal will apt to be very unpopular with air and road transportation related industries/lobbies (9 of the 10 largest companies worldwide either produce autos or petroleum products); nevertheless, it shouldn't be since **additional railroad capacity** alleviates some of their modes' problems also. Hopefully progress and logic will prevail and this **new** transportation mode can develop and thrive despite other interests.

ECONOMIC REASONS FOR TGV (only 2% of world oil reserves are in the US, TGV/HSR transport is all electric using domestic coal and other energy resources)

1. The new GL/TGV system linking to the Northeast corridor/NEC interconnects more than 20 culture rich cities; 7 of the 10 largest and most important in the US. The new line would travel from Great Lakes cities through the Alleghany Mountains on to Philadelphia, New York City, Washington DC and the rest of the Northeast HSR(NEC/Acela) cities.

2. There would be new job creation generated by construction and then for continual operation and maintenance of the GL/TGV route(also, more good CBD/downtown jobs). Rider ship levels should reach and exceed the levels of the French TGV ultimately. The French TGV, with over 20 million annual trips, has revenues of over $2 billion a year.

3. With possible revenues of $4 billion or more a year, the large investment in this line's infrastructure and train sets would be paid for realistically within several years time, similar to the French TGV experience with their revenue streams financing and funding.

4. This new TGV route would augment and strengthen AMTRAK abilities and potential elsewhere on complementary routes and that of the Northeast corridor/Acela. Acela/NEC HSR utilization continues to grow and is AMTRAK'S most profitable and popular route.

5. TGV travel mode would enhance cities' CBDs and integrated rail developments there. Proposed connected cities; Chicago, Cleveland, Pittsburg , and Philadelphia have and are expanding upon their own internal transit rail systems-cities not entirely reliant on autos!

SECURITY REASONS (TGV trains could evacuate an entire large city in 1-2 days)

1. The airline transportation mode is more favored for terrorist attacks(hijackings, bombings, sabotage, poisonings etc.) Assaults are not as likely or as catastrophic with the TGV transportation mode-insurance companies and the public would welcome this.

2. New TGV mode of transport wouldn't call for the necessary extreme expense and trouble of security systems and additional equipment like the airline mode requires.

3. In the advent of an airspace shutdown again or bad weather the HSR corridors would serve as another travel alternative to air/road travel in the northeast US and Great Lakes.

MOBILITY/HEALTH REASONS (TGV/HSR<10% the energy use of like air travel)

1. Tragically and extremely costly, about 50,000 people die and tens of thousands more are permanently disabled yearly from roadway related accidents(less driving=less deaths). Hundreds more people are killed and severely injured yearly in plane crashes too. In France and Japan, HSR hasn't had a fatality in over 60 years combined. Scores of people and billions of $s could be saved by using HSR in lieu of personal vehicles and aircraft.

2. Most TGV right of way could be built adjacent to existing highways and rail lines for environmental considerations and land use purposes(aircraft and road vehicles create much more **noise** and **air** pollutions); TGV land expropriations will likely be inevitable.

3. Over **1/3** of all Americans don't like to fly, therefore leaving long, congesting, costly and hazardous auto/bus modes or intricate AMTRAK schedules as their only alternatives.

4. Airport traffic creates more pollutions/congestions around large population centers. There are potentially a total of 8 congestion adding auto trips to and from airports to pickup and drop-off a flyer at both destinations. Combination rail to walking travel modes are always healthier and favorable to alternative airplane to automobile modes.

5. The new dedicated TGV HSR line would travel the 750 mile Chicago to Philadelphia length in 4-5 hours at 186+ mph speeds(which approaches short jet plane trip speeds) with only 3 stops in between (Cleveland, Pittsburg, and Harrisburg). Continuing on to DC, NYC or Atlantic City would add another 1-2 hours to the total overall trip departing the Chicago/Gary station eastbound. New airport **alternatives** are called for!

6. This mode of travel would be especially relaxing and enjoyable. The ability to personally move about, enjoy sights(especially in Pennsylvania), work, talk, eat and rest in a hassle-free, safe vehicle like a bullet train is truly unsurpassed. Indeed, elderly and ADA citizens would prefer this transport option to auto, bus and airplane travel too.

BENEFITS TO INDIVIDUAL STATES (TGVs reduce airports' congestions also)
(connected cities CBDs will gain significant tourist, business, and personal trip activity)
Illinois
The west end of the GL/TGV corridor, linking Chicago to **100 million** + people . The Chicago CBD is positioned to reach another **30 million** connecting travelers by all modes

from adjoining states to the GL/TGV system stations. Moreover, the GL/TGV reduces the need of airport expansions as well as easing roadway congestions to boot!

Indiana
Gary, IN; the US geographic transport pinch point that filters most modes' traffic east and west. Gary/Chicago airport/region development and more use of the South Shore/IC Railroad infrastructure. The suburban Gary/Chicago TGV station would have multi-modal connections; airlines, commuter and HSR rail and major interstate highways.

Ohio/Michigan (TGVs have a **dual purpose** as transit trains in **Cincinnati** and **Detroit**)
The midpoint of the GL/TGV corridor between Chicago and Philadelphia with additional connections originating from Detroit and also Columbus and Cincinnati into Cleveland.

Pennsylvania
Economic development of Pittsburgh and Philadelphia CBDs and the connection to the Pennsylvania capitol of Harrisburg which is also positioned in the state's mountain resort areas along with many other tourist attractions. The advantages of **two** US HSR systems.

*Transportation is the leading cause of accidental/preventable deaths in the US.
**GL/TGV system would be a prudent, comfortable and safe mode of essential mobility that half the US could access, utilize and appreciate-a vital investment. The US should embrace developing and engineering this efficient, alternative transportation technology.

Data collection

Costs for the French TGV

Infrastructure, Land Expropriation and Trainset costs(converted to $ by USDOL/Bureau Labor Statistics, Producer Price Index to 2002) for the First TGV(PSE) HSR line. (Financed by French and international bonds and SNCF internal self-financing)

B=billion, M=million

ff=French francs, ECU=European currency units

Trainset=2 locomotives, 8 passenger cars/386 seats

- Total cost of developing the trains and building the new TGV-PSE line was just under 1,000 M pounds(1984). (Potter 87)

  In 2003, 1,374 M pounds.

- Infrastructure/superstructure: 7.85 B ff; Trainsets(110): 5.3 B ff(1984). (Strohl 83)

  In 2003, Infrastructure: $1.4 B; Trainsets $1 B.

- Capital costs were about $1.6 B using 1983 dollars. (Vranich 35)

In 2003, $2.25 B.

- TGV-PSE project, which cost 18 B ff at 1989 values. (Perren 5)

In 2003, $3.5 B.

- First stage, 417 km/259 miles, TGV-PSE; 9.798 B ff in 1990 value. (Perren 17)

In 2003, $1.2 B.

- Infrastructure(417 km): $2.04 B, Trainsets(109): $1.7 B(1994). (Lynch, Tim 55)

In 2003, Infrastructure: $2.4 B, Trainsets: $2.0 B.

- Infrastructure/TGV: 7.9 B ECU(till 1992). (High level group 166)

In 2003, $6.9 B.

- New line cost of 8.5 B ff; 87 Trainsets, 5.7 B ff(1985). (ECMT/1986 27)

In 2003, Trainsets: $1 B, Infrastructure: $1.4 B.

- Infrastructure investment of 8.5 B ff(1985). (ECMT/1989 75)

In 2003, $1.4 B.

- Total cost of building TGV line and 87 trainsets, 8.5 B ff(1981). (Meunier 197)

In 2003, $2.5 B.

Works Cited

European Conference of Ministers of Transport/ECMT et al. High Speed Traffic Europe Paris : OECD Publications Service. 1986.

European Conference of Ministers of Transport/ECMT et al. Rail Network Cooperation Paris: OECD Publications Service. 1989.

European Conference of Ministers of Transport et al. Why Do We need Railways? Paris:
OECD Publications Service. 1995.

High Level Group et al. "The European High-Speed Train Network". Brussels: Luxembourg Office for the Official Publications of the European Communities. 1995.

Jones, Joseph. High-Speed Trains in France: Paris-Sud Est. Queen's University at Kingston, Ontario. CIGGT, Transportation Development Centre, Via Rail 1981.

Lynch, Tim. The Economics and Financing of High Speed Rail. Tallahassee, Fl: National Urban Transit Institute. 1995.

Lynch, Thomas. High Speed Rail in the US. Amsterdam : Gordon and Breach, 1998.

Meunier, Jacob. On the Fast Track Westport, CT.: Praeger, 2002.

Perren, Brian. TGV Handbook. Middlesex: Capital Transport Publishing, 1993.

Potter, Stephen. On the Right Lines? New York: St. Martins, 1987.

Strohl, Mitchell. Europe's High Speed Trains. Westport , CT.: Praeger, 1993.

Van Den Berg, Leo. The European High-Speed Train and Urban Development. Aldershot, England: Ashgate, 1998.

STUDY OF THE EFFECT OF HETEROGENETY OF TRAFFIC ON INTERCITY ROADWAYS USING COMPUTER SIMULATION

V. Thamizh Arasan[1] and Shriniwas S. Arkatkar[2]

[1]Professor, Transportation Engineering Division, Department of Civil Engineering Indian Institute of Technology Madras, Chennai-600036, India, PH: +91-44-22574270, FAX: +91-44-22574284 Email: arasan@iitm.ac.in

[2]Research Scholar, Transportation Engineering Division, Department of Civil Engineering Indian Institute of Technology Madras, Chennai -600036, India, Email: s_arkatkar@yahoo.co.in

Abstract

Study of the basic traffic flow characteristics and estimation of PCU values are the pre-requisites to develop standards for capacity and service-volume and to formulate effective traffic regulation and control measures. This is better done by modeling the system, which will enable the study of the influencing factors over a wide range. Computer simulation has emerged as an effective technique for modelling traffic flow due to its capability to account for the randomness related to traffic. This paper is concerned with application of a simulation model of heterogeneous traffic flow, named HETEROSIM, to derive PCU values of the different categories of vehicles, taking a stretch of an intercity road in India as the case for the study. The results of the study, provides an insight into the complexity of the vehicular interaction in heterogeneous traffic.

Introduction

Study of the various characteristics of road traffic is necessary for planning, design, and operation of roadway facilities, in addition to regulation and control of traffic. Study of these characteristics by observing various aspects of traffic flow in the field is difficult and time consuming. Hence, it is necessary to model road–traffic flow for in depth understanding of the related aspects. The study of these complex characteristics that cannot be sufficiently simplified to be amenable to analytical solution requires alternative tools like computer simulation. (Banks and Carson, 2000; Law and Kelton, 2000) Simulation is increasingly becoming a popular traffic-flow modelling tool for analysing traffic operations and presently, several traffic-simulation-program packages are available in the market. The available program packages, however, cannot be directly used to study the characteristics of mixed traffic flow as these are based on homogeneous traffic-flow conditions. The research attempts made earlier to simulate heterogeneous traffic flow (Examples: Khan and Maini, 2000; Marwah and Singh, 2000; Kumar and Rao, 1996; Ramanayya, 1988) were limited in scope as they were location and traffic condition specific. Moreover, these studies did not truly represent the absence of lane and queue discipline in heterogeneous traffic. Hence, an appropriate traffic simulation model, named HETEROSIM, has been developed (Arasan and Koshy, 2005) to replicate

heterogeneous traffic flow conditions. This paper deals with the application of the developed traffic-flow model to study the relationship between traffic volume and speed and estimation of PCU values for different categories of vehicles taking the traffic flow on a stretch of an intercity road as the case for the study.

Simulation Framework

On Indian roads, the traffic comprises vehicles of wide ranging static and dynamic characteristics. Because of this heterogeneity of traffic, it is difficult to enforce lane discipline. Hence, the vehicles occupy positions on any part of the road based on space availability. In view of this, an appropriate modeling technique was developed to simulate the stated conditions of traffic flow. The modeling framework considers the entire road space to be a single unit without specific traffic lanes. The vehicles are represented, with dimensions, as rectangular blocks occupying a specified area of road space. The positions of vehicles are represented using coordinates with reference to an origin. The model was implemented in C++ programming language with modular software design. The model is also capable of showing the animation of simulated traffic movements over the road stretch. The flow diagram illustrating the basic logical aspects involved in the program is shown as Fig. 1.

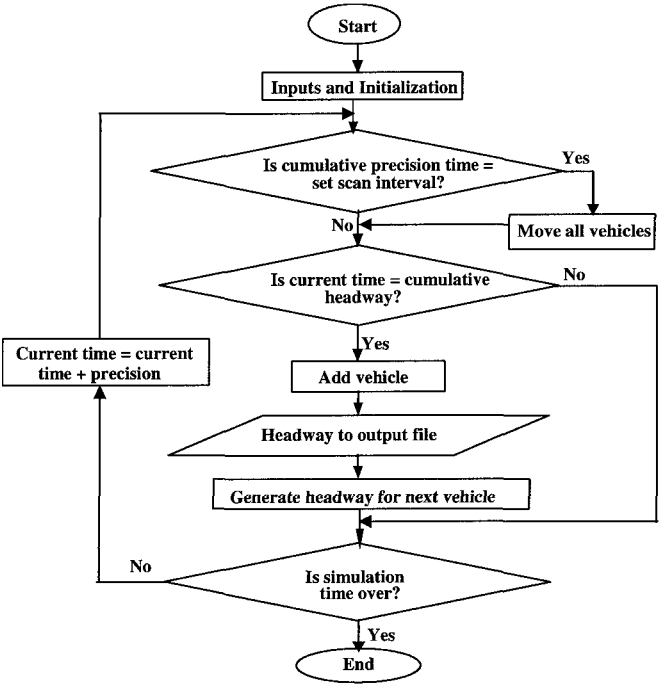


Fig. 1. General Structure of Simulation Model

Traffic Data

The stretch of roadway between km 77.2 and km 77.4, of National Highway No. 45 between the cities, Chennai and Chengalpet, in the southern part of India, was selected for collection of traffic data for the study. The stretch is a four lane divided road with 5m wide central median. The width of main carriageway, with bituminous surfacing, on both sides of the median is 7.5 m. The 1.25 m wide shoulder on both sides is paved with bituminous mix. The field-data inputs required for the model, were collected at the selected stretch using a digital video camera. The video captured traffic data was then transferred to the computer for detailed analysis. The playback of the video made it possible to estimate the volume and composition of traffic. (Nagaraj, et. al, 1990) The data of traffic composition and free speed, as measured at the study stretch of the road and the data of acceleration rate, vehicle dimensions and lateral clearances, as available in the previous study (Arasan and Koshy, 2005), are given in Table 1.

Table 1. Input Data for Heterogeneous Traffic Flow Simulation

Vehicle Category	Percentage Composition %	Free speed (kmph)			Acceleration (m/s^2)	Vehicle dimensions (m)		Lateral Clearance Share (m)	
		Max. Speed	Min. Speed	Std. Devtin.		Length	Width	Min.	Max.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Trucks	34.62	90	53	8	0.9	7.5	2.5	0.3	0.6
Buses	21.28	90	45	10	0.8	10.3	2.5	0.3	0.6
Cars	17.20	115	60	15	1.5	4.0	1.6	0.3	0.5
LCVs	11.40	90	50	6	1.0	5.0	2.0	0.3	0.5
MTW	12.00	75	35	11	1.4	2.0	0.75	0.2	0.4
MThW	1.20	55	45	3	1.0	3.0	1.5	0.1	0.3
Bicycles	2.30	20	10	4.5	0.1	1.9	0.5	0.1	0.3
Tri-cycle	0.00	15	9	3.0	0.07	2.5	1.3	0.1	0.3

*LCVs – Light Commercial Vehicles, MThW - Motorized Three-Wheelers, MTW - Motorized Two-Wheelers

Speed-Flow Relationship

The simulation model was first used to study the basic characteristics of the traffic by giving the data shown in Table 1 as input. During the simulation runs, the traffic volume on the study stretch was varied from 300 to 3000 vehicles/hr. The simulation runs were repeated using three different random number streams to ensure consistency of the results. Through simulation runs, the simulated output values of volumes and speeds were recorded. To draw the speed-flow curve, weighted average

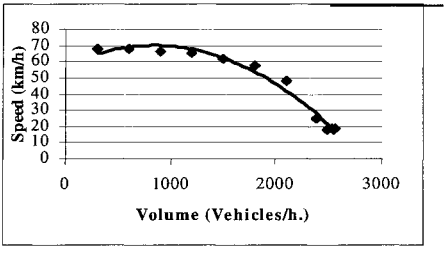


Fig. 2. Speed-Flow Relationship

speeds were calculated for getting the stream speed at each of the volume levels. The speed-flow relationship, thus formulated, is depicted in Fig. 2. From the figure, it can be seen that the typical trend of speed-flow relationship is replicated in this case also.

Estimation of PCU Values

Due to variations in traffic composition, the Passenger Car Unit (PCU) is normally used as the standard unit of measurement of traffic volume. This allows the conversion of a traffic stream of any composition to an equivalent stream of passenger cars so that a common basis for measurement of traffic flow is made available. The PCU may be defined as follows (Arasan and Koshy, 2005). "If on a particular section of road, under particular traffic condition, addition of one vehicle of a particular type per hour will reduce average speed of traffic by same amount as addition of say, X passenger cars of average size per hour, then, one vehicle of this type is equivalent to X cars, i.e., PCU = X for that vehicle". This definition was the basis for estimation of PCU in this study. There have been several attempts to derive PCU values applicable for homogeneous and heterogeneous traffic environments (Elefteriadou, et al., 1997; Chandra and Sikdar, 2000). There is general agreement among researchers that PCU of a vehicle, for a given road width, will decrease with an increase in flow level.

In this study, as a first step in determination of PCU values under mixed traffic streams, it was decided to measure the impact of the different categories of vehicles on the flow of Cars-only traffic stream to have an understanding of the relative influence of the different categories of vehicles on traffic flow. Accordingly, the PCU values for the different categories of vehicles in cars-only traffic streams (simulated for the purpose) were determined for different volume levels for the observed roadway condition. The volume levels considered for simulating the cars-only traffic stream are 500,1000,1500,2000 and 2500 vehicles/hr. The simulation procedure followed for determination of PCU value is as follows. From the cars-only traffic, 50% of cars are removed and an equivalent number of the chosen vehicle type is added (through iterative process) so that they create more or less the same effect on the traffic and the speed of cars remains unaffected. Then, the number of cars removed divided by the number of the subject-vehicle type introduced as replacement for the removed cars, will give the PCU value of the vehicle type.

The results of the above simulation experiment are depicted in Fig. 3 through Fig. 8.The PCU value, along with the cars–only and the mixed traffic (cars and subject vehicles) stream speeds for M.T.W. and M.Th.W, for various volume levels, are shown respectively, in Fig. 3 and Fig. 4. From the figures, it can be seen that the PCU values of M.T.W. and M.Th.W. are increasing initially and then, are decreasing subsequently as the volume increases. This trend may be attributed to the frequent weaving movements of motorized two-wheelers and three wheelers when their movement is constrained due to increase in traffic volume, to maintain the desired speed, which causes significant reduction in the speed of other larger vehicles (cars in this case). When the traffic volume increases beyond certain level (2000 vehicles per hour), however, such movements of these vehicles are not possible and consequently, the relative interference caused to cars will decrease resulting in lesser PCU values

for these vehicles. Fig. 5, 6 and 7 depict the variation of PCU value and the other two parameters, over traffic volume, respectively, for buses, trucks and LCVs. It can be seen that in all these cases, there is a marginal decrease in PCU value with increase in traffic volume, the decrease being significant at the volume level of about 1500 vehicles/hr. In the case of bicycle (Figure 8), it can be seen that the PCU value significantly decreases with increase in traffic volume up to the volume level of about 1500 vehicles/hr. and beyond this volume level, the PCU value increases with increase in traffic volume.

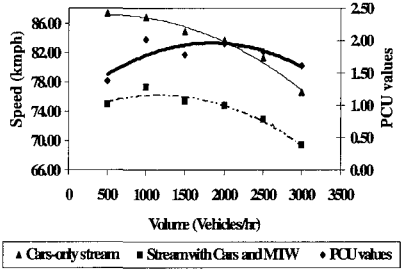


Fig. 3.Variation of PCU value of M.T.W.

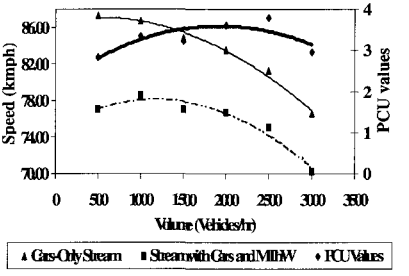


Fig. 4. Variation of PCU value of M.Th.W.

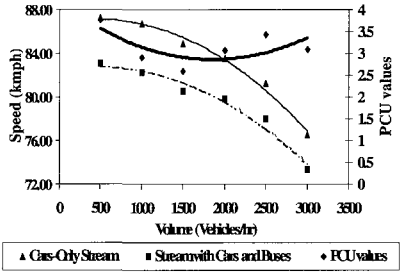


Fig. 5. Variation of PCU value of Buses

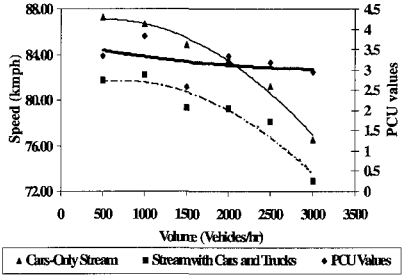


Fig. 6. Variation of PCU value of Trucks.

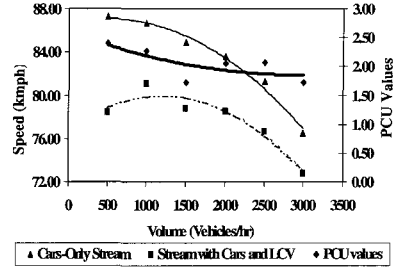


Fig. 7.Variation of PCU value of LCVs.

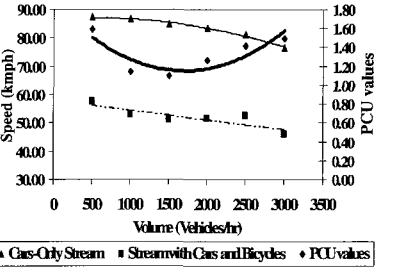


Fig. 8.Variation of PCU value of Bicycles.

Conclusions

The following are the important conclusions of the study:

1. The speed-flow relationship developed for the heterogeneous traffic follows the generally expected trend indicating the usefulness of the simulation-program package, (HETEROSIM) to study the characteristics of heterogeneous traffic flow.
2. The variation of the PCU values depicted in Figures 3 through 8 indicates that while the roadway and traffic conditions (except volume) are held constant, the PCU value of vehicles vary significantly with change in traffic volume.
3. In case of motor vehicles that are smaller than car (M.T.W. and M.Th.W.) the trend lines indicating the change in PCU value, with respect to traffic volume, have convexity upwards. Whereas, the trend lines in the case of motor vehicles that are larger than car (LCV, bus, truck), have concavity upwards. The trend line representing the variation of PCU value for bicycles has a relatively higher order concavity that faces upwards. These variations indicate the highly complex nature of vehicular interaction in heterogeneous traffic flow.

Limitations of the Study

The PCU values estimated through this study is only a factor representing the relative influence of the different categories of vehicles on a cars-only traffic stream. Thus, the study has served the limited purpose of creating a base value of PCU for the different categories of vehicles which may be useful for comparison of the PCU values estimated for conditions wherein the different types of vehicles are simultaneously present in heterogeneous traffic streams.

References

Arasan, V. T., and Koshy, R. Z. (2005). "Methodology for Modeling Highly Heterogeneous Traffic Flow." *J. Transp. Eng.*, ASCE, 131(7), 544-551.

Banks, J. and Carson, J. S. (2000)."Discrete-Event System Simulation." *Pearson Education, Singapore.*

Chandra, S. and Sikdar, P. K. (2000). "Factors affecting PCU in mixed traffic conditions in India". *J. Transp. Eng.*, ASCE, 129(3), 155-160.

Elefteriadou, L., Torbic, D., and Webster, N. (1997). "Development of passenger car equivalents for freeways, two-lane highways, and arterials." *Transp. Res. Rec.*, 1572, TRB, Washington, D.C., 51–58.

Khan, S. I., and Maini, P. (2000)."Modeling heterogeneous traffic flow." *Transp. Res. Rec.*,1678, TRB, Washington, D.C., 234–241.

Kumar, V. M. and Rao, S. K. (1996)."Simulation modeling of traffic operations on two lane highways." *Hwy. Res. Bulletin*, 54, Indian Roads Congress, New Delhi, India, 211-236.

Law, A. M., and Kelton, W. D. (2000)."Simulation modelling and analysis", Mc-*Graw Hill Higher Education, Singapore.*

Marwah, B. R., and Singh, B. (2000)."Level of service classification for urban heterogeneous traffic: A case study of Kanpur metropolis." *Transp. Res. Circular E-C018: Proc., 4th Int. Symp. on Hwy. Capacity*, Maui, Hawaii, 271–286.

Nagraj, B. N., George, K. J., and John, P. K. (1990)."A study of linear and lateral placement of vehicles in mixed 0traffic environment through video-recording." *Hwy. Res. Bulletin*, 42, Indian Roads Congress, New Delhi, India, 105–136.

Ramanayya, T. V. (1988)."Highway capacity under mixed traffic conditions." *Traffic Eng. Control*, 29(5), 284–287.

Analysis of Regional Supply Chain Economic and Environmental Effects of Expansion of the U.S. Freight-Rail System

Chris Hendrickson[1], H. Scott Matthews[2], and Gyorgyi Cicas[3]

[1]Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA 15213-3890, Phone: 412-268-2941, Fax: 412-268-7813, Email: cth@cmu.edu
[2]Departments of Civil and Environmental Engineering and of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA 15213-3890, Phone: 412-268-6218, Fax: 412-268-7813, Email: hsm@cmu.edu
[3]Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA 15213-3890, Phone: 412-268-7889, Fax: 412-268-7889, Email: gcicas@andrew.cmu.edu

Abstract

In this paper, we analyze the regional supply chain economic and environmental effects of shifting 10% of intercity freight carried by trucks to rail. Compared to truck, the freight-rail network provides a cost competitive, more fuel-efficient shipping alternative that can also help to reduce roadway congestion. At the same time, the increased demand for rail transportation raises the need for the expansion of the railroad system. In our case study, we assume that 10 percent of intercity freight moved by trucks in Pennsylvania is shifted to rail, and that this shift requires a 10 percent expansion of the freight-rail system: construction of new tracks, stations, maintenance and repair shops, and manufacturing of new locomotives. Using the national and regional 1997 benchmark economic input-output life-cycle assessment models (EIO-LCA and REIO-LCA) developed at Carnegie Mellon University (www.eiolca.net), we estimate the change of total supply chain economic activity, electricity and fuel use, emission of carbon dioxide and conventional air pollutants (CO, SO_2, NO_x, VOC, PM_{10}) induced by the shift between rail and truck transportation in Pennsylvania and the Mid Atlantic economic region (including DC, DE, MD, NJ, NY, and PA). We find that the shift has environmental benefits considering only the transportation sectors and their supply chains themselves. If the railroad infrastructure investment effects are allocated over a period of time, the overall supply chain effects are also positive.

Introduction

Trucking and rail are modal competitors for freight shipments, as even truck trailers may be transported for part of their trip via rail. In this paper, we analyze the regional supply chain economic and environmental effects of shifting 10 percent of intercity freight carried by trucks to rail. The total volume of U.S. domestic freight has grown 20 percent since 1991, adding to congestion and increased fuel consumption (DoT 2004). Rail accounted for 37 percent and truck transportation 29 percent, of total domestic ton-miles in 2001 (DoT 2004). In 2001, the transportation sector accounted for the majority of national emissions of carbon monoxide (CO) and nitrogen oxide (NO_x) at 83 percent and 55 percent, respectively (DoT 2004). Trucks induce more CO and NO_x emission per ton-mile than trains (DoT 2004 and AAR 2004). The freight-rail network provides a cost competitive, more fuel-efficient shipping alternative that can help also to reduce roadway congestion. At the same time, the increased demand for rail transportation raises the need for the expansion of railroad system.

In this case study, 10 percent of intercity freight moved by trucks is assumed to be shifted to rail and it requires up to a 53 percent expansion of the freight-rail system: construction of new tracks, stations, and manufacturing of new locomotives. Using the regional 1997 benchmark economic input-output life-cycle assessment model (REIO-LCA) developed at Carnegie Mellon University (Hendrickson 2006), our objective is to estimate the change of total supply chain economic activity, electricity and total energy consumption, emission of carbon dioxide and conventional air pollutants (CO, SO_2, NO_x, VOC, and PM_{10}) induced by the shift between rail and truck transportation in Pennsylvania. We will estimate emissions and resource uses for the Mid Atlantic economic region (including District of Columbia, Delaware, Maryland, New Jersey, New York, and Pennsylvania) (BEA 2005). Note that the entire national and international supply chain effects would be larger than the regional effects estimated here.

In our base case year 2000, trucks carried 10,700 million tons, and 2,010 million tons of goods were moved by rail (AASHTO 2005). Therefore, in our hypothetical freight shift case (10% shift scenario) 3,080 million tons will be carried by rail, which implies a 53 percent increase of final demand for the rail transportation sector. The 2000 per-ton costs of freight rail and truck transportation were $0.024 and $0.080, respectively (AASHTO 2005). Thus, the increased demand for rail transportation is $74 million, and the decreased annual cost of trucking is $770 million. Pennsylvania had 5,085 miles of rail in 2003 (AAR 2003), thus we assume that 2,695 miles of new tracks (at a cost of $ 2.5M per mile) and two new stations will be constructed (at a cost of $ 50 M station), and 14 locomotives (at a cost of $ 2 M per locomotive) will be purchased to provide appropriate freight rail transportation services (Hendrickson 2006). Table 1 presents the input-output sectors used in the analysis:

Table 1: Economic input-output sectors and final demands used for the 10% freight shift case

Sector name	Note	IO 1997 number	Base case final demand, M$	Final demand in freight shift case, M$
Manufacturing and industrial buildings	construction of 2 new stations	230210	0	100
Other new construction	track construction	230250	0	6,700
Railroad rolling stock manufacturing	manufacturing of 14 locomotives	336500	0	28
Truck transportation		484000	856	770
Rail transportation		482000	48	74

The total supply chain economic effect representing all economic transactions significantly changes due to the construction of tracks and manufacturing locomotives. The shift from truck transportation to rail reduces economic transactions by $ 300M for trucking and its supply chain, while increasing transactions by $ 60 M for rail and its supply chain. The economy-wide difference between the scenarios is approximately 12% without considering the investment to construction and purchase of locomotives. The per-ton cost of shipping freight by truck is larger by a factor of three than transporting it by rail. Table 2 shows the comparison of economic effects estimated for the base and analysis cases.

The infrastructure investments for additional rail capacity are reported in Table 2 and require considerably more than the annual savings of the freight switch, an amount equal to $ 12,000 M. However, the services of these investments would be available over a long period of time. Also, to the extent that excess capacity exists in the rail sector, these costs might be avoided. Thus, this is a worst case bound on the actual cost.

Table 2: Total supply chain (direct and indirect) economic effects for the base and freight shift cases ($M)

Sector	Base Case	10% Shift Scenario	Shift – Base Difference
Truck Transportation	2,100	1,800	(300)
Rail Transportation	100	160	60
Rail Facility Construction + Manufacturing	0	12,000	12,000

Air Emissions

Table 2 shows our estimates of the air emissions associated with the base 2000 case and for the hypothetical shift of 10% freight from truck to rail. The changes in air

emissions are dominated by the reductions in emissions from truck transportation for carbon monoxide (CO), nitrogen oxides (NOx), and volatile organics (VOC) emissions. For sulphur dioxide (SO2) and particulates (PM10), the rail construction is the predominant effect. As noted above, these construction and manufacturing emissions are due to an initial investment, whereas the truck and rail emissions occur annually.

Table 3: Comparison of total supply chain air emissions for the base and 10% freight shift cases

Sector	CO, mt	NOx, mt	SO2, mt	PM10, mt	VOC, mt
Truck, Base Case	120,000	9,700	1,200	350	9,400
Truck, Shift Scenario	110,000	8,700	1,000	320	8,500
Truck, Shift – Base Difference	(10,000)	(1,000)	(200)	(30)	(900)
Rail, Base Case	640	2,600	500	6	82
Rail, Shift Scenario	990	3,900	760	9	130
Rail, Shift – Base Difference	350	1,300	260	3	52
Construction + Mfg, Shift Scenario	18,000	4,800	5,800	820	1,900

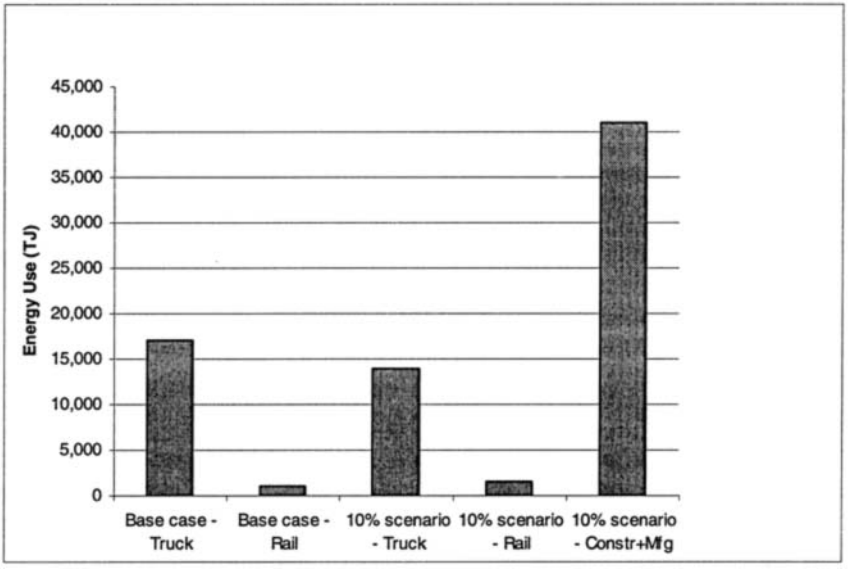


Figure 3: Comparison of total supply chain energy use for the base and 10% freight shift scenarios

Energy Use

Figure 3 shows our estimates of total energy consumption for the various sectors of transportation, construction and manufacturing. Total energy includes fuel use and electricity for the entire supply chain of the different sectors. Regarding the two transportation sectors alone, the switch from truck to rail reduces overall energy use in the economy by roughly 2,400 terajoules per year. However, the railroad infrastructure expansion requires twenty times more than these savings. However, the railroad track, stations and locomotives could very likely be in use for longer than twenty years.

Greenhouse Gas Emissions

Figure 4 shows our estimates of greenhouse gas emissions associated with the base case transportation and the 10% freight shift scenario. In the figure, we report only carbon dioxide emissions, as other greenhouse gas emissions are very small for these sectors. Not surprising, the results are similar to those of energy and air emissions. Within the transportation sectors themselves, the shift reduces greenhouse gas emissions, but the effect of infrastructure investment is significant.

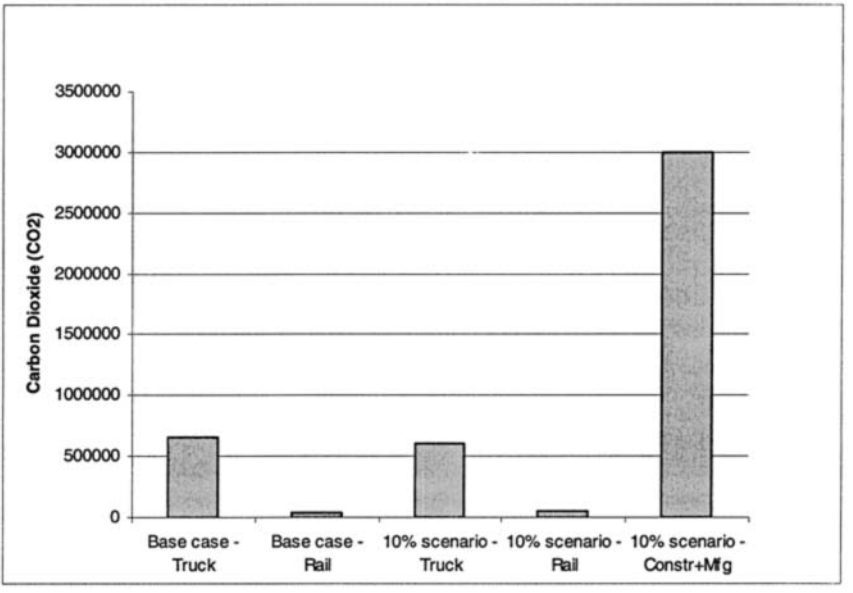


Figure 4: Comparison of supply chain greenhouse emissions

CONCLUSIONS

As might be expected, rail tends to have lower environmental impacts per unit of transport than does truck transportation. A shift of freight from truck to rail reduces the environmental impacts. However, the environmental effects of infrastructure investments to accommodate such shifts are substantial if capacity expansion is required.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant Number: DMII - 0328071. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Association of American Railroads, "RR Industry Info – Railroads and States", (2005) http://www.aar.org/PubCommon/Documents/AboutTheIndustry/RRState_Rankings.pdf, accessed August 14, 2005

Association of American Railroads, "Railroads Announce Surge in Intermodal and Carload Freight", (2005) http://www.tomorrowsrailroads.org/media/display_release.cfm?ID=249, accessed August 14, 2005

American Association of State Highway and Transportation Officials, (2005) "Transportation Invest in America – Freight-Rail Bottom Line Report", http://freight.transportation.org/doc/FreightRailReport.pdf, accessed August 14, 2005

U.S. Department of Commerce, Bureau of Economic Analysis, (2005) "BEA Regions", http://www.bea.gov/bea/regional/gsp/help/OnlineHelp.htm, accessed August 14, 2005

U.S. Department of Transportation, Bureau of Transportation Statistics, (2004) "Transportation Statistics Annual Report", http://www.bts.gov/publications/transportation_statistics_annual_report/2004/, accessed August 14, 2005

Hendrickson, C. T., L. B. Lave, H. S. Matthews, A. Horvath, S. Joshi, F. C. McMichael, H. MacLean, G. Cicas, D. Matthews and J. Bergerson, (2006) "*Environmental Life Cycle Assessment of Goods and Services: An Input-Output Approach*"; Resources for the Future: Washington, DC.

Assessing Benefits of Implementing Integrated Corridor Management Strategies

H. Tanikella[1], B. Park[2], B. L. Smith[3] and M. G. Best[4]

[1] Graduate Research Assistant, Department of Civil Engineering, University of Virginia, P.O. Box 400472, Charlottesville, VA 22903-4742; PH (434) 924-1420; FAX (434) 982-2591; email: ht5t@virginia.edu
[2] Assistant Professor, Department of Civil Engineering, University of Virginia, P.O. Box 400472, Charlottesville, VA 22903-4742; PH (434) 924-6347; FAX (434) 982-2591; email: bpark@virginia.edu
[3] Associate Professor, Department of Civil Engineering, University of Virginia, P.O. Box 400472, Charlottesville, VA 22903-4742; PH (434) 243-8585; FAX (434) 982-2591; email: bls2z@virginia.edu
[4] Graduate Research Assistant, Department of Civil Engineering, University of Virginia, P.O. Box 400472, Charlottesville, VA 22903-4742; PH (434) 924-1420; FAX (434) 982-2591; email: mgb3e@virginia.edu

Abstract

The concept of integrated corridor management (ICM) has received a lot of interest recently because of the perceived potential benefits it can offer. Several regional agencies have initiated the development of ICM plans to facilitate vehicular movements more effectively within the concerned corridors. However, very few of such studies have focused on quantifying the benefits associated with implementation of ICM strategies. This study, therefore, aimed at using traffic simulation as a tool for evaluating ICM strategies to obtain a preliminary understanding of the range of benefits that ICM can offer. Two levels of ICM strategies--low and high--were evaluated against a base case for four different traffic conditions, including normal, incident, special event, and inclement weather for peak traffic volumes. Each level of ICM strategy was compared to the base case using performance indicators such as network performance and person-hours of travel time. Based on the assumptions, the results indicated reductions in travel times in the range of 20 percent from base to low ICM and 27 percent from base to high ICM.

Introduction

Integrated Corridor Management (ICM) refers to the practice of managing corridors as a whole—instead of independently—and providing local solutions (SAIC Team (2005)). ICM is being considered by the Federal Highway Administration (FHWA) as one of its nine major focus areas for improving mobility and safety in surface transportation. FHWA seeks to make an informed decision on the necessity and degree of deployment of ICM strategies by understanding the magnitude of benefits that it can offer. To date, multiple strategies categorized under ICM initiatives have not been evaluated quantitatively to assess their relative benefits under different traffic conditions. To ascertain the exact nature of these benefits, it is necessary to evaluate different levels of ICM strategies under various network traffic conditions. This study, therefore, used a microscopic simulation tool, VISSIM, to model a generic corridor and evaluate several different traffic conditions to provide a preliminary assessment of performance improvements. Two different levels of ICM strategies were evaluated against a base case for four different network traffic conditions (normal, incident, special event, and inclement

weather) for peak volume traffic conditions. The following sections describe the evaluation procedure and the resulting conclusions on ICM benefits.

Network Definition

The first stage in the evaluation of ICM strategies was the selection of a generic corridor. The corridor was so chosen because it contained freeways, arterials, and bus and rail transit lines. Figure 1 shows a snapshot of the network selected for this study along with the freeways and arterials of interest. All the freeways and arterials had bus lines running in both directions. Additionally, both freeways had four lanes in each direction with two additional HOV lanes on Freeway 2 SB. After some preliminary evaluation, only two of the arterials (Arterial 1 and Arterial 2) and the two freeways were chosen for detailed evaluation. The subway was considered in the preliminary analysis, but as it provided consistent travel times with little variability, it wasn't considered for evaluation in the subsequent phase.

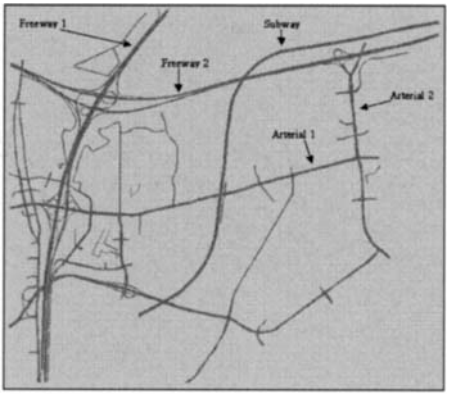


Figure 1. Hypothetical simulation network map.

Network Traffic Conditions & ICM Strategies

Table 1 shows the simulation parameters for each traffic condition. These assumptions were based on some studies in literature (Hanbali (1994), Perrin (2001), Sadek (2004)). Since this was a preliminary evaluation, no effort was made to test the accuracy of these assumptions.

Evaluations

The network for each of the twelve traffic conditions was coded into VISSIM, and the evaluations were conducted using plausible data. Each simulation experiment was conducted for two hours of simulation time with ten repetitions for each condition. The network warm-up time was thirty minutes for each simulation run.

Three key performance measures were chosen to compare the performance of the test network for different levels of ICM strategies. The key indicators chosen for evaluation were network travel time, delay, and individual travel time for freeways and key arterials in person-hours. While network travel time and delay measures are based on all modes and all types of vehicles, it was determined that

in a multi-modal corridor, person-hours of travel time on key routes would provide a better understanding of the performance of the corridor. Therefore, this was chosen as one of the performance indicators. The results of the simulation runs are discussed in the following section based on the selected performance metrics.

Table 1. Traffic Condition Parameters

ICM Level		normal	incident	special event	weather
BASE	-volumes	peak volumes	peak volumes	peak volumes + 3000 veh, 200 buses on freeways; 2000 veh on arterials	peak volumes
	-buses	20 min.	20 min.	20 min. w/add. 20 buses in first 30 min.	20 min.
	-trains	10 min.	10 min.	10 min. w/add. 6 trains in first 30 min.	10 min.
	-other	signal retiming - limited coordination	70% cap. reduction for 30 min.		reduce speeds by 15% and 30%
LOW	-volumes	peak volumes w/10% HOV diversion	peak volumes w/10% HOV diversion	peak volumes w/10% HOV diversion + 3000 veh, 200 buses on freeway; 2000 on arterials	peak volumes w/10% HOV diversion
	-buses	20 min.	20 min.	20 min. w/add. 20 buses in first 30 min.	20 min.
	-trains	10 min.	10 min.	10 min. w/add. 6 trains in first 30 min.	10 min.
	-other	signal retiming - fully coordinated; HOV-2	incident signal retiming; HOV-2; 70% cap. reduction for 30 min.	special event signal retiming; HOV-2	inclement weather signal retiming; HOV-2; reduce speeds by 15% and 30%
HIGH	-volumes	peak volumes w/15% HOV, 10% transit *diversions*	peak volumes w/15% HOV, 10% transit *diversions*	peak volumes w/15% HOV, 10% transit diversions + 3000 veh, 200 *buses on freeway; 2000 on arterials*	peak volumes w/15% HOV, 10% transit *diversions*
	-buses	15 min.	15 min.	15 min. w/add. 20 buses in first 30 min.	15 min.
	-trains	7 min.	7 min.	7 min. w/add. 6 trains in first 30 min.	7 min.
	-other	signal retiming - fully coordinated; HOV-3; ramp metering	incident signal retiming; HOV-3; 70% cap. reduction for 30 min.	special event signal retiming; ramp metering; HOV-3	inclement weather signal retiming; HOV-3; ramp metering; reduce speeds by 15% and 30%

Results

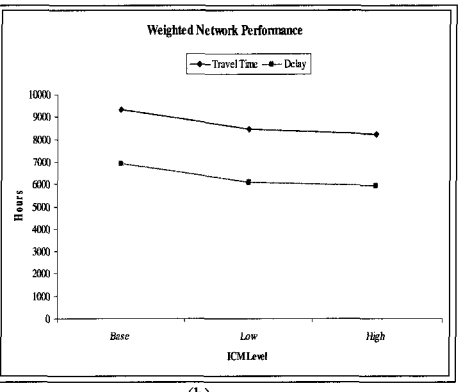
Network Travel Time and Delay. Over the four traffic conditions, network travel time improved by averages of 5.5 percent and 8.7 percent between base and low ICM and base and high ICM levels, respectively (Figure 2(a)). Further, total network delay decreased by averages of 6.6 percent and 10.6 percent between base and low ICM and base and high ICM levels, respectively. Most notably, implementing low ICM in a normal traffic condition reduced average network travel time and delay by 10.5 percent and 14.0 percent, respectively. Figure 2(b) shows the network travel time and delay obtained by weighting with percentage occurrence of each condition (88 percent normal, 10 percent incident, and 2 percent special event). Thus, the overall network performance improvement was 10 percent from base to low ICM and 13 percent from base to high ICM.

Person-Hours of Travel Time. Tables 2 and 3 show the person-hours of travel time experienced on all modes on key corridors. Travel times for each of the traffic conditions were calculated by weighting the individual mode travel times by the occupancy of the mode for key routes. Using these values, the total travel time in person-hours for key corridors in the network was computed for each traffic condition and ICM level. In order to obtain a single trend for the entire network, the total person-hour travel times were weighted by traffic condition occurrence (88 percent normal, 10 percent

incident, and 2 percent special event).

Scenario	Travel Time(Hours)		Delay(Hours)	
Normal	Mean	Std.Dev	Mean	Std. Dev
Base	9286	167	6891	197
Low	8311	120	5928	140
High	8163	127	5835	140
Incident				
Base	9519	312	7198	308
Low	9441	253	7217	286
High	8725	356	6430	387
Special Event				
Base	9685	268	7328	307
Low	9172	166	6754	201
High	9065	163	6703	186
Bad Weather				
Base	9914	183	6765	186
Low	9389	176	6445	221
High	9109	118	6244	134

(a)



(b)

Figure 2. Network performance results for traffic and ICM conditions.

Table 2. Person-Hours of Travel Time on Main Arterials Weighted Based on Mode Occupancy.

Normal	Arterial 1 (EB)	Arterial 1 (WB)	Arterial 2 (NB)	Arterial 2 (SB)
Base	317	534	65	320
Low	264	211	46	326
High	244	224	52	309
Incident				
Base	290	555	49	299
Low	373	394	43	354
High	255	219	56	306
Inclement Weather				
Base	264	340	59	325
Low	116	484	51	344
High	133	894	69	333
Special Event				
Base	285	579	123	597
Low	217	345	73	307
High	297	304	105	313

The results shown in Tables 2 and 3 indicate that, during the normal traffic condition, travel times generally decreased except along Freeway 2 EB where the travel times increased from base to low ICM and from base to high ICM. Thus, individual corridors experienced increases in travel time despite a network-wide reduction in travel time. In terms of percentage change, Arterial 1 WB showed the highest decrease in travel time from base to low ICM with 60 percent and a 58 percent decrease from base to high ICM. However, the travel times on Freeway 2 EB increased by 27 percent from base to low ICM and 32 percent from base to high ICM.

In the incident traffic condition, for several corridors, travel times from base to low ICM actually increased while travel times from base to high ICM remained the same or decreased marginally. In addition, in some cases, freeway performance improved, but the arterial performance declined. In terms of percentages for individual corridors, the maximum decrease in travel time was 29 percent for Arterial 1 WB from base to low ICM and 79 percent for Freeway 1 SB from base to high ICM.

Table 3. Person-Hours of Travel Time on Freeways Weighted Based on Mode Occupancy

Normal	Freeway 1 (NB)	Freeway 1 (SB)	Freeway 1 (SB-HOV)	Freeway 2 (EB)	Freeway 2 (WB)
Base	78	273	145	89	75
Low	56	207	141	114	47
High	58	189	133	118	49
Incident					
Base	63	283	122	322	50
Low	50	272	122	399	53
High	56	158	130	267	47
Inclement Weather					
Base	74	245	405	550	81
Low	72	150	201	521	81
High	69	147	188	561	77
Special Event					
Base	263	290	181	93	117
Low	167	150	139	129	66
High	160	247	130	100	77

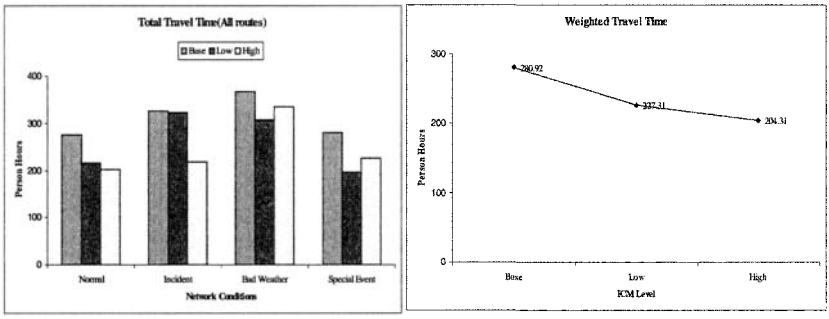
In the special event traffic condition, most of the corridors showed decreases in travel times from both base to low ICM and base to high ICM. Thus, the ICM strategies were very effective in reducing travel times in the corridor. In terms of percentages, the highest decreases in travel time occurred on Arterial 2 SB and Freeway 1 SB of 43 percent from base to low ICM and on Freeway 1 NB of 64 percent from base to high ICM. The maximum increase in travel time occurred on Freeway 2 EB of 38 percent from base to low ICM and 7 percent from base to high ICM.

In the inclement weather traffic condition, the travel times increased from base to low ICM and from base to high ICM for Arterial 1 WB and Arterial 2 SB while there was a small decrease in the freeway travel times from base to low ICM and from base to high ICM. In terms of percentage changes on corridors, the greatest reduction in travel time occurred on Arterial 1 EB with 55 percent from base to low ICM and 115 percent from base to high ICM on Freeway 1 SB (HOV). However, there was an increase in travel time on Arterial 1 WB of 42 percent from base to low ICM and 163 percent from base to high ICM.

Figure 3(a) shows the total person-hours of travel time along key corridors by traffic condition. It can be seen that all the traffic conditions registered a net decrease from base to low ICM. From base to high ICM, all the traffic conditions showed a decrease, but in some cases, the decrease is less than from base to low ICM. Figure 3(b) shows the weighted performance considering the proportion of occurrence of each traffic condition. The percentage decrease from base to low ICM was 19 percent and from base to high ICM was 27 percent.

Discussion

An analysis of the above results shows that ICM strategies can be used to reduce travel times and delay on a corridor network. However, some important issues arise out of this analysis—the selection of the specific ICM strategies, choosing the optimal strategies for corridor management, selection of suitable performance measures, and a more in-depth consideration of the reasons for differences in performance. The results of this analysis demonstrate reduced benefits for the selected traffic conditions based on three performance measures. A more in-depth study would include an evaluation of strategy-specific improvements in performance, e.g., improvements in corridor performance due to signal re-timing or ramp metering. In addition, the selection of suitable performance measures is also another important aspect for evaluation of ICM strategies.



(a) (b)

Figure 3: (a) Total network travel time in person-hours for all modes weighted by occupancy; (b) Total network travel time in person-hours for all traffic conditions weighted by percent of occurrence.

Conclusions

The preliminary analysis of a few ICM strategies for different traffic conditions based on certain assumptions indicates an improvement in the overall performance of the hypothetical corridor. The results indicate that under the assumptions, normal and special event traffic conditions showed considerable benefit while the incident and the inclement weather traffic conditions did not. The overall range of benefits for the entire network was of the order of 20 percent from base to low ICM and 27 percent from base to high ICM. This study demonstrates one particular case of selection of ICM strategies and presents a method of quantitative evaluation of benefits at various levels using simulation. The methodology employed in this study can be used for a more elaborate analysis of the suitability of different ICM strategies. Some important future research directions are evaluation of individual ICM strategies like ramp metering and signal control coordination, selection and formulation of suitable metrics for performance evaluation of multi-modal corridors and a sensitivity analysis of input parameters that are used for defining ICM levels and traffic conditions.

References

Hanbali, R.M. (1994). "Economic Impact of Winter Road Maintenance on Road Users" *Transportation Research Record 1442* TRB, Washington, D.C., pp. 151-161.

Perrin, J., Martin, P.T., & Hansen, B.G. (2001). "Modifying Signal Timing during Inclement Weather" *Transportation Research Record 1748*, TRB, Washington, D.C.

Sadek, A.W., & Amison-Agbolosu, S.J. (2004) "Validating Traffic Simulation Models to Inclement Weather Travel Traffic conditions with Applications to Arterial Coordinated Signal Systems" *Prepared for the New England Transportation Consortium.* http://docs.trb.org/00983592.pdf

SAIC Team (2005) "Integrated Corridor Management-Concept Development and Foundational Research", http://www.itsa.org

Field Evaluation of the Effect of Speed Monitoring Displays on Speed Compliance in School Zones

Kelly G. Ash[1] and Mitsuru Saito[2]

[1] Transportation Engineer (EIT), PBS&J, 7310 North 16th St., Suite 310, Phoenix, Arizona 85020, USA, Phone: 1-602-943-1003, Fax: 1-602-943-1303, e-mail: kgash@pbsj.com.
[2] Professor, PhD, PE, Civil and Environmental Engineering Department, Brigham Young University, 368 Clyde Building, Provo, Utah 84602, USA, Phone: 1-801-422-6326, Fax: 1-801-422-0159, e-mail: msaito@byu.edu.

Abstract

Speed monitoring displays (SMDs) were installed in four reduced speed school zones in Utah in order to evaluate the effectiveness of these dynamic signs at improving speed compliance. Despite the data loss due to the malfunctions of the SMDs and speed data collection equipment, substantial speed data were collected, analyzed, and compared from before and after the SMDs were installed. In some cases the SMDs maintained their effectiveness at increasing speed compliance; in other cases they lost some of their effectiveness possibly due to higher percentages of commuter traffic. For the most part, these SMDs helped improve school zone safety by decreasing speeds and increasing speed compliance as manifested by the decrease in mean speed, standard deviation, 10 mph pace range and the percentage of vehicles exceeding the 20 mph school-zone speed limit.

Introduction

Reduced-speed school zones greatly improve the safety of young children commuting to and from school and provide more appropriate gaps in traffic for children to cross the street. The safety and efficiency of a school zone depends on drivers' compliance to the school-zone speed limit. The main goal of this field study was to evaluate the effects of speed monitoring displays (SMDs) on speed compliance by drivers in school zones in the State of Utah. These SMDs are pole-mounted, fixed SMDs unlike the temporary, portable ones that police departments often use for speed enforcements on residential streets. The field study was conducted between September 2004 and early April 2005.

A public opinion survey conducted in the study (762 participants) concluded that Utah drivers feel there is a need to improve school zones in the State of Utah (Saito and Ash 2005). Since the main reason for speeding in school zones among Utah drivers was because they did not notice the school zone, increasing the visibility of school zones with the use of more noticeable traffic controls can improve speed-limit compliance. UDOT desired to test the use of SMDs as a new addition to the current measures to make school crossings more visible.

Description of Speed Monitoring Display Used for the Study

The SMDs used in this experiment were all pole-mounted and displayed the speed of vehicles in the speed signs. Figure 1 shows a pole-mounted SMD just after installation. The covered area contains a speed sign consisting of the legend "Your Speed" and the variable speed display area. The dimensions of the speed signs are 36 in. by 48 in., and the

dimensions of the variable speed display area are 24 in. by 30 in. The speed of vehicles is displayed with fluorescent yellow-green (same color as school zone signs) sliding disks that slide in and out depending on the number. The signs were programmed to function only during the school zone times. They were also programmed so that small LED lights in the numbers would flash to attract drivers' attention if they were speeding more than 5 mph over the school-zone speed limit (25 mph or more). The signs with solar power were used at all the locations in the study except for one site. Unfortunately, most of the solar-powered signs experienced difficulties from time to time due to not having sufficient power to function properly; however, sufficient data were collected to make a comparison. The SMDs were all installed on the shoulder of the road between the crosswalk and the school-zone speed-limit sign (measuring from the crosswalk, approximately 60 percent of the distance from the crosswalk to the school-zone speed-limit sign).

Figure 1: Pole-mounted, SMD just after installation

Study Sites

Four sites were selected for the field data collection: 400 North (SR-89) and 400 East in Logan, Utah, State St. (SR-89) and 1110 South in Salt Lake City, Utah, 100 East (SR-146) at 1800 North in Pleasant Grove, Utah, and US-6 at Reference Post (RP) 153.8 in Goshen, Utah. Logan and Pleasant Grove are medium-sized cities, Salt Lake City is a large urban area, and Goshen is a small rural town.

The Logan site serves children crossing 400 North to walk to and from Adams Elementary School which is located one block north of the school zone (see Figure 2). The roadway (400 North) has four-lanes and a two-way left-turn lane (TWLTL) with a downward slope for westbound vehicles. The road also has sufficient shoulder widths to allow cars to park along both sides of the street. Even though the street is a very busy road, houses still align its shoulders. The approach speed limit to the school zone is 35 mph.

The Salt Lake City site serves children crossing State Street to travel to and from Lincoln Elementary School, which is located one block east of the school zone (see Figure 3). In this area, State Street receives heavy volumes of traffic and consists of six lanes of traffic with a raised median on a flat grade. The road also has shoulder widths large enough for cars to park along both sides of the street. The street is aligned with both residential and commercial land uses. The approach speed limit to the school zone is 35 mph.

The Pleasant Grove site serves a very large number of children crossing 100 East (SR-146) at 1800 North to commute to and from Manila Elementary School (see Figure 4). SR-146 is a two-lane highway with a TWLTL on the south side of the intersection and no median on the north side of the intersection. The road has a slight downhill grade for southbound traffic. On the south side of the intersection, SR-146 is lined with sidewalks on both sides of the street. However, the north side of the intersection has unpaved shoulders. The approach speed limit to the school zone is 35 mph.

The Goshen site allows students to cross SR-6 to walk to and from the Goshen Elementary/Middle School. Going through Goshen, SR-6 is a two-lane level highway with a painted TWLTL and extremely wide shoulders (see Figure 5). Due to the width of the roadway, there is a significant distance that children must travel to cross the street. Along the

street, there are a few homes and small businesses surrounding the school. The approach speed limit to the school zone is 35 mph.

These sites are equipped with typical school-zone traffic control devices. Both approaches have a School Advance Warning Assembly, which consists of signs S1-1 and W16-9p of the Manual of Uniform Traffic Control Devices (MUTCD) (FHWA 2003). Following the School Advance Warning Assembly on both sides of the street, there is a School Speed Limit Sign with two flashing beacons above it. This sign is S5-1 of the MUTCD. These signs state that the speed limit is 20 mph when flashing and advise drivers that children may be crossing the street. At the crosswalk, for both directions of traffic there is a School Crosswalk Warning Assembly consisting of signs S1-1 and W16-7 of the MUTCD. The school zone ends shortly after the crosswalk with a End School Zone sign (sign S5-2 of the MUTCD).

Figure 2: Westbound Approach to the School Zone (SR-89) in Logan

Figure 3: Northbound Approach to the School Zone (SR-89) in Salt Lake City

Figure 4: Southbound Approach to the School Zone (SR-146) in Pleasant Grove

Figure 5: Eastbound Approach to the School Zone (US-6) in Goshen

Speed Study Results

Because of the space limitation we use only the morning data to discuss the SMDs' long-term effects. The reader is encouraged to infer the effectiveness of SMDs in reduced-speed school zones from the data presented in the tables. See Saito and Ash (2005) for the afternoon data and the data on short-term effects. Data collection dates are found at the top of the tables.

SR-89 (400 North) at 400 East in Logan. Table 1 presents speed data from before and after the SMDs were installed at this site. For both directions of traffic the SMDs were effective at reducing the mean speed, standard deviation, 85th percentile speed, and speed-limit compliance. As seen in the table, the signs had a significant impact on drivers traveling westbound, which is a downhill grade. The mean speed decreased by about 3 mph and the 85th percentile speed dropped by about 4 mph. The standard deviation decreased in both cases, suggesting a tighter distribution of speeds (i.e., resulted in less outliers). The percent

of vehicles exceeding the 20 mph speed limit was reduced by about 40 percent and the 10 mph pace decreased by about 3 mph for westbound traffic. The percentage of vehicles in the pace increased as well; which also suggests an increase in compliance. The SMDs also had an impact on the vehicles traveling eastbound; however, due to the uphill grade and already compliant speeds, the changes were not quite as significant as they were for the westbound traffic. The distribution of speeds for the westbound traffic suggested that all of the vehicles were influenced by the SMDs; whereas of the vehicles traveling eastbound, the faster vehicles were more greatly impacted.

Table 1. Speed Results for Westbound SR-89 (400 North) at 400 East in Logan, Utah

Statistics	Before *(09/13/04 – 09/16/04)*	After *(03/28/05 – 03/31/05)*
Westbound Morning – 7:30 to 8:30 AM		
Mean (mph)	23.24	19.68[a]
Standard Deviation	4.59	3.30
85th Percentile (mph)	26.2	22.1
% Exceeding 20 mph	75.9%	35.7%
10 mph Pace (% in Pace)	17 – 27 (85.9%)	14 – 24 (91.1%)
Sample Size	809	1001
Eastbound Morning – 7:30 to 8:30 AM		
Mean (mph)	19.86	18.70[a]
Standard Deviation	4.88	3.50
85th Percentile (mph)	23.3	20.7
% Exceeding 20 mph	32.2%	18.1%
10 mph Pace (% in Pace)	13 – 23 (84.3%)	13 – 23 (91.9%)
Sample Size	699	717

[a] *Difference was statistically significant from the "before" mean speed based on a normal approximation test at a 95 percent confidence level*

SR-89 (State Street) at 1110 South in Salt Lake City. Table 2 presents speed data from before and after the SMDs were installed at this site. Due to the location's urban downtown setting, speeds were already very compliant to the school-zone speed limit despite the functional classification of the roadway. Vehicles tended to arrive at the school zone in platoons; therefore drivers were less likely to speed due to the constraints of other vehicles around them. Since the speed-limit compliance during the "before" condition was already excellent, little difference in speed compliance was observed after the SMDs were installed.

SR-146 (100 East) at 1800 North in Pleasant Grove. Table 3 presents speed data from before and after the SMDs were installed at this site. The results of the analysis seemed to vary for the two directions of traffic, but may have been caused by the dysfunction of the southbound SMD. The results for the northbound morning traffic show that the SMD may have lost effectiveness over time. Perhaps the lack of long-term effectiveness for the morning time period was the result of more commuters who see the sign everyday. This site experienced frequent SMD malfunctions.

US-6 in Goshen (RP 153.8). Table 4 presents speed data from before and after the SMDs were installed at this site. Since the SMDs at the Goshen location were hard wired instead of powered by solar panels, no apparent difficulties were encountered with the SMDs. The results of the eastbound data collection actually showed a slight increase in speeds through the school zone (1 to 2 mph). Reasons for the increase are unknown. Regardless of the

increase in the mean speed, drivers traveling through this school zone were still very compliant to the reduced-speed school-zone speed limit. The results of the westbound analysis in Goshen were somewhat different than the results of the eastbound analysis, showing essentially no change in the mean speed. The SMD may have been more effective if the shoulders of the road had not been so wide. The speed distributions at this location showed that during the morning period, the SMDs had a noticeable influence on excessive speeds through the school zone despite the increase and lack of change in the mean speeds.

Table 2. Speed Results for SR-89 (State Street) at 1110 South in Salt Lake City, Utah

Statistics	Before *(09/13/04 – 09/16/04)*	After *(03/28/05 – 03/31/05)*
Northbound Morning – 7:30 to 8:25 AM		
Mean (mph)	19.06	19.85[a]
Standard Deviation	3.23	3.70
85th Percentile (mph)	21.4	22.4
% Exceeding 20 mph	24.1%	37.0%
10 mph Pace (% in Pace)	14 – 24 (91.9%)	14 – 24 (90.7%)
Sample Size	1223	1069
Southbound Morning – 7:30 to 8:25 AM		
Mean (mph)	19.83	18.94[a]
Standard Deviation	4.36	4.23
85th Percentile (mph)	22.7	21.3
% Exceeding 20 mph	40.4%	22.8%
10 mph Pace (% in Pace)	14 – 24 (86.0%)	13 – 23 (87.7%)
Sample Size	463	838

[a] *Difference was statistically significant from the "before" mean speed based on a normal approximation test at a 95 percent confidence level*

Table 3. Speed Results for SR-146 (100 East) at 1800 North in Pleasant Grove, Utah

Statistics	Before *(10/04/04 – 10/07/04)*	After *(03/21/05 – 03/24/05)*
Northbound Morning – 8:15 to 8:50 AM		
Mean (mph)	19.88	19.80
Standard Deviation	4.44	3.78
85th Percentile (mph)	24.2	23.3
% Exceeding 20 mph	40.7%	34.0%
10 mph Pace (% in Pace)	14 – 24 (77.4%)	15 – 25 (86.6%)
Sample Size	513	777
Southbound Morning – 8:15 to 8:50 AM		
Mean (mph)	21.58	22.69[a]
Standard Deviation	4.92	4.09
85th Percentile (mph)	25.7	26.1
% Exceeding 20 mph	60.6%	76.0%
10 mph Pace (% in Pace)	16 – 26 (75.2%)	17 – 27 (83.0%)
Sample Size	926	317

[a] *Difference was statistically significant from the "before" mean speed based on a normal approximation test at a 95 percent confidence level*

Table 4. Speed Results for US-6 in Goshen, Utah (RP 153.8)

Statistics	Before *(09/27/04 – 09/30/04)*	After *(03/21/05 – 03/24/05)*
Eastbound Morning – 8:10 to 9:00 AM		
Mean (mph)	19.75	21.66[a]
Standard Deviation	4.69	3.18
85th Percentile (mph)	23.4	24.4
% Exceeding 20 mph	38.2%	63.9%
10 mph Pace (% in Pace)	15 – 25 (81.9%)	16 – 26 (88.2%)
Sample Size	144	119
Westbound Morning – 8:10 to 9:00 AM		
Mean (mph)	22.09	21.77
Standard Deviation	6.36	4.56
85th Percentile (mph)	28.3	27.3
% Exceeding 20 mph	55.2%	50.9%
10 mph Pace (% in Pace)	14 – 24 (68.0%)	16 – 26 (76.9%)
Sample Size	125	108

[a] *Difference was statistically significant from the "before" mean speed based on a normal approximation test at a 95 percent confidence level*

Conclusions

The effectiveness of SMDs at increasing speed-limit compliance in four school zones in Utah was analyzed (one in a rural area, two in middle-sized cities, and another in a large city). The SMDs analyzed in this study proved to increase speed compliance in most cases. According to the distribution of speeds at essentially every location, excessive speeds were reduced. For the most part, these SMDs helped improve school-zone safety by decreasing speeds and increasing speed compliance as manifested by the decrease in mean speed, standard deviation, 10 mph pace range and the percentage of vehicles exceeding the 20 mph school-zone speed limit. Malfunctions of the solar-powered SMDs hampered data collection significantly at the beginning of the study and they might have influenced drivers' responses to short-term effects. Also, since the results of the field study differed by location, further research should be conducted to determine if other factors and conditions may contribute to and influence the effectiveness of SMDs.

References

Saito, M. and K. G. Ash. (2005). *Evaluation of Four Recent Traffic Safety Initiatives, Volume IV: Increasing Speed Limit Compliance in Reduced Speed School Zones.* Brigham Young University: for the Utah Department of Transportation. Final Report, Report No. UT-05.13.

FHWA. (2005). *Manual on Uniform Traffic Control Devices (MUTCD).* Federal Highway Administration, U.S. Department of Transportation, 2003 edition.

Development of Group Decision-Support Model and System for Combined Transportation and Utility Construction

Chien-Cheng Chou[1]; Carlos H. Caldas[2], M. ASCE; James T. O'Connor[3], M. ASCE, and Grant K. Goldman[4]

[1]Ph.D. Candidate, Dept. of Civ. Engrg., Univ. of Texas, Austin, TX 78712. E-mail: ccchou@mail.utexas.edu
[2]Assistant Professor, Dept. of Civ. Engrg., Univ. of Texas, Austin, TX 78712. E-mail: caldas@mail.utexas.edu
[3]C.T. Wells Professor, Dept. of Civ. Engrg., Univ. of Texas, Austin, TX 78712. E-mail: jtoconnor@mail.utexas.edu
[4]Graduate Research Assistant, Dept. of Civ. Engrg., Univ. of Texas, Austin, TX 78712. E-mail: ggoldman@mail.utexas.edu

Abstract

As more and more transportation projects are located in highly congested metro settings, many projects require that adjacent utilities be adjusted to make room for new or expanded highway facilities. The adjustment of utilities prior to highway construction is a highly challenging operation from many perspectives. One major strategic approach that has emerged over the last 15 years is for state Departments of Transportation (DOTs) to combine utility adjustment work with the highway contractor's scope of work, thereby eliminating or reducing some of the associated complications and risks. While many benefits can result from this combined approach, it does have its disadvantages and own set of challenges. Hence, in the highway planning and design phase, a decision support model and a Group Decision-Support System (GDSS) were proposed to provide guidance to both state DOTs and utility decision-makers as to when the combined approach could be applied. This paper discusses the design of the decision support model and its system. With the use of GDSS, state DOTs and utility decision-makers can quickly realize the most important problems that need to be addressed and discuss solutions. A scenario using the proposed system on a hypothetical utility adjustment is presented.

Introduction

Modern highway projects often involve adjusting surrounding utilities in order to make room for new or expanded highway facilities. The conventional approach used by most state DOTs and utility owners to implement utility adjustments requires that each involved utility owner adjust its own utilities prior to the highway construction project. In this manuscript, this method is referred to as the "Conventional Approach." As an increasing number of highway projects are located in highly congested metro settings, schedule slippages and increased costs associated with highway construction will be expected if the utilities are not adjusted in a timely manner (GAO 1999). A recent research study also showed that the most frequently cited reason for delays in highway construction is caused by utility adjustments delays due to the difficulty in multi-party coordination (Ellis and Thomas 2003). Adjusting utilities based on the uncompleted highway design may compel utility owners to relocate again because of any slight change of highway alignment. Utility owners are reluctant to begin adjustment work unless the detailed design of highway facilities is finalized and confirmed. Therefore, using the conventional approach may turn more utility adjustment activities into the critical path of the entire highway project, which apparently lengthens the highway project duration.

The CTUC approach. The strategic approach that has emerged over the last 15 years is for state DOTs to combine utility relocation work with the highway contractor's scope of work, thereby

eliminating or reducing some of the previously mentioned complications and risks. This approach is referred to as the "Combined Transportation and Utility Construction (CTUC) Approach." One of the CTUC approach's main benefits is that since both the utility adjustments and highway construction activities are controlled by the highway contractor, the activities requiring similar resources could be scheduled to be performed at the same time, and thus would save a lot of resources-allocating time and efforts which are commonly wasted in the conventional approach (AASHTO 2004). While many benefits can result from the CTUC approach, it does have its disadvantages and own set of challenges. For example, compared to their subcontractors in the conventional approach, utility owners feel less control on the highway contractor and lack confidence in the highway contractor's capability. Since utility owners have to maintain utility service reliability and quality, if the highway contractor has not adjusted the same type of utility facilities before, it is difficult to for state DOTs to convince utility owners that the CTUC approach is the best choice. Given the challenges faced by state DOTs and utility owners coupled with the impact of external, unexpected events, a clear need emerges for a decision support system that encompasses all decision variables driving or impeding implementation of the CTUC approach. In addition, to evaluate the potential benefits and challenges to implementing CTUC on their projects, state DOTs and utility owners considering CTUC need a systematic and transparent method for analysis and decision-making on the applicability of CTUC.

Research Objectives. The objectives of the research presented on this paper are: (1) to identify the decision variables for evaluating the applicability of the CTUC approach; (2) to develop a decision support model to improve the quality of CTUC decision-making; (3) to develop a GDSS to assist project stakeholders in considering possible use of the CTUC approach on a particular project, based on the decision variables identified. To achieve the above objectives, our main research steps are introduced as follows: (1) perform literature review of utility adjustment approaches currently used by state DOTs and utility owners; (2) conduct interviews with state DOTs and utility companies' representatives to identify CTUC decision variables and assess their impact on the CTUC decision; (3) review current decision support systems and technologies; (4) develop the CTUC decision support model and system; (5) test and validate the CTUC decision support model and system.

Literature Review

Past research studies have indicated that in the conventional approach, delays to highway construction projects caused by utility adjustments result in longer completion times and increased costs (Cisneros 1996) (Ellis 1996) (TRB 2004). Other research projects have also documented the CTUC approach as the next logical solution to the problems occurred in the conventional approach (Quinn 1997) (AASHTO 2004) (Vidalis and Najafi 2002). Through the above studies, the problems of the conventional approach and CTUC case studies have been well documented and discussed. However, none of these research studies discuss why utility owners do not want to use the CTUC approach, and the CTUC decision variables are not well addressed in the literature. In other words, most research projects focus on the problems of the conventional approach and implementation details of the CTUC approach. This research attempted to design a GDSS that can incorporate both state DOTs and utility owners' opinions and assist them in making the CTUC decision.

CTUC Decision Support Model

Since DSS technologies are widely used in almost every business domain, providing better CTUC decision recommendations may need to reuse or integrate current DSS technologies. Hence, the

architecture of each major DSS is discussed first so that the appropriate architecture of DSS can be defined and developed further.

In model-driven DSS, decision analysts (i.e. professionals who do not have domain knowledge but have the ability to perform decision analysis) examine data, identify decision variables, and develop a mathematical model to best describe the problem domain. Model-driven DSS can then use the model to perform simulation under varied events in order to aid decision-makers (i.e. professionals who have domain knowledge and are primary users of DSS) in analyzing a situation (Holsapple and Whinston 1996). However, the general model regarding the human decision-making process consists of four stages (Simon 1960) (Forgionne 2000):

(1) Intelligence: Observe reality. Gain problem understanding. Acquire needed information.
(2) Design: Develop decision criteria. Develop decision alternatives. Identify relevant events. Specify the relationships between criteria, alternatives, and events.
(3) Choice: Logically evaluate the decision alternatives. Develop recommended actions that best meet the decision criteria.
(4) Implementation: Develop an implementation plan. Secure needed resources. Put implementation plan into action.

The architecture of model-driven DSS doesn't fit CTUC decision-making requirements because the core task of CTUC decision-making would be the Intelligence stage, which is assumed to be done by decision analysts without any help of model-driven DSS.

In Executive Information System (EIS), a time-series of data exists and will be automatically organized into specified board categories, which are defined by decision analysts and decision-makers. The decision-maker will then view (slice or dice) the data from interesting perspectives (Power 2004). The EIS directly supports the Intelligence stage; however, CTUC may not have so many data to be analyzed.

In Machine Learning System (MLS), the computer system will simulate learning, organize problem data, and structure the learning model. The decision-makers can then reuse the model provided by MLS (Power 2004). Hence, it is not the type of DSS that can be applied in CTUC because CTUC needs a mechanism to collect each party's opinions in advance and if possible, to compare each party's opinions to the opinions from experts so that decision-makers will know the right direction they should follow, just like the way EIS provides to its managers. The decision analyst will identify several potential decision variables, and then EIS will use a lot of real data to perform the trend analysis, eliminate unimportant variables, and finally show the net effect of these important variables. While in CTUC, if the real data can be replaced by experts' opinions, and parties' opinions can be seamlessly exchanged, the CTUC DSS can be said as an integrated system that has group communication and EIS functionalities.

CTUC Decision Variables. The CTUC decision variables have been identified based on the literature review and experts' opinions from interviews. Experts then assessed the impact level of each variable on the CTUC decision. The basic elements of each CTUC decision variable in the proposed model are described as follows:

(1) Decision Variable Name: The name of the decision variable. For example, HAZMAT.
(2) Decision Variable Description: The short description of the decision variable. For example in HAZMAT, it means when HAZMAT-related work (e.g. asbestos, leaking underground storage tanks, contaminated soils, contaminated groundwater, or unknown substances) ONLY applies to the utility adjustment work.

(3) Possible Value: A list of all possible values pertaining to this decision variable. For example in HAZMAT, Yes or No.
(4) Pro/Con: The preferred approach for each value of the decision variable. There are five possible choices of Pro/Con, namely: (a) Pro-CTUC only; (b) sometimes Pro-CTUC and sometimes Neutral; (c) Neutral; (d) sometimes Anti-CTUC and sometimes Neutral; (e) Anti-CTUC only. For example, experts from state DOTs may prefer not to use the CTUC approach when HAZMAT only applies to the utility adjustment work, while experts from the utility industry may prefer to use the CTUC approach under the same situation.
(5) Impact Level: The impact on the CTUC decision. When Pro/Con is "Pro-CTUC" or "Pro-CTUC or Neutral," the possible impact levels are: High, Medium, Low, or No Impact. When Pro/Con is "Anti-CTUC" or "Anti-CTUC or Neutral," the possible impact levels are: Show-Stopper, High, Medium, Low, or No Impact. Note that Show-Stopper should be marked only when the circumstance precludes further analysis of CTUC; in other words, the conventional approach would definitely be used for the project. From experts' perspective, they were asked to indicate the preferred approach first under the specified project circumstance. Then, they assessed the impact level of this situation on the CTUC decision. Finally, the average impact levels of all experts' assessments were calculated, classified, and stored in the knowledge base of the CTUC Decision Support System.
(6) Situation Resolvable: Whether or no the situation can be changed to facilitate CTUC. It indicates the controlling party responsible for possible process changes to facilitate CTUC. As mentioned before, since the CTUC decision is the result of a series of negotiation activities, decision-makers need to know whether or not the value of this decision variable can be changed. If the project circumstance can be changed so that "Anti-CTUC" can become "Neutral" or "Pro-CTUC", or "Low Impact" of "Pro-CTUC" can become "High Impact," the efforts to make the change should be persuaded. The possible answers are: "Yes, the process can be changed; but the responsible party is unknown"; "State DOT is responsible for the process change"; "Utility is responsible for the process change"; "Other parties are responsible for the process change"; "No, there is no chance to change the process to facilitate CTUC."
(7) Confidence Level: How confident the decision-maker feels about this answer. The possible choices are: "Sure Answer"; "Somewhat Sure Answer"; "Don't Know at This Time".

CTUC Decision Support System

A Decision Support System based on the proposed model is being developed. When CTUC decision-makers use the system, the first thing is to select the correct value of each decision variable that can best describe their specific project circumstance. Then, they evaluate the confidence level of this answer. The system will show Pro/Con-CTUC, Impact Level on the decision, Situation Resolvable from previous assessment results from experts. CTUC decision-makers can therefore review these data and may make adjustments to the above elements based on specific project conditions. As discussed before, the assessment results from experts are in the general context. Decision variables may behave differently depending on the value of the other decision variables. For example, in the experts' assessment level, the decision variable of unacceptable specifications usually means utility owners cannot provide a set of specifications that is acceptable to state DOTs in terms of assignment of responsibility, liability, and risk. Hence, it is usually Anti-CTUC and has High Impact on the CTUC decision. However, if the utility is a public utility and has a good relationship with its state DOT (the other decision variables), the above situation become Neutral because the public utility is normally

willing to let the state DOT manage the utility adjustment project, including hiring utility adjustment design consultants.

Two major reports of the system are shown in Figure 1 and Figure 2. Figure 1 summarizes the results from a DOT and a utility owner for the CTUC decision on a hypothetical utility adjustment. The top 6 Pro-CTUC and Ant-CTUC decision variables from DOT's perspective are displayed, compared with Utility ABC's perspective. DOT and Utility ABC may have different impact levels, e.g., both DOT and the Utility ABC think "Specification Consistency = High" is Pro-CTUC; however, DOT thinks it has high impact while Utility ABC thinks it has low impact on the CTUC decision. Sometimes DOT and Utility ABC don't have the same opinion on the same decision variable. For example, DOT thinks "HAZMAT = only on utility" while Utility ABC thinks "HAZMAT = on both sides." In this case, further clarification and negotiation activity is necessary because both parties may have different understandings on the same issues.

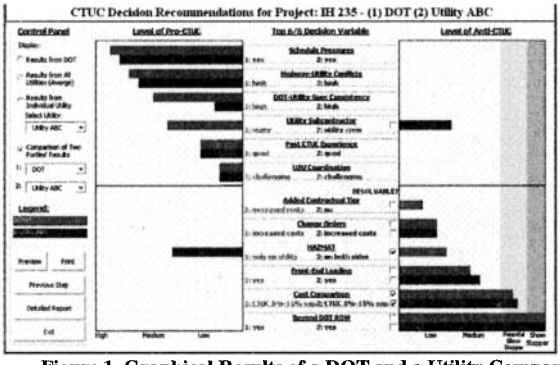


Figure 1. Graphical Results of a DOT and a Utility Company

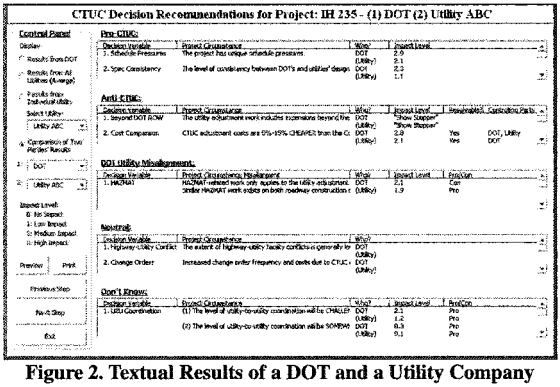


Figure 2. Textual Results of a DOT and a Utility Company

For Anti-CTUC decision variables, "Show-Stopper" implies that the adjustment work should not be considered to use the CTUC approach. The system may suggest DOT that they should negotiate the issue, "Cost Comparison = CTUC is 5% - 15% more expensive than Conventional," with Utility ABC first because it is resolvable from experts' perspective. For the decision variables which are either unknown to decision-makers currently or decision-makers cannot be sure about, the system will also display the textual report listing these Pro-CTUC or Anti-CTUC decision variables with high impact levels from experts' perspective as shown in Figure 2. The text report acts as a checklist for decision-makers to make sure none of the Anti-CTUC decision variables become valid when the project is going.

Conclusion

Successful implementation of the CTUC approach requires systematic analysis and early decision-making based on CTUC decision variables. Both the conventional approach and the CTUC approach should be treated without bias. The CTUC decision support model was designed to capture project circumstances and to highlight the factors that impact the decision on the most appropriate project delivery method for a given utility adjustment. The system itself serves as an intelligence tool that assists decision-makers in capturing and evaluating the project circumstances in a systematic way.

References

AASHTO. (2004). "Strategic Plan Strategy 4-4: Right-of-way and Utilities Guidelines and Best Practices." *Research Report*, American Association of State Highway and Transportation Officials, Washington, DC.

Cisneros, L. (1996). "Timely Coordination of Utility Relocation for Highway Purposes." *Fifth National Highway Utility Conference*, FHWA, Washington, DC, 35-38.

Ellis, D. (1996). "Timely Coordination of Utility Relocation for Highway Purposes." *Fifth National Highway Utility Conference*, FHWA, Washington, DC, 43-46.

Ellis, R.D., and Thomas, H.R. (2003). "The Root Causes Of Delays in Highway Construction." *82nd Annual Meeting of the Transportation Research Board*, TRB, Washington, D.C.

Forgionne, G.A. (2000). "Decision-Making Support System Effectiveness: The Process to Outcome Link." *Information Knowledge Systems Management*, IOS Press, Amsterdam, Netherlands, 2(2), 169-188.

GAO. (1999). "Transportation and Infrastructure: Impacts of Utility Relocations on Highway and Bridge Projects." *GAO/RCED-99-131*, United States General Accounting Office, Washington, DC.

Holsapple, C.W., and Whinston, A.B. (1996). *Decision Support Systems: A Knowledge-Based Approach*, West Publishing, St. Paul, MN.

Quinn, S.B. (1997). "Design/Build: An Alternative Delivery System." *Proceedings of the XIIIth World Meeting of the International Road Federation*, Toronto, Canada.

Simon, H. (1960). *The new science of management decision*, Harper & Row, New York.

TRB. (2004). "Economics of Including Utility Relocations in Roadway Construction Contracts." *Committee Research Problem Statements: AFB70 RPS05*, Transportation Research Board, Washington, DC.

Vidalis, S., and Najafi, F.T. (2002). "Innovative Construction Methods for Highway Construction Projects in Minimizing Cost and Time Overruns." *International Journal of Pavements*, CEFET-SP, 1(3), 58-68.

A Study of Lane Change Frequency on a Multilane Freeway

Vivek Goswami[1] and Ghulam H. Bham[2]
[1] Graduate Student, Univ. of Missouri Rolla, PH (573) 341-6548 (vg796@umr.edu)
[2] Assistant Professor, Univ. of Missouri Rolla, PH (573) 341-6286 (ghbham@umr.edu)

Abstract

Traffic flow characteristics like volume, density and speed are used to calibrate and validate current simulation models. However, these models do not validate microscopic characteristic like lane changing behavior of drivers. To get realistic results in terms of each lane of a freeway, the frequency of lane changes should be accounted for in validation of simulation models. A study of lane change frequency on a multilane freeway on I-80 California is presented in this paper. A half a mile section with five lanes, an on-ramp and a shoulder lane drop is studied. The numbers of lane changes in terms of origin and destination lanes and their percentages with respect to the entrance volume are presented for each freeway lane. Comparisons of lane change frequency towards the shoulder lane and towards the median lane are carried out. The lane changes are classified into anticipatory, mandatory and discretionary lane changes on the basis of observed data. Due to mandatory lane changes, near the vicinity of entrance gore, a zone of intense lane change has been identified. This zone is prone to higher frequency of accidents as a result of higher number of conflicting points. As the shoulder lane drops, a late mandatory merge is identified downstream of the section. This late mandatory merge accounts for about 29% of the total vehicles moving from the shoulder lane to the adjacent lane.

BACKGROUND AND MOTIVATION

Current traffic simulation models are considered valid if their outputs are in agreement with the field data in terms of volume, density and average speed. It is important to compare results from the simulation models with field data taking individual lanes into account rather than the overall facility. Due to the presence of on-ramps and off-ramps on a multilane freeway, a higher number of lane changes, particularly mandatory lane changes are made. Because of these lane changes there is a direct affect on driver behavioral parameters like speed, acceleration/deceleration, accepted gaps, etc. The lane changing vehicles affect the vehicles in the immediate vicinity as well as the traffic stream. On a multilane freeway it is, therefore, important to study the number of lane changes in each lane and simulation models should account for them to get more realistic results.

It is important to study the frequency of lane changes on multilane freeways to understand and develop capacity models for freeways and improve the geometrics of the multilane freeways to increase safety. Lane changing processes also play an important role in freeway capacity and safety (Sparmann, 1979). Lane changes are a major cause of traffic instability (Leutzbach, 1988). Average lane change frequencies exhibit a weak but systematic relationship to traffic and roadway conditions (Worrall and Bullen, 1970). For multilane freeways, lane change frequencies of exiting vehicles close to their intended off-ramp and frequencies of through vehicles are a function of distance from the off-ramp. The highest frequencies of lane changes of exiting vehicles were found to be toward the right lane and corresponding frequency of through vehicle lane changes is towards the left lane (Pahl, 1970). Lane change crashes account for 4% to 10% of total crashes on freeways (Lee et. al., 2004). Redelememeier and Tibshirani (2000) presented that 90% of lane change crashes are due to driver error rather than failure of the vehicle or roadway.

DATA COLLECTION AND SITE DETAILS

This paper uses a thirty minute data set collected on I80 in Emeryville, CA. The data set was collected as part of the NGSIM project and consists of detailed vehicle trajectories collected at every 66.77 milliseconds for a half-mile section. The site is shown in Figure 1. There are five main lanes and a shoulder lane (lane 6) which drops and merges with the off-ramp.

LANE CHANGE TYPE

Lane changes are classified as: anticipatory, mandatory and discretionary lane changes. Anticipatory lane changes are made in anticipation of merging traffic from a downstream on-ramp. They are made to avoid speed reduction due to merging vehicles. Mandatory lane changes are made to take an exit, merge with the freeway traffic and to avoid a lane closure. Discretionary lane changes are made to enter a desired lane. Drivers make discretionary movement when they are unsatisfied with current driving situations and favorable conditions exist in adjacent lanes. Discretionary lane changes are not discussed in detail in this paper.

Table 1. Number of Lane Changes from Origin to Destination Lane as Percentage of Entering Volume

To / From	No. of Vehicles Entering	1	2	3	4	5	6	Off-Ramp (8)
1	**870**	**736 (84.6%)**	108 (12.4%)	21 (2.5%)	1 (0.1%)	1 (0.1%)	0	2 (0.2%)
2	**773**	105 (13.7%)	**506 (65.5%)**	117 (15%)	30 (3.9%)	9 (1.1%)	0	6 (0.7%)
3	**591**	28 (4.9%)	164 (27.7%)	**268 (45.3%)**	100 (16.9%)	14 (2.4%)	0	16 (2.7%)
4	**715**	20 (2.8%)	96 (13.4%)	228 (31.8%)	**253 (35.4%)**	67 (9.2%)	0	52 (7.3%)
5	**719**	20 (2.8%)	42 (5.6%)	110 (15.4%)	255 (35.5%)	**219 (30.4%)**	0	74 (10.3%)
6	**751**	2 (0.2%)	5 (0.6%)	28 (3.6%)	99 (13.1%)	394 (52.4%)	0	224 (29.8%)
On-Ramp(7)	**307**	6 (1.9%)	10 (3.2%)	37 (12%)	86 (28%)	155 (50.5%)	0	12 (3.9%)
Total	**4726**	**917**	**931**	**809**	**824**	**859**	**0**	**386**

Note: Number of through vehicles in each lane is in bold.

DISCUSSION OF RESULTS

Total Number of Lane Changes

Table 1 presents the total number of lane changes in terms of origin and destination lane and their percentages with respect to the entrance volume. Due to a lane drop there is an increase in the number of vehicles leaving from each lane as compared to the number of vehicles entering in each lane over the half hour of data. It is observed that in lane 1 and lane 2, a higher number of lane changes are towards right side. From lane 3 and onwards there is a higher number of lane changes towards left side. It is observed that vehicles moving towards the shoulder lane show a continuous decrease in number and percentages of through vehicles. This is shown in bold in Table 1. It is observed that the percentage of lane changes is higher when

moving towards the median lane as compared to moving towards the shoulder lane. It is also observed that shoulder lane (lane 6) to lane 5 has the highest percentage (52.4%) of lane changes with respect to the entering volume. This is because the shoulder lane drops and all the vehicles on the shoulder lane have to move to lane 5 or exit the section. Figure 1 presents more details of lane changes made in subsections of 200 feet for every lane.

Single and Multiple Lane Changes

Table 2 presents the total number and percentages of single and multiple lane changes made in the direction of median and shoulder lanes separately. It is observed that multiple lane change maneuvers (2 and more) made by a single driver towards the median lane (1399) are about four times more as compared to the shoulder lane (350). This shows the traffic movement from the on-ramp. After entering the section, these drivers make multiple lane changes to come in the desired lanes. In case of vehicles moving towards the shoulder lane and planning to exit from the off-ramp, higher numbers of vehicles were close to the shoulder lane before start of the study section and therefore these vehicles have to make fewer multiple lane changes. Similarly, number of single lane changes towards median lane (1301) is about three times more as compared towards the shoulder lane (466). Overall in the section it is observed that lane changes towards median lane are about three and a half times more as compared to those towards shoulder lane. This signifies the difference in driver behavior entering from the on-ramp and moving towards the median lane, and vehicles moving towards the shoulder lane.

Table 2. Number of Vehicles making Single and Multiple Lane Changes

Lane Change Maneuver	No. of lane changes towards shoulder lane		No. of lane changes towards median lane		Total
(Col. 1)	(Col. 2)	Col. 3 = (Col. 1 x Col. 2)	(Col. 4)	Col.5 = (Col. 1 x Col. 4)	Col. 6 = Col.3+Col.5
Single	466	466 (57.1%)	1301	1301 (48.2%)	1767 (50.2%)
Double	117	234 (28.7%)	419	838 (31%)	1072 (30.5%)
Triple	26	78 (9.6%)	127	381 (14.1%)	459 (13%)
Quadruple	7	28 (3.4%)	35	140 (5.2%)	168 (4.8%)
Pentuple	2	10 (1.2%)	8	40 (1.5%)	50 (1.4%)
Total		816 (100%)		2700 (100%)	3516 (100%)

Lane Change Frequency towards Shoulder Lane

A total of 816 lane changes were made towards the shoulder lane (Table 2). It is observed that overall, there is a continuous increase in the frequency of lane changes from the beginning of the section to the subsection ending at 1100 feet, and then there is a continuous decrease in the frequency of lane changes until the end of the section. The increment is due to the reason that vehicles tend to reach the shoulder lane by making mandatory lane changes to take an exit downstream. The reduction in lane changes after the end of subsection at 1100 feet is because most of the mandatory lane changes were made upstream. Also the stream of vehicles move towards median lane after entering from the on-ramp and shoulder lane (due to lane drop) and this gives less favorable conditions for vehicles to make lane changes towards the shoulder lane.

Lane Change Frequency towards Median Lane

A total of 2700 lane changes were made towards the median lane (Table 2). Out of these, 835 (31% of 2700) mandatory lane changes lane changes were from lane 6 to lane 5, which are because of the lane-drop. Following were the discretionary lane changes: 699 (26%) lane changes were from lane 5 to lane 4, 602 (22%) were from lane 4 to lane 3, 392 (14.5%) were from lane 3 to lane 2 and 183 (6.5%) were from lane 2 to lane 1. Higher number of lane changes were observed towards the median lane as compared towards the shoulder lane due to the following reasons: vehicles merging from the on-ramp make single and multiple lane changes to come to the desired lane; due to the presence of the on-ramp, 222 (31.8%)

anticipatory lane changes are made from lane 5 to lane 4 to avoid speed reduction; and due to the lane drop vehicles have to make mandatory lane changes from lane 6 to lane 5.

Frequency of Anticipatory Lane Changes

Figure 1 shows the frequency of anticipatory lane changes; from lane 5 to lane 4 (four thick arrows). It is observed that 31.8% (222) of the lane changes from lane 5 to lane 4 are anticipatory lane changes. Vehicles moving from lane 6 to lane 5 before start of the entrance gore are not accounted as anticipatory lane changes as those move to lane 5 to avoid the off-ramp. These lane changes are, therefore, mandatory lane changes.

Frequency of Mandatory Lane Changes

Out of 386 exiting vehicles from the off-ramp, 150 (39%) made single or multiple mandatory lane changes (Table 3). The remaining 224 (61%) vehicles were already on the shoulder lane before start of the section and exited from the off-ramp. Also 12 vehicles entered from the on-ramp and exited using the off-ramp.

Table 3. Number of Mandatory Lane Changes towards the Off-Ramp.

Lane Change	Origin-Destination Lane	No. of Lane changes (%)
Single	5 – 6	77 (51.4%)
Double	4 - 6	48 (33%)
Triple	3 – 6	17 (11.3%)
Quadruple	2 – 6	6 (4%)
Pentuple	1 - 6	2 (1.3%)
Total		150 (100%)

There were 593 mandatory lane changes from lane 6 to lane 5 and 242 vehicles merging (late mandatory merge discussed later) to lane 5 from lane 6 due to the lane drop.

Late Mandatory Merge

A late mandatory merge is observed downstream of lane 6 as a result of the lane drop. This region of 35 feet lies between 2325 feet and 2360 feet and vehicles enter lane 5 from lane 6. A total of 242 vehicles (29%) out of 835 vehicles moving from lane 6 to lane 5, enter lane 5. The late mandatory merge represents behavior of the more aggressive drivers who wait until the end of the section to merge in. Out of 242 vehicles only 63 vehicles (20% of 307 vehicles entering from on-ramp) were among the vehicles entering from on-ramp. This shows that 179 (80%) vehicles making the late merge were in lane 6 before the on-ramp. These 179 drivers (21% of the total 835 drivers making mandatory lane change from the shoulder lane to lane 5) can be classified as most the aggressive drivers as they wait till the end of the section to merge in lane 5 even though they were in lane 6 from the start of the study section.

Zone of Intense Lane Changing

Figure 1 shows the frequency of all the lane changes. A zone of intense lane changing between lane 6 and lane 5 is identified at the vicinity of the entrance gore (600 feet) to a subsection ending at 1800 feet. This intense lane-changing zone is created because vehicles enter the section from the on-ramp (lane 7) to lane 6 and move from lane 6 to lane 5 making mandatory lane changes as there is a lane-drop downstream. Similarly, vehicles from lane 5 make lane changes to lane 6 to make an exit from the off-ramp. In this zone of 1200 feet, 60.2% (357) out of 593 (excluding 242 vehicles making late mandatory merge) lane changes from lane 6 to lane 5 are made. Similarly, 68% (117) out of 172 lane changes from lane 5 to lane 6 are made in this zone, thereby making this zone a high intensity lane change zone, spatially. In total, 474 lane changes are made in 1200 feet in half an hour.

CONCLUSIONS

A study of lane change frequency on a multilane freeway with an on-ramp and a lane drop is conducted in this paper. The observations shows that there is a continuous increase in

lane changes until 1100 feet of the section moving towards the shoulder lane and after 1100 feet a continuous decrease in the number of lane changes is observed. A total of 150 mandatory lane changes (39% of the total 386 vehicles exiting from off-ramp) towards off-ramp are observed. This shows that 61% of the vehicles were already in the shoulder lane before the start of the section. There is a continuous increase in mandatory lane changes moving towards the off-ramp. A 1200 feet zone of intense lane changes, starting at 155 feet before the entrance gore, has been identified due to mandatory lane changes between lane 6 and adjacent lane, which lies between 600 feet to 1800 feet of the section. Sixty percent lane changes from lane 6 to lane 5 and 68% of lane changes from lane 5 to lane 6 occur in this region. Further, due to the lane drop of lane 6, a late mandatory merge is observed between 2325 feet and 2360 feet of the section and 242 (29%) of movements from lane 6 to lane 5 occur in this region. Out of these 242 vehicles, 63 vehicles (20.5% of 307 vehicles entering the freeway from on-ramp) are from vehicles entering from on-ramp and remaining 179 (23.8% of 751) are vehicles entering in lane 6 at the start of the section. A total of 222 vehicles (31.8%) make anticipatory lane changes by moving from lane 5 to lane 4 to avoid the traffic stream entering lane 5 from lane 6.

RECOMMENDATIONS

Proper pavement markings can weaken the zone of intense lane change. From the entrance gore and up to 1200 feet, a solid line is proposed towards the shoulder lane and between the shoulder lane and lane 5 a broken line is proposed towards lane 5. This will prevent vehicles from shoulder lane making lane change to lane 5, whereas vehicles from lane 5 can make a lane change to the shoulder lane. After 1200 feet, 200 feet of broken line is proposed on both sides. This provides opportunity to safely complete lane changes from lane 5 to lane 6. After the end of this 200 feet line the broken line towards lane 5 is discontinued and a solid line is proposed thereby preventing vehicles from making a lane change to lane 6, whereas the broken line towards lane 6 is continued till 1800 feet. In this manner, we are able to create no-lane changing zones alternatively for both the lanes, and reducing the area of intense lane changing. In this way, lane-changing maneuvers can be made more safely compared to the original situation. Additionally, the drivers do not have to face the traffic stream from two directions as in the no-lane change zone; drivers are assured that vehicles from the other lane are not going to enter their lane.

The number of lane changes and the origin-destination of vehicles on a multilane freeway with an on-ramp and a lane drop may be utilized in calibration and validation of traffic simulation models. By validating the model, in terms of individual lane, with detailed information on frequency of lane changes, more realistic results from simulation models can be obtained.

REFERENCES

Lee, S.E.; Olsen, E.C.B. and Wierwille, W.W. (2004). "A Comprehensive Examination of Naturalistic Lane-Changes" Report No. DOT HS 809 702.

Leutzbach, W. (1988). "Introduction to the Theory of Traffic Flow." Spinger- Verlang, Berlin,

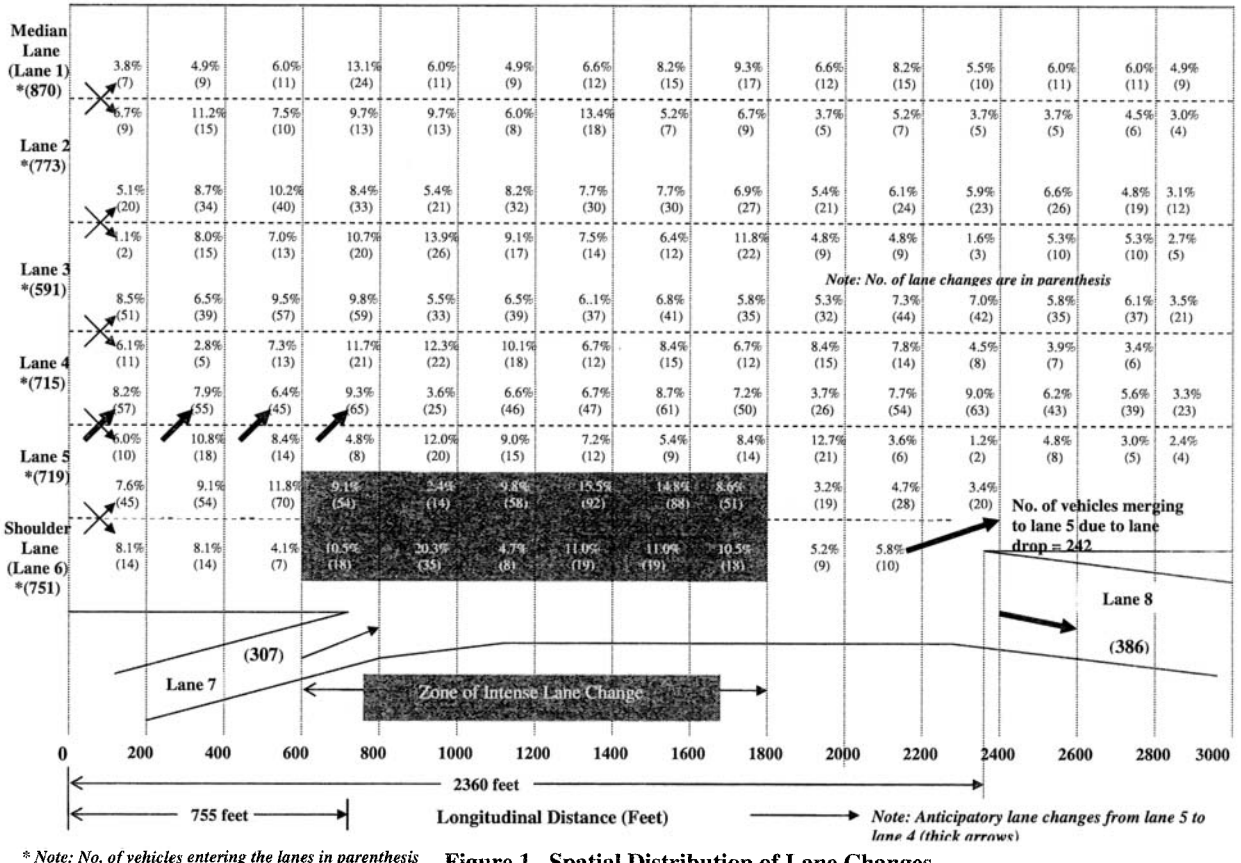
NGSIM Data Set. (2005). I-80 California. Cambridge Systamatics, California. (http://ngsim.camsys.com/)

Pahl, J. (1972). "Lane Change Frequencies in Freeway Traffic Flow." Highway Research Record 409, pp 17- 25.

Redelmeier, D.A. and Tibshirani, R.J. (2000). "Are Those Other Drivers Really Going Faster?" Chance, Vol. 13, No. 3.

Sparmann, U. (1979). "The Importance of Lane Changing on Motorway." Traffic Engineering and Control, 20 (6). pp 320-323.

Worrall, R. D. and Bullen, A.G.R. (1970). "An Empirical Analysis of Lane Changing on Multilane Highways." Highway Research Record. Volume 303. pp 30- 43.



** Note: No. of vehicles entering the lanes in parenthesis below the lane numbers for half an hour data*

Figure 1. Spatial Distribution of Lane Changes

Evaluation of New Coordinated Feedback-Based Freeway Ramp Metering Strategy Using Macroscopic and Microscopic Simulation Models

I. Yasar,[1] K. Ozbay,[2] and P. Kachroo[3]

[1]Rutgers University, Civil and Environmental Engineering, 623 Bowser Road, Piscataway, NJ 08854-8014; Phone: (732) 445-3162; Fax: (732) 445-0577; email: ilgin@eden.rutgers.edu
[2]Rutgers University, Civil and Environmental Engineering, 623 Bowser Road, Piscataway, NJ 08854-8014; Phone: (732) 445-2792; Fax: (732) 445-0577; email: kaan@rci.rutgers.edu
[3]Bradley Dept. of Electrical & Computer Engineering, Virginia Tech., Blacksburg, VA 24061-0111; Phone: (540) 231-2976; Fax: (540) 380-3383; email: pushkin@vt.edu

Abstract

In this paper, C-MIXCROS and D-MIXCROS, 2 new feedback-based coordinated ramp metering strategies, that explicitly consider ramp queues, are proposed and then evaluated using both macroscopic (Rutgers Macroscopic Simulation Environment) and microscopic (PARAMICS) simulation models (on an 11-mile-long corridor of I-295 in South Jersey) under different demand conditions. In addition to the newly proposed coordinated ramp metering strategies, a well-known coordinated strategy (METALINE (5)) and 3 other local strategies (ALINEA (6), New Control (1), and MIXCROS (1)) are also implemented using the same network and results are compared. The proportional-derivative state feedback control logic and direct regulation of on-ramp queues are employed in the derivation of this new proposed coordinated ramp metering strategy. The results from the microscopic simulation are consistent with the macroscopic simulation, where D-MIXCROS and C-MIXCROS both perform better than all other control strategies tested for all the demand scenarios. The deteriorating effect of large on-ramp queues on the total travel time is especially observed for METALINE results; the total travel time is approximately 22% greater than those of C-MIXCROS results. MIXCROS (1) is also successful in keeping the on-ramp queues at a reasonable level for each ramp. However, because it is a local ramp metering strategy, coordinated versions of MIXCROS are observed to be more beneficial both for the ramp system and at the network level.

1. INTRODUCTION

The continuous increase of traffic demand has led to increasingly severe congestions, both recurrent (occurring daily during rush hours) and nonrecurrent (resulting from incidents). One of the most efficient and direct control measures that are typically employed in freeway networks is ramp metering. Ramp metering provides improvement of freeway flow by breaking up platoons, and allowing for more efficient merging, reduction of accidents, and fuel consumption. Freeway control can be open loop (in general time-of-day dependent) or closed loop (traffic responsive). In the first case, controls are derived from a priori known traffic data such as demands and occupancies (e.g. Demand Capacity (2)), whereas closed-loop controls directly react to existing traffic conditions (e.g., ALINEA (6), MIXCROS (1)).

Overall, there are 2 types of ramp metering. Local ramp metering considers an isolated section of the network consisting of a freeway section with 1 on-ramp, and the controller responds only to changes in the local conditions. Coordinated ramp metering is the application of ramp metering to a series of entrance ramps with the goal of coordinating the response of all the ramps in the system. The disadvantage of local ramp metering is its lack of coordination between ramps aiming for optimization of the freeway facility. Coordinated ramp metering, on the other hand, can become very complex and expensive to implement and maintain.

The ramp metering controls that are evaluated in this paper, namely ALINEA, New Control, METALINE, MIXCROS, and the coordinated version of MIXCROS, are briefly described below.

ALINEA is a linearized local feedback control algorithm that adjusts the metering rate to keep the occupancy downstream of the on-ramp at a prespecified level (called the occupancy set point) (6).

METALINE is the coordinated version of the local ramp metering strategy ALINEA. It has been implemented on certain freeways in France, the United States, and the Netherlands. The control logic of METALINE is Proportional-Integral state feedback. The main challenge to the successful operation of METALINE is the proper choice of the control matrices and the target occupancy vector (Table 1). There is no direct consideration of queue overflows in METALINE (5).

MIXCROS, a traffic-responsive local ramp metering control law proposed by Ozbay et al. (2003; (1)), is developed to maximize the throughput on the freeway without creating long queues on the ramp via the use of

carefully calibrated weight parameters for the freeway and ramp, namely w_1, w_2. The control logic of MIXCROS is Proportional-Derivative state feedback. Complete derivation of MIXCROS is given in (1). MIXCROS proved to be very effective in reducing the congestion on the ramp system while keeping the on-ramp queue at an acceptable level. However, because it is a local feedback-based ramp metering strategy, it produced very little improvement at the network level (3).

2. DESCRIPTION OF THE COORDINATED VERSION OF MIXCROS

The basic model used for the design of the coordinated MIXCROS control law is shown in Figure 1.

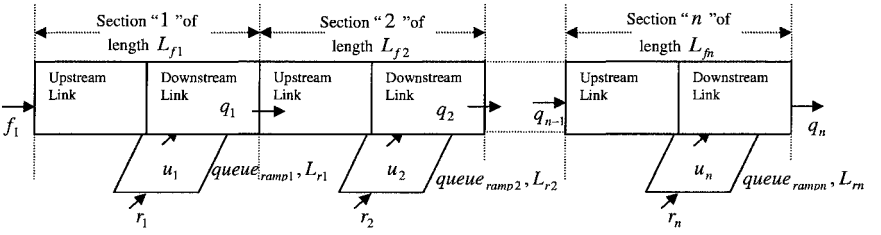


Figure 1. "n" Freeway sections with 1 on-ramp.

In this figure, n is the number of freeway sections, f_i (veh/hr) is the flow entering the freeway at the first section, q_i is the flow leaving the freeway section i (veh/hr), r_i is the flow entering the ramp (veh/hr), u_i is the metered flow (veh/hr), ρ_i is the freeway density (veh/mi), $\rho_{c(i)}$ is the critical density (veh/mi), T T is the time step duration (hr), $w_{1(i)}, w_{2(i)}$ are the weight factors ($w_{1(i)} + w_{2(i)} = 1$), $queue_{rampi}$ is the queue length on the ramp (veh/mi), K_i is the control gain ($0 < K < 1$), L_{fi} and L_{ri} are the length of the freeway and ramp section i (mi), respectively.

The coordinated version of MIXCROS is applied to a number of on-ramps to provide network-wide improvements. It aims to maximize the throughput on all the freeway sections without creating long queues on all the metered ramps. The proportional state feedback control logic (Equation 2) and direct regulation of on-ramp queues are employed in the derivation of this new proposed coordinated ramp metering strategy.

This coordinated traffic-responsive ramp metering control achieves its goal by minimizing the following error function (Equation 1). Therefore, the control objective of coordinated MIXCROS is to make the following error function go to 0 as the time, t, goes to infinity:

$$e(t) = \sum_{i=1}^{n} \left(\left| w_{1(i)} x_{1(i)}(t) \right| + \left| w_{2(i)} x_{2(i)}(t) \right| \right)_i, \tag{1}$$

where $x_{1(i)}(t) = \rho_i(t) - \rho_{cr(i)}$, $x_{2(i)}(t) = queue_{rampi}(t)$, $i = 1,2,...n$ i : section index

The error function (Equation 1) is defined as summation of the absolute values of the state variables $x_{1(i)}$ and $x_{2(i)}$. The state variable $x_{1(i)}$ represents the freeway section of the network, and $x_{2(i)}$ represents the on-ramp queue. Equation 1 takes these state variables into account and determines how much importance should be given to freeway density and queue length on the ramp with the help of weights, w_1 and w_2.

The control objective can be achieved by designing a control law that makes the system follow the below closed-loop dynamics.

$$\dot{e}(t) + Ke(t) = 0 \tag{2}$$

The system can be in 2^n regions, where n is the number of freeway sections in the network because of the first term in the error function ($\left| w_{1(i)}(\rho_i(t) - \rho_{cr(i)}) \right|$). These 2^n regions can be combined to form a control law that is applicable in all regions with the help of a function *sign* such that

$$sign = \begin{cases} 1 & if\ \rho_i(t) > \rho_{cr(i)} \\ -1 & else \end{cases}.$$

The control variables should satisfy is the following condition:

$$Gu(t) = -F(t) - Ke(t),$$

where

$$F(t) = \sum_{i=1}^{n}\left(w_{1(i)} \times sign \times \left(\frac{1}{L_{fi}} \times (-q_i(t) + f_i(t)) \right) + w_{2(i)} \left(\frac{1}{L_{ri}} \times r_i(t) \right) \right)$$

and

$$Gu(t) = \sum_{i=1}^{n} (sign \times w_{1(i)} \times \frac{1}{L_{fi}} - w_{2(i)} \times \frac{1}{L_{ri}}) \times u_i(t).$$

The control law can be designed in a decoupled way or coupled way. The decoupled control law (D-MIXCROS) for each freeway section is:

$$u_i(t) = (sign \times w_{1(i)} \times \frac{1}{L_{fi}} - w_{2(i)} \times \frac{1}{L_{ri}})^{-1}(-F_i(t) - Ke_i(t)).$$

The coupled control law (C-MIXCROS) is:

$$u_i(t) = \alpha_i \times (sign \times w_{1(i)} \times \frac{1}{L_{fi}} - w_{2(i)} \times \frac{1}{L_{ri}})^{-1} \times (-F(t) - Ke(t)),$$

where $\sum_{i=1}^{n} \alpha_i = 1, F = \sum_{i=1}^{n} F_i(t)$, and $e(t) = \sum_{i=1}^{n} e_i(t)$.

The complete derivation of the above control law, which is outside the scope of this paper, is offered in (1).

In Table 1, the calibration parameters for each ramp metering strategy are presented. As it is seen from this table, local controls, namely ALINEA and New Control, require the least amount of parameters. On the other hand, coordinated control, namely D-MIXCROS, has only 7 calibration parameters, giving it an advantage over the other coordinated control evaluated, METALINE. The comparison of the performances of each ramp metering strategy is provided in the simulation results section based on the macroscopic simulation environment modeling.

Table 1 Calibration Parameters for Each Ramp Control Implementation

Ramp Metering Strategy	Calibration Parameters for Each On-ramp	Total number of Calibration Parameters for 6 On-ramps
ALINEA	K_R	6
New Control	K_R	6
MIXCROS	K_R, w_1, and w_2	12
METALINE	K_{R1} and K_{R2}	12
D-MIXCROS	K_R, w_1, and w_2	7
C- MIXCROS	K_R, α, w_1, and w_2	13

D-MIXCROS is similar to the local version of MIXCROS because it takes each on-ramp system into account separately. However, it differs from the local version because D-MIXCROS uses the same control gain K_R for all the ramps, which ensures a unity in the actions of metered ramps. From a practical view, it is easier to implement D-MIXCROS than the local version of MIXCROS on a number of ramps. D-MIXCROS implementation on 6 on-ramps requires 7 parameters to be calibrated, whereas MIXCROS uses 12 calibration parameters (Table 1). Hence, it is easier to implement D-MIXCROS compared with both the local version of MIXCROS and C-MIXCROS on a number of ramps.

In C-MIXCROS, the control effort is distributed among all on-ramps using distribution factor α_i, which provides the communication between on-ramp systems. With "this allocation ratio α_i", the nature of the congestion in each ramp system can be handled in a very detailed manner. For example, if the congestion on the 2nd ramp is

propagating towards the other on-ramp locations, the allocation ratio, α_2, for the 2nd on-ramp can be reduced; making it less than the other ratios ($\alpha_2 < \alpha_1, \alpha_3$).

3. MACROSCOPIC SIMULATION

The proposed coordinated ramp metering strategies C-MIXCROS and D-MIXCROS, as well as ALINEA, MIXCROS, and METALINE, are tested on 6 consecutive ramps along a corridor. Each ramp system consists of a 1-lane (1 mile) freeway link and a 1-lane (0.5 mile) ramp link. The simulation duration for each tested case is 300 min.

In both macroscopic and microscopic simulation models, for ALINEA, New Control, and METALINE implementations, a queue threshold of 35 vehicles is used. New Control and METALINE, as well as all versions of MIXCROS, perform satisfactorily without a queue override strategy that shuts off the ramp metering and creates unwanted fluctuations. In ALINEA implementation, for the values of parameter K_R above 240 veh/hr, on-ramp queues are decreased, whereas ramp metering provides no improvement on the downstream traffic conditions in the ramp systems. So, the purpose of the control, which is to keep the downstream freeway section at the set level, is not met. Therefore, the regulator parameter, K_R, is kept between 70 veh/hr and 240 veh/hr for all the ramps.

All the tested ramp metering strategies maintain the freeway outflow close to the capacity, while keeping the traffic density below critical density. All the controls except D-MIXCROS and C-MIXCROS experience high fluctuations in the traffic density within the1st 50 min of simulation. Queue override tactics employed in these controls mainly cause this problem. That is, these controls use the storage capacity of the on-ramps, which leads to increased traffic flow on the freeway sections. This increased traffic flow results in congestion in downstream locations, causing more restrictive ramp metering rates so as to serve additional throughput from the upstream ramp systems.

Among all tested controls, METALINE has the largest on-ramp queues owing to its restrictive metering. Decreasing freeway demand by 5.26% (Demand Scenario 1) leads to increase in the freeway maximum outflow (throughput) with each ramp metering strategy. With reduced freeway demand, each ramp metering strategy results in approximately the same total travel time. To observe the behavior of the controls in the presence of large ramp demand (Demand Scenario 2), ramp demand is increased by 67% compared with the base demand scenario, lowering the freeway demand by 33% because of the limited capacity of the freeway segments. With this demand configuration, all versions of MIXCROS provide superior individual ramp performance results (e.g., increased average freeway downstream flow, speed, and density) compared with all the strategies tested. In Demand Scenario 3, ramp demand is increased only by 33% whereas freeway demand is lowered by 33%; all ramp metering strategies, both local and coordinated, provide almost the same improvements on the network level. Because of light ramp demand, METALINE is also able to keep the on-ramp queues at reasonable levels with the help of a queue override tactic (Table 2). In this table, A refers to the total travel time on all 6 ramps (veh.hr/hr) and B stands for the total travel time in the network (veh.hr/hr).

Table 2. Overall Network Results for 4 Demand Scenarios

	Base Demand		Demand Scenario 1		Demand Scenario 2		Demand Scenario 3	
	A	B	A	B	A	B	A	B
ALINEA	13.73	247.17	13.02	239.58	19.62	239.70	15.21	198.29
New Control	12.70	246.21	12.22	238.36	19.24	239.57	15.29	198.60
MIXCROS	11.92	248.65	12.06	238.44	18.37	238.77	14.91	198.17
METALINE	79.87	314.61	13.44	243.43	459.95	680.33	19.55	201.43
D-MIXCROS	11.92	248.65	12.08	238.43	18.37	238.76	14.91	198.17
C-MIXCROS	11.92	244.56	12.06	238.44	18.37	238.77	14.91	198.18

4. MICROSCOPIC SIMULATION

A PARAMICS model of the section of I-295 in South Jersey is created using the available geometric and traffic demand data. The calibrated and validated model of the 11-mile-long 3-lane freeway section includes the junctions of I-295 with Route 38, State HWY 73, State HWY 70, and Berlin Rd. Each on-ramp has 1-lane. Then, an Application Programming Interface (API) is coded to assign green times based on each tested control law to all 4 on-ramps in PARAMICS. In the API file, it is ascertained that the calculated green phase duration is within specified limits (i.e. minimum and maximum values are 2 and 15 seconds, respectively). Statistics are collected for 3-hr simulations from the detectors located downstream and upstream of the ramp and 2 additional detectors, 1 at the exit and 1 at the entrance of the ramp. In the microscopic simulation model, the proposed ramp metering controls (namely D-MIXCROS and C-MIXCROS), ALINEA, and MIXCROS are evaluated and compared with the No

Control scenario using 3 demand scenarios, whose congestion levels are listed in Table 3. The congestion level is the percent of the time that the downstream link occupancy is greater than the critical occupancy. All simulations are run for 3 hr with different seed values for the statistical analysis of the results (which ensures a 95% confidence level).

After implementing various versions of MIXCROS using PARAMICS API, a series of simulation runs are carried out to determine the gain parameter K_R and weight factors w_1 and w_2, and because, it is complicated to analytically determine these values that impact the performance the ramp metering control law. The same procedure is followed for the calibration of ALINEA. This calibration approach is similar to the one adopted in (11). The critical occupancy values for each metered ramp is determined using occupancy flow plots obtained from the simulated data generated using PARAMICS. These values are kept constant during the course of the simulation for each scenario. The changes in the capacity experienced in real-world operation are taken into account in another study conducted by the authors of this paper (13).

Table 3. Congestion Levels on Each Ramp

	1st Ramp	2nd Ramp	3rd Ramp	4th Ramp
1st Demand Level	27%	60%	68%	20%
2nd Demand Level	53%	39%	58%	9%
3rd Demand Level	0%	17%	35%	24%

Ramp metering controls seem to be more effective under certain demand patterns than others. As traffic demand increases, ramp metering tends to be more effective in reducing system travel time. The reason for reduced ramp metering performance for the 3rd demand scenario is the low level of congestion on each ramp system. It was also claimed in other studies that the effectiveness of ramp control varies depending on the severity of congestion (12). Table 4 summarizes the main findings of this implementation. All controls tested except ALINEA reduced the average travel time regardless of the demand scenario compared with No Control.

During the preliminary calibration runs, it is observed that ALINEA tends to cause large on-ramp queues at the expense of relieving congestion on downstream freeway links. Therefore, an on-ramp queue threshold (> # of 35 vehicles) is used. This solution also employed by the developers of ALINEA (6) frequently causes green time to be set to maximum and releases an excessive number of vehicles on the ramp to prevent spillbacks onto the arterial causing fluctuations of freeway traffic. This in turn reduces system-wide benefits obtained due to the deployment of ALINEA.

When studying Table 4 that depicts the system performance, ALINEA results obtained from our simulation study for the overall system show resemblance to other studies conducted independently by other groups (e.g., (14) and (3)), that also found that the improvements of mainline freeway traffic conditions due to ramp control do not overweigh the deterioration of the traffic performance on the on-ramps as a result of the use of ALINEA. For the demand scenarios tested, both C-MIXCROS and D-MIXCROS lead to maximum improvement for all the performance criteria.

Table 4. Overall Network Results for 3 Demand Scenarios

	Demand Scenario 1		Demand Scenario 2		Demand Scenario 3	
	Tot. Travel Time (veh.hr)	Mean Speed (mph)	Tot. Travel Time (veh.hr)	Mean Speed (mph)	Tot. Travel Time (veh.hr)	Mean Speed (mph)
No Control	3723.57	55	4119.18	49.6	3408.15	57.8
ALINEA	3901.73	52.13	4137.14	49.43	3368.08	58.23
Change (%)	4.78	-5.23	0.44	-0.35	-1.18	0.74
MIXCROS	3674.04	55.63	3976.06	51.4	3354.74	58.25
Change (%)	-1.33	1.14	-3.47	3.63	-1.57	0.78
D-MIXCROS	3552.71	57.28	3801.35	54.1	3394.75	57.83
Change (%)	-4.59	4.14	-7.72	9.07	-0.39	0.43
C-MIXCROS	3546.94	57.25	3938.86	52.15	3354.98	58.05
Change (%)	-4.74	4.09	-4.38	5.14	-1.56	0.43

5. CONCLUSION

Evaluation of the new coordinated ramp metering strategy is performed to demonstrate its characteristics and eventually its impact on the ramp system and whole network in 2 phases. The 1st phase includes the macroscopic testing of the proposed coordinated ramp metering controls using RMSE (Rutgers Macroscopic Simulation Environment) to compare it with 3 local (ALINEA, New Control, and MIXCROS) and 1 coordinated (METALINE) ramp metering control under the various demand scenarios. The 2nd phase involves evaluating the proposed methodology using a microscopic simulation environment (PARAMICS) under 3 different demand scenarios.

From these implementations, it is found that the system performs better after the implementation of the coordinated version of MIXCROS (1), namely C-MIXCROS and D-MIXCROS, compared with other ramp metering controls. As expected, the mainline freeway experiences better traffic conditions when any of the tested ramp metering controls is implemented. However, when the queue thresholds are used in ALINEA, New Control, and METALINE to prevent the ramps from being overloaded, the system benefits of these strategies are reduced. C-MIXCROS and D-MIXCROS significantly improve system performance compared with other controls under various demand conditions and they are proven to be quite effective. Well-tuned parameters are critical to successful for good ramp metering performance. As opposed to some coordinated ramp metering controls that employ optimization techniques, parameter calibration for C-MIXCROS and D-MIXCROS is comparatively not burdensome.

REFERENCES

1. Kachroo, P., and K. Ozbay (2003). *Feedback Ramp Metering in Intelligent Transportation Systems*. Kluwer Academics.
2. Masher, D. P., D. W. Ross, P. J. Wong, P. L. Tuan, and S. Peracek (1975). Guidelines for Design and Operating of Ramp Control Systems. In: Stanford Research Institute Report NCHRP 3-22, SRI Project 3340, SRI, Menid Park, CA.
3. Ozbay, K., I. Yasar, and P. Kachroo. Comprehensive Evaluation of Feedback Based Freeway Ramp Metering Strategy by Using Microscopic Simulation: Taking Ramp Queues into Account. In: *Transportation Research Record*, No. 1867, TRB, National Research Council, Washington, D.C., 2004, pp. 89-96.
4. Papageorgiou, M. Modelling and Real-Time Control of Traffic Flow on the Southern Part of Boulevard Peripherique in Paris. Internal Report. INRETS-DART, Arcueil, France, 1988.
5. Papageorgiou, M., J-M. Blossville, and H. Hadj-Salem. *Modelling and Real-Time Control of Traffic Flow on the Southern Part of Boulevard Peripherique in Paris, Part II: Coordinated On-Ramp Metering*. In: *Transportation Research*, Vol. 24 A, 1990, p. 361-370.
6. Papageorgiou, M., H. S. Habib, and J. M. Blosseville. ALINEA: A Local Feedback Control Law for On-ramp Metering. In: *Transportation Research Record*, No.1320, TRB, National Research Council, Washington, D.C., 1991, p. 58-64.
7. Papageorgiou, M., and A. Kotsialos. In: *IEEE Transactions on Intelligent Transportation Systems*, Vol. 3, No. 4, December 2002.
8. Yasar, I., Modelling and PARAMICS Based Evaluation of New Local Freeway Ramp Metering Strategy that Takes into Account Ramp Queues, M.S. Thesis, Rutgers University.
9. Yin, Y., Liu, H., and Benouar, H. A Note on Equity of Ramp Metering. In: *IEEE Intelligent Transportation Systems Conference*, Washington, D. C., USA, October 3-6, 2004.
10. Zhang, H. M., and W. W. Recker. On Optimal Freeway Ramp Control Policies for Congested Traffic Corridors. In: *Transportation Research, Part B: Methodological*, 33, 6, 1999, p. 417-436.
11. Zhang, M., T. Kim, X. Nie, W. Jin, L. Chu, and W. Recker. Evaluation of On-ramp Control Algorithms, California PATH Research Report, UCB-ITS-PRR-2001-36, December 2001.
12. Jin, W. and M. Zhang. Evaluation of On-ramp Control Algorithms. In: California PATH Working Paper, UCB-ITS-PWP-2001-14, 2001.
13. Ozbay, K., I. Yasar, and P. Kachroo. Development and Evaluation of Online Estimation Methods for Feedback-Based Freeway Ramp Metering Strategy, Presented in 85th TRB Annual Meeting, 2006.
14. Gardes, Y., A. Kim, and A. May. Bay Area Simulation and Ramp Metering Study-Year 2 Report, California PATH Report, 2003.

Continuous network design with emission pricing as a bi-level optimization problem

Tom V. Mathew[1] and Sushant Sharma[2]

[1] Assistant Professor, Department of Civil Engineering, Indian Institute of Technology Bombay, India; PH: (9122) 2257 7349; FAX: (9122) 2257 7302; email: vmtom@iitb.ac.in
[2] Research Scholar, Department of Civil Engineering, Indian Institute of Technology Bombay, India; PH: (9122) 2576 7349; email:sushantsharma@iitb.ac.in

1. Abstract

Traffic network design problem attempts to find the optimal network expansion policies under budget constraints. This can be formulated as a bi-level optimization problem: the upper level determines the optimal link capacity expansion vector and the lower level determines the link flows subject to user equilibrium conditions. However, in the context of environmental concerns, driver's route choice includes travel time as well as emission pricing. This study is an attempt to solve network design problem when the user is environment cautious. The problem is formulated as a bi-level continuous network design problem with the upper level problem determines the optimal link capacity expansions subject to user travel behavior. This behavior is represented in the lower level using the classical Wardropian user equilibrium principles. The upper level problem is an example of system optimum assignment and can be solved using any efficient optimization algorithms. Genetic Algorithm is used because of its modeling simplicity. The upper level will give a trial capacity expansion vector and will be translated into new network capacities. This then invokes the lower level with these new link capacities and the output is a vector of link flows which are passed to upper level. The upper level then computes the objective function and GA operators are applied to get a new capacity expansion vector and the process is repeated till convergence. The model is first applied to an example network and the optimum results are shown. Finally the model is applied to a large case study network and the results are presented.

2. Introduction

The foremost option that strikes transportation engineers and planners, alike, while considering the growth in traffic demand is to expand the capacity of existing congested links or build new links. In such cases, selecting new links and adding capacity to existing links becomes an interesting problem. The planner has to make decision while considering, on the one hand, minimization of total system cost under limited budget and, on the other hand, behavior of the road users. This problem is stated as network design problem (NDP). The objective of NDP is to achieve a system optimal solution by choosing optimal decision variables in terms of link improvements. The decision taken by the planner affects the route choice behavior of road users and need to be considered while assigning traffic to the network. Network design models concerned with adding indivisible facilities (for example a lane addition) are said to be *discrete*

NDP, whereas those dealing with divisible capacity enhancements (for example road widening) are said to be *continuous* NDP. Among the several route choice models, the most popular one is based on Wardrop's first principle and can be formulated as an equivalent minimization problem.

This equilibrium network design problem can be formulated as bi-level programming problem. At the upper level, the total system travel time, subjected to the construction cost available for the link capacity expansion is minimized. At the lower level, the user equilibrium flow is determined using Frank Wolfe algorithm. Although, there are few attempts to implement these models, a case study illustrating the application of the model to a large network is absent. Therefore, this study is an attempt to find optimal capacity expansion for a large city network

3. Literature Review

The first discrete NDP was developed by LeBlanc (1975) and later it was extended to the continuous version (LeBlanc and Abdulaal, 1979). Hook-Jeeves pattern search algorithm (HJ Algorithm) was used for solving these models. Another kind of network design problem was proposed by Yan and Lam (1996) for optimizing road tolls under condition of queuing and congestion. Optimization of reserve capacity of whole signal controlled network as bi-level programming problem in network design was first time attempted by Wong (1997). Meng et al. (2001) have proposed an equivalent single level continuously differentiable optimization model for the conventional bi-level continuous network design problem. Ziyou and Yifan (2002) combined the concept of reserve capacity with continuous equilibrium network problem. Chen and Yang (2004) came out with two models that considered spatial equity and demand uncertainty in the network design problem. Both models were solved by a simulation based genetic algorithm. Meng et al. (2004) did a comprehensive study of static transportation network optimization problems with stochastic user equilibrium constraints.

Genetic algorithm (GA) based approach for optimizing toll and maximizing reserve capacity was attempted by Yin (2000). They summarized the advantages of using GA based approach as efficient and simple. Ceylan and Bell (2004) used GA-based approach for optimizing traffic signal, while including driver's routing. The major drawback of GA-based model is the expensive fitness evaluation in each generation for the large route network design problem, which is combinatorial in nature (Aggarwal and Mathew, 2004).

To consider the environmental parameters in network design problems various studies have been carried out with traffic assignment like a multi-objective decision model with system optimum conditions (Tzeng and Chen, 1993). A new kind of assignment called *system equitable traffic assignment* was attempted taking generalized environmental cost function (Bendek and Rilett, 1998). Nagurney (2000) considered a multi-criteria traffic network model with emission terms in objective function.

4. Model Formulation

Determining the optimal link capacity expansions is formulated as a bi-level continuous network design problem with the upper level problem minimizes the system travel

cost subject to user's travel behavior. This behavior is represented in the lower level using the Wardropian user's equilibrium principles. However, the link cost function is modified to represent both the link travel time as well as the environmental concerns. The upper level problem is an example of system optimum assignment and can be solved using any efficient algorithms. Although the link cost function is modified, it is still convex and is therefore solved using Frank-Wolfe algorithm.

The following notation has been used for continuous NDP formulation: A is the set of links in the network, ω is the set of OD pairs, q is the vector of fixed OD pair demands, $q_{rs} \in q$ is the demand from node r to node s , R is the set of paths between OD pair r and s, f the vector of path flows, f = $[f_k^{rs}]$, x the vector equilibrium link flows, x = $[x_a]$, y is the vector of link capacity expansions, y = $[y_a]$, B the allocated Budget for expansion, θ conversion factor from emission to travel cost, e_a the emission at link a, Δ the link-path incidence matrix, $\Delta = [\delta_{a,k}^{rs}]$. The bi-level problem can be mathematically represented as below:

Upper Level

$$\text{Minimize } Z_y = \textstyle\sum_a x_a t_a(x_a, y_a) \tag{1}$$
$$\text{subject to}$$
$$\textstyle\sum_{\forall a} \gamma_a y_a \leq B;\ y_a \geq 0\ : \forall a \in A$$

Lower Level

$$\text{Minimize } Z_x = \textstyle\sum_a \int_0^{x_a} \theta_t t_a^w(x_a) + \theta_c e_a(x_a)\, dx \tag{2}$$
$$\text{subject to}$$
$$\textstyle\sum_{\forall k} f_k^{rs} = q_{rs}\ : k \in K; r, s \in q;$$
$$x_a = \textstyle\sum_r \sum_s \sum_k \delta_{a,k}^{rs} f_k^{rs}\ : r, s \in q; a \in A; k \in K$$
$$f_k^{rs} \geq 0\ : r, s \in q; k \in K$$
$$x_a \geq 0\ : a \in A$$

5. Solution Approach

A flowchart of the solution approach is given in Fig. 1. The algorithm starts in the upper level by reading all the inputs like network details, demand matrix, budget, link expansion cost functions, travel time function, and emission cost functions. The upper level algorithm will give a trial capacity expansion vector and will be translated into new network capacities. The lower level algorithm is then invoked with these new link capacities. In this level, the demand matrix is assigned into the network considering both travel time function and emission cost function. The output of this model is a vector of link flows which is passed to the upper level. The upper level then computes the objective function which is the system travel time from link flows and assigns penalty if the budget constraint is violated. This objective function value is given to the upper level algorithm which supplies a new capacity expansion vector and the process is repeated till convergence.

The upper level problem is solved using genetic algorithm which is an iterative procedure that borrows the ideas of natural selection and can easily solve many complex problems.

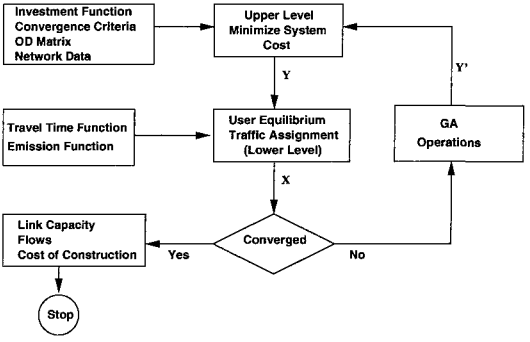


Figure 1: Flowchart representing solution approach

The emission function has been incorporated along with the travel time by converting that into equivalent cost. It is supposed that the user is being charged for producing the emissions in the network. The amount of pollution on link a is given by

$$e_a = x_a e_f d_a \tag{3}$$

where x_a is the traffic flow on link a, e_a is pollution by vehicles on link a in grams (gm), e_f is emission factor (gm/km), d_a is length of link a (km). The emission generated is multiplied by θ_c which is a factor for converting emission into cost.

6. Motivating example

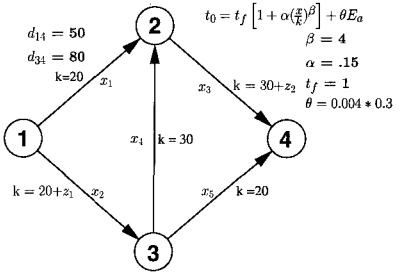


Figure 2: Bi-level example network

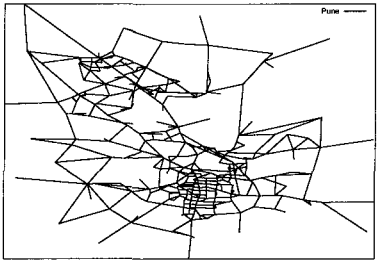


Figure 3: Pune city network

To appreciate the working of the bi-level problem, an example network having 4 nodes and 5 links (Fig. 2) is solved. The two OD pairs involved are q_{3-4} and q_{1-4} and has 50 and 80 trips respectively. The budget is taken as 10 units. The link expansion vector is initialized to 0. User equilibrium is performed to get the required link flows x1*, x2*, x3*, x4*, x5*. Now these flows are input to upper level from where we get

Table 1: Example network results: link flow values, capacity expansion vectors, and total system travel time (TSTT) for six iterations

No	x0*	x1*	x2*	x3*	x4*	UE	TSTT	y1*	y2*	SO
1	42.92	37.96	54.76	11.84	76.68	405.47	846.01	4.56	5.43	785.93
2	40.93	39.81	56.93	16.00	74.27	391.06	764.72	4.68	5.31	764.71
3	40.88	39.85	56.87	15.99	74.33	391.05	764.62	4.72	5.28	764.62
4	40.87	39.86	56.85	15.97	74.35	391.06	764.66	4.73	5.27	764.66
5	40.86	39.88	56.85	15.99	74.35	391.05	764.57	4.73	5.27	764.57
6	40.86	39.88	56.85	15.99	74.35	391.05	764.57	4.73	5.27	764.57

a new set of link expansion vectors y1*, y2* which minimizes the system travel time. Same values of y1*, y2* were used as input for new run of User equilibrium. This iteration is repeated until the total system travel time from lower level and upper level converges. The results are tabulated in Table 1. The convergence of the solution (y1, y2) to the optimum value illustrates the working of the model.

7. Case Study

The network of Pune city was considered for the case study. This city has an area of 138 square kilometers. The city is divided into 97 zones out of which 85 are internal 12 are external zones. There are 273 road nodes 1131 road links, including walking and non walking links. Fig. 3 shows the network of the city.

Various traffic flow characteristics like α, β, free flow speed and running speed for all the links were found after surveying. Link characteristics were collected and trip table was generated for the whole network after validating the OD counts. The emission factor for all the vehicle has been supposed as 25gm/Km. The emission which stays in and causes pollution is supposed to be 30% of whole pollution generated. This is multiplied by a factor which gives the external cost of emission in environment which is 0.004 Rs/gm. In order to study the network expansion, a budget of Rs. 600 crores and a maximum expansion of 100% is considered.

8. Conclusion

Finding optimal network capacity expansion where the user is environment conscious for a transport network is attempted in this study. The problem is formulated as a continuous bi-level optimization problem: the upper level solves the problem of finding the optimal network expansion values and the lower level represents the user travel behavior. The former is solved using GA while Frank-Wolfe algorithm is used in the later case. An example network is used to illustrate the working of the model. A large case study network is solved which demonstrates the ability of the solution methodology to handle large problem. Further study is needed to incorporate a better representation of environmental concerns.

References

Aggarwal, J. and Mathew, T. V. (2004), "Transit route network design using genetic algorithms", *Journal of Computing in Civil Engineering* , Vol. 18(3), pp. 248–256.

Bendek, C. M. and Rilett, L. R. (1998), "Equitable traffic assignment with environmental cost functions", *Journal of Transportation Engineering* , Vol. 124(1), pp. 16–22.

Ceylan, H. and Bell, M. G. H. (2004), "Traffic signal timing optimisation based on genetic algorithm approach, including driver's routing", *Transportation Research Part B* , Vol. 38, pp. 329–342.

Chen, A. and Yang, C. (2004), "Stochastic transportation network design problem with spatial equity constraint", *Transportation Research Record* , Vol. 1882, TRB, Washington, DC.

LeBlanc, L. J. (1975), "An algorithm for discrete network design problem", *Transportation Science* , pp. 183–199.

LeBlanc, L. J. and Abdulaal, M. (1979), "Continuous equilibrium network design problem", *Transportation Research Part B* , Vol. 13, pp. 19–32.

Meng, Q., Lee, D.-H., Yang, H. and Huang, H.-J. (2004), "Transportation network optimization problems with stochastic user equilibrium constraints", *Transportation Research Record* , Vol. 1882, TRB, Washington, DC.

Meng, Q., Yang, H. and Bell, M. G. H. (2001), "An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem", *Transportation Research Part B* , Vol. 35, pp. 83–105.

Nagurney, A. (2000), *Sustainable Transportation Networks*, Glos,UK.

Tzeng, G.-H. and Chen, C.-H. (1993), "Multiobjective decision making for traffic assignment", *IEEE Transactions on Engineering Management* , Vol. 40(2).

Wong, S. C. (1997), "Group-based optimisation of signal timings using parallel computing", *Transportation Research Part C* , Vol. 5(2), pp. 123–139.

Yan, H. and Lam, W. H. K. (1996), "Optimal road tolls under conditions of queuing and congestion", *Transportation Research Part A* , Vol. 30(5), pp. 319–332.

Yin, Y. (2000), "Genetic algorithms based approach for bi-level programming models", *Journal of Transportation Engineering* , Vol. 126(2), pp. 115–120.

Ziyou, G. and Yifan, S. (2002), "A reserve capacity model of optimal signal control with user equilibrium", *Transportation Research Part B* , Vol. 36, pp. 313–323.

Resectorization Analysis of Airspace Sector Capacity

Sui-Ling Li[1]

[1]Associate Professor, Department of Shipping & Transportation Management, National Penghu University,300, Liow-Her Road, Maa-Gong, Penghu Shiann,88042 Taiwan, R.O.C.; Tel:886-09-9264115-3213; Fax: 886-06-9279812; e-Mail:suiling@npu.edu.tw

Abstract

Measuring airspace sector capacity is the basic foundation for researching airspace management problems. It's important to measure the effects of the flow management in airspace by the research of the right estimator. Therefore, this study is basically on the safe and effective purpose in airspace to develop appropriate measuring models of sector capacity that can predict. This study researches into those key factors that can construct mutually influential model in order to measure the capacity of airspace before/after resectorization, factors such as airspace sector characteristics, controller's workload, air traffic flow, and the procedure of air traffic control. At the same time, this research is applied with factor analysis methods to measure and analyze influence factors of sector capacity. This research proposes two-stage regression model to measure sector capacity, analyze sector capacity before/after resectorization, realize the utility of sector capacity, and reflect the issues of combination/decombination for sector airspace. The outcomes of factor analysis show the most evident factors that influence of airspace sector workloads are as follows: sector complex traffic, air traffic control, sector characteristics, time period, and inter-sector traffic. The regression outcomes show that per hour threshold of sector capacity before resectorization is more than per hour threshold of sector capacity after resectorization.

Introduction

In the current air traffic control environment, sectors are fixed, meaning that aircraft may be delayed or rerouted to handle excess demand. The number of specific flights rerouted depends on timing. In this way, the flow through each sector is held at or below capacity, but flights are rerouted from their preferred paths and air traffic controller's workloads rise because of the need for increased coordination, communication, and effort on rerouting aircrafts. Increased demand of international air traffic causes heavy workloads on controllers of some air routes and terminals.

The air traffic controllers play an important role for the utility of airspace capacity. It is essential to develop a measuring operational capacity of air traffic controllers to analyze the actual utility of airspace capacity and to realize the safe operations and efficiency of airspace operation.

Many established analysis tools of sector capacity used developed countries assess the operating performance of air traffic management in airspace. Increased demand of international air aviation for passengers and air cargos cause heavy workloads of controllers in some air routes and terminals. It must develop a measuring model of sector capacity to analyze actual utility of sector capacity and to realize safety operations and efficiencies of sectors. Owing to the sectors in Taipei Terminal Control area are urgent to align suitable sector boundaries when current traffic flows into or out of airports to balance workload. These flight data indicate the complexity of demand, which must be balanced with sector capacity. Therefore, this study bases on sector demands and capacities, and different sector boundaries to measure the workload of terminal airspace resectorization.

Therefore, this paper first reviews the literatures of measuring the workload of airspace sector and airspace resectorization. Then, it studies the key factors of airspace sector characteristic, air traffic, flight mix, and the procedure of air traffic control for airspace capacity, and analyzes these factors of each other influence relation to construct the model's variables of airspace sector workloads. Part three, this research also is applied with factor analysis and two-stage least square regression to measure model of air controller's workload and airspace sector capacity, and to evaluated airspace sector's capacity before/after resectorization. Part four, by means of Taipei Terminal Control Area study researches into the issues of related airspace capacity management. Part five, the conclusions and suggestions.

The Key Factors of Airspace Sector Capacity Estimation

Measuring airspace sector capacity is the basic foundation for researching airspace management problems. It's important to realize and analyze the effects of the flow management by the right estimation. Amab Majumdar(2002,2004) provide a qualitative review of research on the effect of air traffic control complexity, quality of equipment, individual difference and controller cognitive strategies to understand the relation between air traffic controller workload and theses factors. But the air traffic controller's workload is more difficult to define and even more difficult to measure (Dimitris Bertsimas, 1998; Obrad Bab,Dirk Schaefer, 2001). The measuring approach must consider operation procedure specific to individual factors that influence the air traffic controller's response to be potentially difficult. The air traffic is mainly complexity factors, including conflict position, phase of flight, height of conflict, airspace structure, speed variation, type of conflict, weather, and traffic mix. The sector factors are including sector size, sector shape, boundary location, and airway configuration.

Despite of measuring importance of airspace capacity, few extant studies have focused on airport but not on airspace. Therefore, this study proposed factor analysis to construct and measure the workload of airspace sector controllers. The airspace capacity measurement is not only relative important to catch the key factors of air controller's workloads, but also set the weight value of these key factors. In advance, using two-stage least square regression can measure the relation between airspace sector's workloads and airspace sector capacity before/after resectorization.

The Characteristics of Taipei Terminal Control Area

There are three sectors in the Taipei Terminal Control Area, including north (NR), west (WR), and east (ER) sectors as shown in figure 1. North sector controls the flight take-off/landing at Chiang Kai-Shek (CKS) international airport.West sector controls the flight departure at Sungshan airport. East sector controls the flight arrival at Sungshan airport. See figure 1 for air route map.

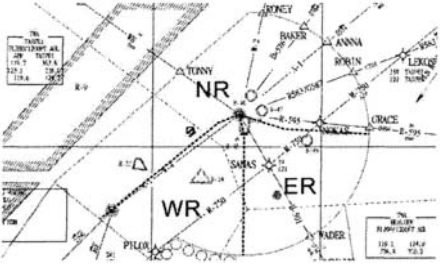


Figure 1. Three sector before resectorization of Taipei Terminal Control Area

The Factor Analysis of Airspace Sector Workload

This paper only study the static capacity of airspace sector, thus weather condition and capacity facilities are not included. This paper includes all the important influencing variables and factors of airspace sector capacity reported by Majundar (2002)and Sui-ling LI(2006). These important factors are as follows: the factors of air traffic controller workload, air traffic flow, air traffic control rule, and sector characteristics. An airspace sector controller's workload is measured by the workload time of air traffic controllers to serve flights per hour entry/exit each airspace sector. Air traffic controller's workload is defined as the time interval between the first flight entry time and the last flight exit time in each airspace sector per hour from a radar data. The air traffic flow factor is defined as the number of climbing aircraft in each airspace sector per hour, which includes descending, continuous cruising, cruise-climbing, cruise-descending, continuous descending, descend-climbing, and continuous climbing. The air traffic rule is the difference of air flight level and the difference of flight speed. Thus an airspace sector characteristics is composed of number of flight arrivals in each airspace sector per

hour, number of flight departures, and number of neighboring sectors flight entry/exit, certain variables are appropriate for particular countries.

This study collects seven days' radar data in 2005 June and July to be divided into hourly traffic data of north, west, and east three sector. This study selects above twenty five variables through factor analysis to get the five factors. Table 1 shows eigenvaluea and the percent of variance accounted for the solution. The after resectorization 1 is that combines east and north sector's boundary to be a new sector and the west sector not to change. The after resectorization 2 is that combines east and west sector's boundary to be a new sector and the north sector not to change. Table 2 shows that the outcome of factor analysis gets the five key factors, which are sector complex, air traffic control, sector characteristics, time period, and inter-sector traffic.

Table 1 Eigenvalue for Principal Components Analysis of Workload Measures

Factor	Before resectorization		After resectorization 1		After resectorization 2	
	Eigenvalue	%of Variance	Eigenvalue	%of Variance	Eigenvalue	%of Variance
1	13.242	52.966	14.361	57.443	14.589	58.355
2	3.071	12.282	2.588	10.353	2.501	10.002
3	1.931	7.724	1.831	7.325	1.610	6.438
4	1.210	4.839	1.066	4.264	1.156	4.624
5	1.022	4.088	0.927	3.706	1.047	4.187

Table 2 Varimax-Rotated Component of Workload Measures Before/After Resectorization

Factor	Before resectorization	After resectorization 1	After resectorization 2
1	Sector complex traffic	Sector complex traffic	Sector complex traffic
2	Air traffic control	Sector characteristics	Sector characteristics
3	Sector characteristics	Air traffic control	Air traffic control
4	Time period	Time period	Intersector traffic
5	Inter-sector traffic	Inter-sector traffic	Time period

Estimation Model Analysis of Airspace Sector Capacity

The numbers of flight per hour enter/exit of north and west sectors are more than east sector, at the same time, the hourly workloads of east airspace sector are less than north and west airspace sectors. Owing to low operation efficiency of ER sector, this study combines the workloads of east and north sectors or the workloads of east and west sectors to compare with the workloads of current sector size. From this point of view improves planning strategies of airspace resectorization.

Above five factor's scores of influence controller's workloads are taken into regression model to construct two-stage least square regression models in order to predict the airspace sector workload and capacity before/after resectorization as

follows: Function (1) represents the estimator of airspace sector workload before resectorization. Function (2) represents the estimator of airspace sector workload after resectorization for type 1. Function (3) represents the estimator of airspace sector workload after resectorization for type 2. Thus, Function (4) represents the estimator of airspace sector capacity before resectorization. Function (5) represents the estimator of airspace sector capacity after resectorization for type 1. Function (6) represents the estimator of airspace sector capacity after resectorization for type 2.

$$\hat{Workload} = 24.220 + 16.434Traffic + 3.834ATC - 3.052Sector + 1.512Inter\,sec\,tor - 1.333Time \quad (1)$$

(t=132.424) (89.752) (20.937) (-16.666) (8.256) (-7.28)

$R^2 = 0.953, F = 1778.54, N = 448$

$$\hat{Workload} = 32.943 + 20.169Traffic - 1.273ATC + 9.836Sector + 4.117Inter\,sec\,tor + 8.093Time \quad (2)$$

(t=77.302) (47.257) (-2.983) (23.047) (9.647) (18.963)

$R^2 = 0.907, F = 645.167, N = 336$

$$\hat{Workload} = 32.551 + 20.702Traffic + 2.621Sector + 3.262Inter\,sec\,tor - 0.73Time \quad (3)$$

(t=152.461) (96.820) (12.259) (15.248) (-3.412)

$R^2 = 0.967, F = 2442.133, N = 336$

$$\hat{Capacity} = 1.005\hat{Workload} \quad (4)$$

(t=161.282) $R^2 = 0.983, F = 26011.79, N = 448$

$$\hat{Capacity} = 0.9964\hat{Workload} \quad (5)$$

(t=149.528) $R^2 = 0.985, F = 22358.78, N = 336$

$$\hat{Capacity} = 1.003\hat{Workload} \quad (6)$$

(t=183.969) $R^2 = 0.990, F = 33844.72, N = 336$

Where

Workload : represents per hour workload for each airspace sector.
Traffic : represents the factor of per hour sector complex traffic
ATC : represents the factor of per hour air traffic control
Sector : represents the factor of per hour sector characteristics
InterSector : represents the factor of per hour inter-sector traffic
Time : represents the factor of each hour time
Capacity : represents per hour capacity for each airspace sector.

Thus, these regression models are goodness of fit. The one outcome shows that per hour workloads after resectorization is heavier than per hour workloads before resectorization. The other outcome can predict the average threshold of per hour capacity before/after resectorization. These outcomes show that the sector capacity after resectorization is no more than the sector capacity before resectorization.

The Result and Discussion

The outcomes can predict the sector capacity and show that the workloads after resectorization are heavier than the workloads of before resectorization. At the

same time, the outcomes also can show that the average threshold of per hour capacity before resectorization is more than the average threshold of per hour capacity after resectorization. These outcomes also are essential capacity references and tools under different airspace structure to improve strategies of airspace management, such as adding new sectors or combining sectors. Furthermore, this study procedure can be developed to determine how to plan airspace sector boundaries and define sector configuration without disrupting traffic flows to keep traffic density at manageable levels.

Conclusion

This study measures the capacity and workload in three sectors and proposes improving low operation efficiency of ER sector to keep traffic densities at management levels. The major findings from this study are summarized in the following: first, the most evident factors that influence of airspace sector workloads are as follows: sector complex traffic, air traffic control, sector characteristics, time period, and inter-sector traffic. Second, this two-stage least square regression can predict sector capacity and shows the threshold of per hour capacity before/after resectorization. Third, the outcomes show that the workloads of combined east and north sectors or combined east and west sectors are heavier than the workloads of current sector size. Finally, the measuring sector capacity model and resectorization are not only useful basic research for airspace management but also essential references and tools for improving planning strategies of airspace management, such as adding new sectors or combined sectors.

References

Bab Obrad, Krsti Tatjana.(2000). "Airspace Daily Operational Sectorization by Fuzzy Logic." Fuzzy Sets And System,116(1), 49-64.

Bertsimas Dimitris and Stock Patterson Sarah.(1998). "The Air Traffic Flow Management Problem With Enroute Capacities."Operation Research, 46(3) ,406-422.

Li, Sui-Ling(2006). "Measurement and Assessment the Operation Efficiency of Air Sector." TRB 2006 Annual Meeting CD-ROM, USA, Washington, D.C.

Majumdar.Arnab (2005). "En-route Sector Capacity Estimation Methodologies:An International Survey." Journal of Air Transport Management 11,375-387.

Majumdar Arnab, Ochieng Washington, Polak John.(2002). "Estimation of European Airspace Capacity from a Model of Controller Workload." Journal of Navigation, 55(2), 381-404.

Majumdar Arnab and Ochieng.W.Y. (2002). "The Factors Affecting Air Traffic Controller Workload: a Multivariate Analysis Based upon Simulation Modeling of Controller Workload." Transportation Research Record, 1788, 58-69.

Schaefer Dirk, Magill Adrian, Pirard Ben, and Aligne Florence.(2001). "Air Traffic Complexity as a Key Concept for Multi-Sector Planning."IEEE,7.E.5,1-12.

Impact of information on Housing Relocation using Analytical Hierarchy Process and Interactive GIS

P. S. Sriraj, Ph.D.[1], Mark Minor[2], and Piyushimita (vonu) Thakuriah[3], Ph.D.

[1]Research Assistant Professor, Urban Transportation Center, University of Illinois at Chicago, PH: (312) 413 7568, Fax: (312) 413 0006, email: sriraj@uic.edu.
[2]Research Assistant, Urban Transportation Center, University of Illinois at Chicago, PH: (312) 996 4820, Fax: (312) 413 0006, email: mminor2@uic.edu
[3]Associate Professor and Interim Director, Urban Transportation Center, College of Urban Planning and Public Affairs, University of Illinois at Chicago, PH: (312) 355 0447, Fax: (312) 413 0006, email: vonu-pt@uic.edu

Abstract

The problems of the urban poor, especially that of housing/shelter have been studied and documented by many researchers and social scientists. The state of affordable housing has been a cause for concern in the state of Illinois and especially in the six-county Chicago region. This concern stems from various factors. While studies have shown that a significant portion of a household's expenditure is dedicated to housing/shelter, it has also been documented that low-income families that make minimum-wage struggle to cope in the more expensive urban housing markets such as Chicago. This coupled with the inadequate affordable owner-occupied and renter homes is contributing to the complexity of the situation. In such a situation where both the supply and demand are not adequate, one needs to maximize the supply with a judicious reallocation of resources (demand).

Knowledge about the location of affordable housing, travel times to job destinations, along with familial factors such as the quality of local transit access, proximity of daycare centers, schools, and crime in a neighborhood are all viewed as important attributes in the relocation decision of a household. The lack of clear, transparent, and timely information at a disaggregate geography makes this task daunting. The research team at the Urban Transportation Center (UTC) at UIC (Sriraj, et.al 2006), developed a spatial decision support system to facilitate individuals to rank census tracts and in turn neighborhood based on their personal preferences about the various criteria mentioned in the beginning of the paragraph.

This paper extends the concept of this spatial decision support system, by describing the template for a web-enabled interactive spatial decision support system (web-SDSS). The objective of this web-SDSS is to facilitate the housing relocation process for mobility counselors using decision criteria such as transit access, affordable housing, daycare, schools, crime, and jobs. These various criteria and sub-criteria are used in a hierarchical manner framed by the Analytical Hierarchy Process (AHP) (Saaty, 1997), (Johnson, 2001).

The web-SDSS is intended for the primary use of mobility counselors assisting individuals/families seeking to relocate within the six-county metropolitan Chicago region. The web-SDSS will be designed using Mapserver 4.0 and linked with the decision-support system and the database housed at UTC.

Introduction

Housing relocation happens rather frequently in the United States, with the average family relocating once every six years. Between March 1999 and March 2000, 28 percent of the families living below the poverty level shifted their residence (Schafft, 2004). When they move, it is typically short distance and to housing of similar condition and limited economic advantage. Shumaker and Stokols (1982) suggest that while relocation or housing mobility is definitely an issue for the urban poor, there is a rationale behind the move and that the families tend to weigh the consequences of their relocation carefully before deciding to move. Various researchers have also studied the impact of transportation information on relocation choice. The objectives of this paper are two-fold: (1) to demonstrate the usefulness of the SDSS, along with the Analytical Hierarchy Process (AHP) in ranking census tracts for relocation of individuals/families in the six-county Chicagoland region, and (2) to provide the template for the web-based SDSS using Mapserver (http://mapserver.gis.umn.edu, 2005).

The current practice of relocation assistance varies from agency to agency and between the public and private sector. In the public sector, agencies such as the Chicago Housing Authority (CHA) facilitate housing relocation for Section 8 voucher recipients based on the availability of Section 8 voucher accepting units. The CHA does take into account the needs of the client but these are not comprehensive. The process however is driven predominantly by the racial composition and the poverty level of the neighborhood. In the private sector, many websites provide detailed information pertaining to the neighborhood demographics and environment, but either do not have a customized application for the low-income population groups, or/and do not provide a tool for performing tradeoffs based on the various factors deemed essential by the client. The Spatial Decision Support System (SDSS) developed at the Urban Transportation Center (UTC) at the University of Illinois at Chicago (UIC) aims to address these disadvantages embedded in the state of the practice by adopting the Analytical Hierarchy Process to facilitate pair-wise comparison of the different criteria and allowing the individuals to make tradeoffs based on informed choice (Johnson, 2001).

The objective of the paper is to provide a template for web-enabling the SDSS using Mapserver technology and to study the impact of transportation information on the relocation choices of individuals as compared to those made with access to information pertaining to housing, schools, jobs, and crime. Section 8 voucher recipients looking to relocate were recruited to participate in the exercise. A survey instrument was developed and administered to obtain the rankings for the pair-wise comparison of the criteria. This pretest was conducted on-site without the web-SDSS, and will serve as a benchmark for testing the efficiency and ease-of-use of the web-SDSS.

Spatial Decision Support System (SDSS)

SDSS can be defined as "an interactive, computer-based system designed to support a user or a group of users in achieving higher effectiveness in decision-making while solving a semi-structured spatial decision problem" (Malczewski, 1999). Spatial Decision Support Systems (SDSS) have strong potential in assisting stakeholders and decision-makers when faced with multiple alternative courses of action relating to a given problem situation.

The SDSS typically has two modules: (1) The Decision Support System (DSS) module, and (2) the GIS module. The basis for the DSS module for this version of the software is the Analytical Hierarchy Process (AHP). AHP is a ranking methodology suited for problems with a definitive hierarchical structure. In traditional AHP research, the objective is to find a

preferred alternative, from amongst a set of alternatives. Respondents are asked to assign weights to different criteria by comparing each criterion pair-wise with every other criterion. This pair-wise comparison is performed at each level of the hierarchical tree and the process is worked backwards to find the most preferred alternative (Figure 1). For this problem situation, the alternatives are the census tracts in the six-county region, and the criterion/attributes are the variables discussed above.

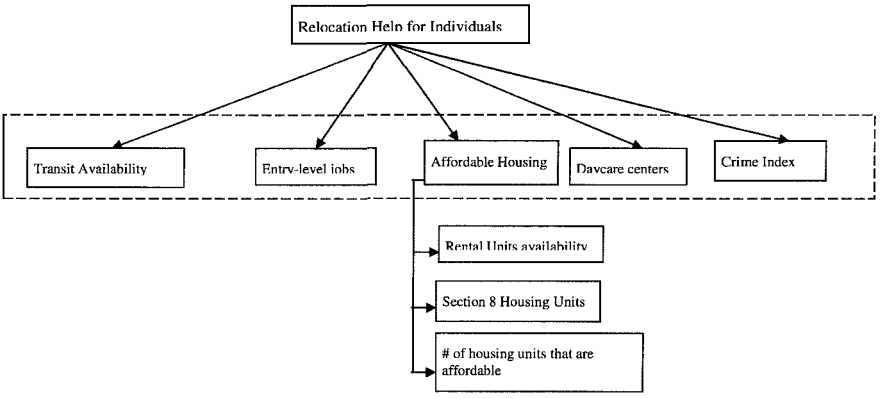


Figure 1. AHP Flow-chart for Housing Relocation

The GIS module is developed on an ESRI® platform, using the ARC-suite of tools. For this version, ARC GIS 9.0 is used.

Methodology

The spatial decision support system (SDSS) makes use of the Analytical Hierarchy Process (AHP) developed by Tom Saaty to solve multi-criteria decision-making problems. The problem situation in this instance is that of relocation of Section 8 voucher recipients. Various factors are taken into consideration when individuals/families move and AHP allows for weighing these factors against one another and facilitate tradeoffs. AHP is a systematic method for comparing a list of objectives or alternatives. The problem needs to be structured in a hierarchical manner (Figure 1) with different levels of hierarchy. The first level denotes the overall goal of this exercise. In this case, it is to provide help with relocation decision process for individuals (Johnson, M., 2001).

The main criteria included in this analysis are: transit, labor market, housing, daycare centers, and crime. The user will complete pair-wise comparison rankings for each of the main criteria. The outcome of the AHP is a prioritized ranking or weighting of each decision alternative.

The alternatives in the AHP are the census tracts and it is the objective of the process to rank the census tracts in the six-county northeastern Illinois region based on the AHP weights to reflect an individual's priorities with respect to the various criteria. The criteria are (1) Availability of transit, (2) Jobs, (3) Affordable housing, (4) Daycare centers, (5) Schools, and (6) Crime. The Housing criterion has three sub-criteria (Figure 1): the number of rental units in a tract, the number of section 8 units, and the number of affordable units in a tract. These criteria are compared pair-wise by the user and the results are aggregated at the census tract level/community area, with the help of the SDSS.

The weights for the different criteria were obtained with the help of pair-wise comparisons from the respondents with the help of the survey instrument. These weights were then input into the AHP matrix to rank the census tracts for each respondent. It is in this context that the interactive SDSS assumes significance. The process of computing the weights and the ranking of the census tracts can be executed off a server at the UTC and the results mapped for the end-user instantaneously with the help of interactive SDSS.

Interactive SDSS

The interactive SDSS is different from the more common web mapping technology that is prevalent. Web-based spatial decision support systems can be classified into three categories: (1) Server-side Web-SDSS, (2) Mixed Client-and server-side Web-SDSS, and (3) Client-side Web-SDSS (Rinner, 2003). The three categories are differentiated depending upon where the basic functions of the SDSS are presented – at the server side or the client's computer. In the case of our SDSS, the first version is envisioned to be a server-side SDSS with all the computations, the program logic, and the database housed in the server at UIC.

This paper improves on the SDSS concept by making it accessible through the World Wide Web (web), expanding the possibilities for its use. Web-enabling the SDSS will make it easier and more convenient for stakeholders to access the information and make trade-offs. It will be readily available for those looking to relocate, especially those with Section 8 vouchers. Moving the concept to the web requires setting up an Internet map server and re-writing the programs to run on the server with customized versions of the IIS, PostgresSQL, and Mapserver software packages, for web-serving, mapping, and processing, respectively.

The user is then able to weight the ranked criteria against each other, according to their personal preference for which criteria are most important. A web site will be constructed to test the concept, and to demonstrate the ability to implement such a tool with available software. Before web-enabling the methodology, it was important to ensure the appropriateness of the methodology to the given problem situation. With this objective in mind, the research team developed a survey questionnaire to get input from users regarding the pair-wise comparison of criteria for the AHP. These pair-wise comparisons allow for the census tracts to be ranked in order of importance for a given user.

AHP Pre-test and Results

The research team worked with a local advocacy group (Latinos United) to identify the pool of participants for this exercise. This led the team to the Spanish Coalition for Housing

(SCH) that is contracted by the Chicago Housing Authority (CHA) to provide Mobility counseling for families/individuals looking to relocate or are looking to get into the Section 8 voucher program. Apart from helping the tenants, the SCH also provides help to landlords in advertising the rental units, evaluating the eligibility of the units for inclusion in the Section 8 program etc.

Over a period of two days, the research team administered surveys to Section 8 voucher recipients at the Spanish Coalition for Housing. The respondents were predominantly Hispanic and were individuals looking to get their Section 8 voucher re-certified and looking to relocate. There were a total of 14 responses.

The respondents ranked Crime, Housing, and Schools as the top 3 important criteria when looking to relocate. The ideal travel time to work was 30 minutes (60%), and the maximum commute time the respondents were willing to endure was 120 minutes (60%). The maximum indicated travel time was used to create buffers around the work zip-code of the respondents, assuming an average travel speed of 20 mph.

The sample was divided into two groups: (1) those who worked and hence were spatially constrained by their work location, and (2) those who did not work, depended on public assistance, and were open to relocation anywhere in the six-county region. The results of the AHP are discussed next for group 1.

Almost 30% of the respondents indicated that they were currently working. For these people, a buffer was drawn around their work zip code by taking into account their threshold for maximum commute time. The census tracts covered within the buffer are all assumed to be suitable for relocation based on the individual's travel time preference. This first cut allows a user to eliminate tracts in the region that are too far from their work location. The pair-wise comparison weights assigned by the respondent for each criterion are then used in the AHP to rank the census tracts. From amongst the composite rankings of census tracts, the tracts with the top 5 weights are selected as candidates for relocation for that user.

The work locations for this group are aligned in and around the downtown of Chicago, and towards the western edges of the city. Not surprisingly enough, for the former category there was no difference in the top-ranked census tracts with or without transit information as part of the AHP. This can be attributed to the connectivity of the street network in the city being high as well as the fact that the transit travel times are similar to the auto travel times. This could also be reflective of the importance given to transportation in comparison to more important criteria such as crime, housing, and schools in a neighborhood. A closer look at the user response reveals that the latter is indeed the case. For those work destinations not near the downtown area, the top-ranked tracts with and without transit information, the top-ranked tracts were more dispersed amongst the north, south, and west sides of the city and suburbs.

Conclusions

The survey of low-income families/individuals and the subsequent analysis using the SDSS revealed that the process of mobility counseling can be made robust with the help of an interactive web-SDSS, and that the decision-support methodology involving AHP is appropriate for this problem situation. Based on this exercise, certain broad conclusions have been arrived at.

1. Need for pre-testing the website: The web-SDSS will have to be pre-tested by the mobility counselors to get feedback about the ease of use and to iron out any glitches in the AHP computations.
2. Need for training and education: It was evident from administering the survey, that the typical end-user for the SDSS will need a clear understanding of the user-interface and the pair-wise comparison process. This can be achieved by providing training to the agencies that provide mobility counseling. The mobility counselors currently take into account the poverty level, and to some extent the racial composition of the neighborhood, and the crime and quality of schools. A rigorous training will provide an efficient means for mobility counseling. Most respondents initially had some difficulty in understanding the repetitive nature of the survey, caused because of the pair-wise comparison of criteria. The research team is working on refining the SDSS and developing a training course such that community organizations such as the Spanish Coalition for Housing can evidence efficiencies in their operation.

References

Keels, M., et.al, Fifteen Years Later: Can Residential Mobility Programs Provide A Permanent Escape from Neighborhood Segregation, Crime, and Poverty?, Unpublished Document, Northwestern University, 2003.

Malczewski, J., GIS and Multicriteria Decision Analysis. New York: John Wiley & Sons, INC. 1999.

http://mapserver.gis.umn.edu, accessed December 2005.

Saaty, T., That is Not the Analytical Hierarchy Process: What the AHP Is and What Is Not?, Journal of Multi-Criteria Decision Analysis, Volume 6, pp. 320-339, 1997.

Schafft, K., Residential Mobility of Low Income Households and the Effects on Schools and Communities, Rural New York Initiative Policy Brief Series, Cornell University, April 2004.

Shumaker, S. A., and D. Stokols, Residential mobility as a social issues and research topic, Journal of Social Issues, Vol. 38, No.3, 1982.

Johnson, M., Decision Support for Family Relocation Decisions under the Section 8 Housing Assistance Program Using Geographic Information System and the Analytical Hierarchy Process, Journal of Housing Research, Volume 12, Issue 2, 2001.

Rinner, C., Web-based Spatial Decision Support: Status and Research Directions, Journal of Geographic Information and Decision Analysis, Vol. 7, No.1, 2003. Pp.14-31.

Sriraj, P. S., M. Minor, and P. Thakuriah, Spatial Decision Support System for low-income families: A Relocation Tool for the Chicago-land Region, Transportation Research Record (forthcoming) (2006).

Automated Identification of Traffic Patterns from Large Data Archives

Ramkumar Venkatanarayana[1], Brian L. Smith, M.ASCE[2], Michael J. Demetsky, M.ASCE[3], and Guimin Zhang[4]

[1] Dept. of Civil Engineering, University of Virginia, PO Box 400742, Charlottesville, VA 22904; PH (434) 906-5677; Fax (434) 982-2951; email: ramkumar@virginia.edu
[2] Dept. of Civil Engineering, University of Virginia, PO Box 400742, Charlottesville, VA 22904; PH (434) 243-8585; Fax (434) 982-2951; email: briansmith@virginia.edu
[3] Dept of Civil Engineering, University of Virginia, PO Box 400742, Charlottesville, VA 22904; PH (434) 982-2325; Fax: (434) 982-2951; email: mjd@virginia.edu
[4] Dept. of Civil Engineering, University of Virginia, PO Box 400742, Charlottesville, VA 22904-4742; PH (434) 906-5678; Fax (434) 982-2951; email: gz8d@virginia.edu

Abstract

The notion of traffic patterns plays a vital role in the transportation profession. Demand patterns, for example the time-series of traffic volumes experienced on highways during a.m. and p.m. peak commuter travel, are used in many applications such as performance measurement, planning and operations. However, due to limited data collection in the past for the specific purposes, there was no need for automated algorithms to identify these patterns. Now, huge data archives are increasingly available for many locations with the vast deployments of ITS systems across the nation. Taking advantage of this unique opportunity, this paper identifies and applies solutions to automatically identify traffic patterns. Clustering algorithm, logistic regression, and tree-based methods are adapted and applied. Based on application to real datasets, clustering algorithms hold high potential for use in automated traffic pattern identification.

Key words: traffic pattern, normal traffic, pattern theory, cluster, data archive, ITS

Introduction

It is widely accepted that traffic demand over a period of time is cyclical – following distinguishable patterns. For example, commuters are well conditioned to expect congested conditions during a.m. and p.m. peak travel periods. Other common patterns observed and used frequently in the transportation field include daily traffic patterns, event traffic patterns etc. The existence of these patterns is acknowledged and used in a variety of ways. Traffic operations, engineering and safety analyses,

planning, performance measurement and simulation are some of the prominent benefactors of proper identification of traffic patterns. For example, traffic operators can use the knowledge of patterns to help quickly identify any departure from "normal" (for example, an incident).

Although the concept of "normal" traffic patterns is intuitively understood and even used, no algorithms exist currently to obtain these patterns from large datasets. The widely used 'simple historic average' algorithm is heavily weighed down with incidents and "abnormal" events. A different approach is presented here. With the vast deployments of Intelligent Transportation Systems (ITS) around the nation, large datasets are often available. These rich datasets can be effectively used to identify representative normal traffic patterns at a location for any time period. In this research, pattern theory and shape analysis are explored and applied to identify traffic demand patterns. Three advanced algorithms: clustering, logistic regression, and tree-based methods, were adapted, applied and evaluated using real datasets under varied conditions. The paper concludes with a discussion on important findings and issues related to automated pattern identification in large-scale transportation data sets.

Background and Theory

This paper focuses on the application to vehicular traffic on highways, although the theory and application could be easily extended to other types of traffic data. The findings from exploratory data analysis of traffic data are first presented. This is followed by relevant background on pattern theory and shape analysis.

Findings from Exploratory Data Analysis

Several results are discerned from the literature, and from observation of the data. These findings are presented here. No two days of traffic are "exactly" the same, although some pattern (however loosely defined) usually emerges over time. These traffic patterns vary significantly from one place to another, and from one time period to another. And in some cases, there may not be even a single complete "normal" day in an archive of several months of data. The concept of "patterns" is often used in a very subjective manner, based on the expert judgment of the individual using them. If any, the context and purpose alone define the basis of such classifications. This purpose translates quantitatively as the scale of interest (in data or features).

"Abnormal traffic" is defined reasonably well, as that related to incidents, weather, or other abnormal events. However, manually identifying these incidents and removing the related traffic data is time consuming and unreliable, as many incidents are not even recorded (Christens, 2003). Obtaining normal traffic patterns by removing "all" the incidents is therefore impractical. Direct pattern identification is needed.

Pattern Theory

"Pattern recognition" is a well-known and developed subject, where new data is compared to some expected or known characteristics. In pattern identification, the pattern characteristics are not known yet, and pattern theory needs to be employed. Pattern theory may formally be defined as a way to apply mathematical formalisms

towards the understanding of patterns (Grenander, 1996). The concept of pattern presupposes the existence and understandability of a structure. Pattern theory helps to uncover these structures, and develop algorithms towards finding them. The rules for the typical structure, and the rules for allowed variability within this structure are discovered. In essence, pattern theory provides the fundamental background for defining patterns. A pattern is defined through a set of meaningful primitives (features) and rules for manipulating them into formal structures. Shape analysis is employed here to identify features of importance in the traffic domain.

Shape Analysis

Shape analysis is a broad topic that extends from digital data collection of shapes to their recognition, and even their classification (Costa and Cesar, 2000). Shape analysis is studied here for the proper selection of good shape features. A shape is defined as all the geometrical information that remains when location, scale and rotational effects are filtered out from an object (Dryden and Mardia, 1998). It is claimed that the process of identifying proper features is even more difficult and important that the classification itself (Costa and Cesar, 2000). Everitt (1993) further acknowledges that the initial choice of variables is itself a categorization of the data for which there are limited statistical and mathematical guidelines. Some important properties for the proper selection of features include (1) highly discriminative features, (2) low correlation among selected features, and (3) small set of features. Two good features applicable and selected for this research are:

- Variance: The statistical sample variance of the data within a time period. During incidents, traffic data is known to exhibit high fluctuations in volume over time.
- Autocovariance: At a pre-defined lag, a time series data can be compared with itself. The sample autocovariance function at lag h is defined as follows:

$$\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}\left(x_{t+|h|} - \bar{x}\right)\left(x_t - \bar{x}\right), -n < h < n.$$

Candidate Solutions

Based on the theory presented above, three advanced computing algorithms were adapted and applied to real traffic data, to identify "normal traffic patterns": Clustering, Logistic regression, and Trees-based method. These are explained below.

Clustering

Clustering algorithms organize similar entities together. The main advantage of a data-clustering algorithm is that the patterns need not be known prior to analyses i.e. they work in an unsupervised environment. K-means clustering, employed in this study, is a well-developed algorithm. At first, a set of pre-defined or some random days of traffic data are selected as the initial cluster centroids. The distance for each day of traffic in the entire data set from all these cluster centroids is evaluated. A day is attributed to the cluster to whose centroid the day is closest. The new centroids are then calculated based on all the days within each cluster. The clusters and the cluster centroids are iteratively adjusted until the distances are minimum across the entire data set. Alternately, the program is halted after a pre-defined maximum number of iterations are performed. Along with the identification of proper features, a distance

measure is also required for a clustering algorithm. The three final distance measures (along with features) used in the current study are:

- Euclidean distance of the direct traffic data series,
- Difference in the variance between the two series of data,
- Difference in the auto covariance of the two series of data, at different lags

Logistic Regression

Logistic regression allows identification of the most important variables, and is therefore considered a very good method for data analysis and inference. At each stage, modeling can be performed by eliminating those variables that exhibit insignificant z-scores. For this reason, logistic regression is considered a parsimonious model. The model is based on the apriori probabilities associated with the features and the classes. In case of a two-class problem, the data is associated to that class which has the higher maximum likelihood. The current research study is considered as a two-class problem, to partition traffic data as either a normal or an abnormal pattern.

Tree-based Method

Tree-based methods partition the data points into a series of rectangles. Recursive partitioning, used in this research, repeatedly splits the space into two regions based on some condition. The initial tree structure is generated by evaluating an input condition on one variable. The next condition is then evaluated and this branching continues till the impurity function is minimized. Misclassification rate method for minimizing node impurity is selected for this study. Some of the documented disadvantages of the trees model are instability due to their hierarchical nature, and their accuracy is often low. One of the main advantages of trees is its interpretability (Hastie et al, 2001).

The above three algorithms are evaluated with two types of metrics – (1) Quantitative metrics: Accuracy/error rates, and (2) Qualitative metrics: Ease of setting up the algorithm, its convergence, and speed.

Case Studies

Traffic data for this study originated from the Virginia Department of Transportation's Traffic Monitoring Stations (TMS). 15-minute link volumes were used. A link is defined as the representation of all the lanes at one location on the road, along a given direction. To analyze a diverse set of traffic conditions, four datasets were used. Datasets 1 and 3 (links 90272, 150012) are from urban locations in different parts of the state. Dataset 4 came from a rural area (link 190093). Dataset 2 came from a different time period (2002) for link 90272. The other three datasets (1, 3 and 4) belonged to the year 2004.

The traffic data was first manually classified into three classes – normal, abnormal and ambiguous. It is supposed that even an expert could classify these ambiguous days of data as either normal or abnormal, depending on the purpose of classification. For stricter training of the algorithms, the ambiguous data were also

treated as abnormal dataset. The datasets were divided into five time frames of the day. 0:00-5:30; 5:30-9:30; 9:30-14:30; 14:30-19:30; and 19:30-24:00. These divisions were based on approximately where the data undergoes significant shape changes. Based on preliminary analyses, a lag of 10-15 minutes was selected for the autocovariance function for optimum results. For each combination of dataset, time frame of the day, and distance measure, the cluster algorithm was run 5 times to check convergence of the results. In case of the logistic regression model, any variables that do not contribute significantly for classification can be removed. This tuning was performed three times so that the best logistic model results are available for comparison with the other models.

Results

For the most part, clustering algorithms converged by consistently grouping between 40-60% of all the days densely into 1 to 4 clusters (out of a possible 10 selected in the beginning). The other two models had significant convergence issues. One main reason is that the training was not performed for each data set separately. The test data associated with the training dataset (dataset 1) performed quite well, as presented in Table 1.

TABLE 1 Accuracy and Error Rates of All Models

		% Abnormal in 'normal cluster'				% Ambiguous in 'normal cluster'				Overall % Error	
dataset	Time of Day	cluster-euclidean	cluster-variance	cluster-autocovariance	Actuals, in the data	cluster-euclidean	cluster-variance	cluster-autocovariance	Actuals, in the data	Logistic Regression	Trees Model
1	0:00-5:30	3.1	1.6	0.8	10.3	7	3.2	5.5	7.1	13.9	25
2		17.6	0	11.3	16.1	2.3	7.1	0	1.8		
3		2	3.6	7.2	6.2	1.5	3.1	5.2	3.1		
4		0.8	1.9	1.9	8	1.5	0	0	1.6		
1	5:30-9:30	32.1	61.6	51.6	51.6	18.2	14.7	15.9	15.9	19.4	33.3
2		7.3	69.1	21.4	41.1	7.1	7.7	17.9	10.7		
3		0	6.3	5.1	15.4	3.5	2.1	2.5	3.1		
4		0.8	14.8	16.7	16.8	5.1	3.6	1.7	4		
1	9:30-14:30	2.1	16.4	18.6	11.9	5.6	7.7	14.9	9.5	22.2	19.4
2		0	22	7.4	8.9	4.2	30	21	17.9		
3		4	4.6	4.1	15.4	4	2.7	4.1	3.1		
4		3.2	13.3	15.8	9.6	6.1	12.9	6.7	8		
1	14:30-19:30	4.8	12	12.6	12.7	5.6	8.1	7.3	6.3	16.7	19.4
2		1.6	6.1	19	12.5	3.6	33.4	3.5	21.4		
3		7.9	13.2	9.3	14.6	7.1	5.7	6	6.9		
4		4.7	17.8	10.9	18.4	3.3	3.7	3.6	4.8		
1	19:30-24:00	2.8	2.4	5.4	7.9	1.4	7.9	4.7	6.3	30.6	13.9
2		8.8	12.8	14.4	14.3	23.2	34.7	23.5	25		
3		6.6	0.7	1.6	7.7	5.5	6.1	6.6	5.4		
4		0.8	7.2	6.6	15.2	0.8	1.8	3.2	4		

- Accuracy/error rates: Logistic regression and tree models did well with the test dataset associated with the training dataset. About 80-85% of the data was classified correctly. It is interesting to note that this classification rate was high even when the data was quite variable, as in 5:30-9:30 time frame. But beyond dataset 1, their performance was very bad (about 10-20%) and hence these other results are not analyzed fully.

- On the whole, Euclidean distance function performed slightly better than the other two features. There are a few combinations of time frames where even variance and autocovariance functions also seem promising (early AM 0:00-5:30 time frame). All the algorithms performed poorly when the data itself actually had a huge percentage of ambiguous and/or abnormal data (dataset 2, and dataset 1 for time frame 5:30-9:30).
- Speed was not an issue for any of the algorithms. However, convergence and transparency are not high for any of these algorithms either.
- The accuracy of obtaining 'normal pattern' through the cluster algorithms is almost always better than considering the entire dataset directly (due to reduction in the false positives in the final clusters, compared to the original data). The entire datasets are used in obtaining historic average, a popular method for obtaining traffic pattern currently. Even when major incidents are removed, a ambiguous data still weigh down the historic average considerably.

Conclusions

Pattern theory fundamentals have been identified and applied to the transportation domain successfully. Some important traffic data features, distance measures, and trade off measures for algorithms were identified to obtain normal traffic patterns. Three important conclusions emerge from this study. Firstly, supervised algorithms are very accurate with the original dataset (used in training), even when the data itself exhibited an unnatural trend. However, this extra tuning created convergence problems while adapting to other data sets. On the other hand, unsupervised clustering could not cope with the unnatural trend, but performed much better with unknown datasets and time periods.

Secondly, application of clustering algorithm still results in many clusters that cannot be semantically understood readily. Thirdly, partitioning a traffic day into different time frames is useful in cases where parts of a day were normal and other parts were not. The results are promising to further improve the scientific basis and to develop unsupervised algorithms to identify normal traffic patterns from large traffic data archives.

References

1. Christens, P. (March 31, 2003) *Statistical Modeling of Traffic Safety Development*. PhD Dissertation, Danish Transport Research Center and Informatics and Mathematical Modeling
2. Grenander, U. (1996) *Elements of Pattern Theory*. Johns Hopkins University Press, Baltimore
3. Costa, L. da F., and R. M. Cesar Jr. (2000) *Shape analysis and Classification: Theory and Practice*. CRC Press, Boca Raton, Florida
4. Dryden, I. L. and K. V. Mardia. (1998) *Statistical Shape Analysis*. John Wiley & Sons, Ltd. (UK)
5. Everitt, B S. (1993) *Cluster Analysis*. 3rd Ed, Edward Arnold, London
6. Hastie, T., R. Tibshirani, and J. Friedman. 2001. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, Canada.

Development of Spatial Data Tools to Manage Transportation Networks

Mukund Dangeti[1], Srinivas S. Pulugurtha[2], and Shashi S. Nambisan[3]

[1]Transportation Research Center, University of Nevada, Las Vegas, 4505 Maryland Parkway, Las Vegas, NV 89154-4007; Tel. (702) 895-1362; Fax (702) 895-4401; Email: mukund@trc.unlv.edu

[2]Department of Civil Engineering, University of North Carolina, Charlotte, 9201 University City Boulevard, Charlotte, NC 28223-0001; Tel. (704) 687-6660; Fax (704) 687-6953; Email: sspulugu@uncc.edu

[3]Transportation Research Center, University of Nevada, Las Vegas, 4505 Maryland Parkway, Las Vegas, NV 89154-4007; Tel. (702) 895-1325; Fax (702) 895-4401; Email: shashi@ce.unlv.edu

Abstract

This paper presents a summary of a process used to develop a spatial data system to help manage transportation networks. Issues related to the system architecture, and identification and integration of software and hardware elements are addressed. Commercial off-the-shelf software and hardware, along with customized interfaces are used to develop the system. Hardware considered includes portable digital assistants, Tablet PC, and Laptop. Key aspects that were considered in selecting the hardware include the ease with which they can be used in the field, their durability, effectiveness, and portability. Compatibility and ability to integrate with other application software were also critical considerations. The software is selected based on the ease with which they can be used and integrated with web authoring software, thus enhancing collection, processing, and dissemination of data. In addition, the compatibility of the software with various data formats (i.e., prior versions of databases), and the potential for such compatibility to continue into the future is also important. Software considered and tested includes Autodesk MapGuide, ESRI ArcIMS, ESRI ArcPad, and Terrasync software. Using selected hardware and software, a pilot project is used to demonstrate the use of the system to assimilate, manage, process, and communicate transportation infrastructure related data.

Introduction

Transportation networks are complex systems that are owned and operated by a diverse group of organizations. The management of such systems requires substantial amounts of data to be monitored, acquired, managed, analyzed, modeled, and maintained. In general, data currently collected by typical transportation agency can be loosely classified into three different categories: planning, engineering, and operational. Such data relate to the network, vehicles/users, and the results of the network related activities. They include on-network characteristics such as supply or demand indicators at node or link levels, infrastructure condition related information, and, safety and environment related attributes resulting from the interactions between the users and the network. Other data of interest include off-network characteristics such as proximate population and demographic characteristics, and land use characteristics.

In transportation engineering, near-real time data updates play a key role in the functioning of the system. Many a process ranging from the design of a simple actuated

signal to a complex emergency medical response system depend on near-real time data from the field. The process involved in updating the transportation system databases with near-real time data depends on the type of data collected and also the quantity of data to be updated. The capabilities afforded by technologies such as the global positioning system (GPS) and geographic information systems (GIS) offer significant potential for collecting data pertinent to the spatial and temporal characteristics of transportation networks. In addition, software such as CORBA has been used in various transportation projects by organizations such as FedEx, the American Automobile Association (AAA), and American Airlines (Hu, 2002). Quiroga, et al, (2001) describe the architecture of a prototype GIS/GPS Internet platform for the management of utility data that takes into consideration both utility data and roadway network characteristics. The model includes a GPS/GIS-based prototype to represent utility facilities located within a highway right-of-way (ROW) and associated attribute data such as ownership, purpose, size, type, and shared/multiple uses.

This paper summarizes the development of a decision support system to manage transportation system data. This involves the identification, comparison, and integration of software and hardware elements. The system helps assimilate, analyze, manage, and communicate near real-time and static data. It uses commercial off-the-shelf software and hardware, along with customized interfaces, GIS and GPS tools. Further details of the system development process along with an illustration of its application follow.

Methodology

An overview of the system architecture is presented in this section. The system architecture comprises the following steps.

1) Selection of suitable hardware and software,
2) Development of the applications or tools for data collection, and
3) Data processing, analysis and archiving

1) Software and Hardware Process

The first step is the selection of suitable hardware and software to address the needs of the system. These, in general, depend on the functional need of the users. As an example, hardware such as loop detectors and CCTV cameras are used to collect traffic data in near-real time whereas data related to traffic work orders may be collected using handheld units such as portable digital assistants (PDAs), Laptop computers, and Tablet PCs. Key aspects that need to be considered in selecting the hardware include the ease with which they can be used in the field, their durability, effectiveness, and portability. Compatibility and ability to integrate with other application software are also critical considerations.

The software used can be classified according to the purpose they serve. They can be used for data collection and analysis, or for displaying the data on the World Wide Web for public/private use. In addition, some of the additional capabilities afforded by the internet mapping software include spatial queries and interactive displays. One of the requirements of the system developed was for it to facilitate web-based analyses, services, and communications. Thus, the ease with which the software can be used and integrated with web authoring software is important for collection, processing, and dissemination of data. The software selected should also be compatible with instruments or hardware selected for data collection. Yet another consideration is the backward compatibility of the software with various data formats (i.e., prior versions of databases), and the potential for such compatibility to continue into the future.

2) Development of Applications

The application development process is typically based on a modular approach. Different application modules are developed to serve and achieve different purposes / goals. Listed below are two major purposes to be served by the applications specified for the system.

1. To be used in the field by field personnel either to collect new data or edit existing data. Applications for this process can be developed using database software such as ORACLE, Access, and Excel or using ESRI ArcPad and ESRI ArcPad Application Builder Software. Such applications should automate the data collection process to the maximum possible extent. Transferring data from the field to the office and the database server could be done either in near real-time (in a few minutes) using wireless technologies or by physically connecting the instrument after returning to the lab.
2. To be used in the office by staff who interface with internal and external customers to record and track works-orders (i.e., complaints and service requests). Applications developed using a web authoring software and Active Server Pages (ASP) are generally used to record complaints and service requests.

3) Data Process, Analysis and Archiving

This step includes data identification, update, storage, distribution, and archiving. This task begins with the identification of various types of data needed for the required analyses, their sources, scales, formats, and other attributes such as the required frequency of updates, archival needs, etc. The exact data incorporated into the system depends on the desired accuracy, speed, and the available resources such as personnel, time, and budget. In some cases when the desired data are not readily available (based on the aforementioned criteria), surrogate data may be used. Based on the data incorporated into the system and their characteristics, specific data linkages are developed to support the analyses. Next, methods are devised to link outputs of one process that are needed as inputs to other processes / models, and to communicate the results of the analyses. Data archiving is driven by considerations such as access to the information in the future including the types of users, frequency of use, the need for raw data versus the appropriateness of synthesized data, and the period for which the archive needs to be maintained for each data element.

Pilot Project

The Las Vegas metropolitan area in Clark County, Nevada has been and continues to be one of the fastest growing urbanized areas in the country. As the Las Vegas valley's population and the number of road miles continue to grow, so do the traffic volumes on local roadways. The state and local government agencies responsible for managing the transportation infrastructure in the area have significantly expanded the road network in order to accommodate the population and economic growth. For roads under the jurisdiction of Clark County, the Clark County Department of Public Works - Traffic Management Division (CCDPW-TMD) is responsible for conducting the studies to determine what traffic control devices are needed, as well as designing, overseeing the installation and maintaining these control devices. In addition, the CCDPW-TMD also maintains streetlights, and supports the investigation of the fatal and serious injury vehicle accidents within unincorporated areas of Clark County. Such activities require access to information related to the transportation system and factors influencing transportation demand.

The database maintained by CCDPW-TMD can be used as an example for the transportation database which needs to be maintained by a local governmental agency with jurisdictional responsibility over road networks. This central database contains several component sections such as right-of-way inventory, roadway and pavement characteristics, traffic characteristics, and, roadway and roadside infrastructure inventory. The requirements

of the system presented herein were to assist CCDPW-TMD with their field data collection and work order management activities.

Hardware and Software

Based on the requirements identified by CCDPW, the following hardware elements were identified and evaluated for use by the agency: (1) Personal Digital Assistant (PDA), (2) GPS Handheld, (3) Tablet PC, and, (4) Laptop. All four different instruments have various advantages and disadvantages when compared to one another. As an example, the expected battery life (which depends on various factors such as the ambient temperature) of the Tablet PC and Laptop is significantly lower than that of the PDA and GPS handheld. The Tablet PC and Laptop are also heavy when compared to PDA and GPS handheld. However, the Tablet PC and Laptop are easier and more comfortable to use with graphical applications because of their bigger screen size and better memory features.

One of the vital factors for the selection of the software is the compatibility consideration for developing Internet mapping applications in an easy-to-use web browser – for example, to integrate local data sources with Internet data sources to display, query, and analyze spatial information. The two widely used software programs that are germane to this research are ESRI ArcIMS (ESRI, 2005) and Autodesk MapGuide 6.0 (Autodesk, 2005). These are also the two programs that are widely used in the field of internet mapping applications. Both Autodesk and ESRI provide technologies to view and query the data on the web, as required by most spatial data users. However, the powerful capabilities, scalability features, and ability to work with ESRI products made the MapGuide a better choice of implementing a web-based system. The second criterion for the selection of software, as required by the CCDPW-TMD, is the simplicity and ease of use of the software in field instruments like PDA and GPS handheld. The two software programs which were used in this study for collection of data in the field are ArcPad software (3) and Terrasync Software (Trimble, 2005). Both these products exhibit a wide range of capabilities. However, developing a data collection interface appears to be easier using ESRI ArcPad software. Both the programs were tested in the field to determine their efficiency in the process of data collection, by using them on different hardware.

Development of Applications

Two different types of applications were developed. They are: 1) Application for field data collecting personnel, and 2) Application for the office-based staff for work-order management.

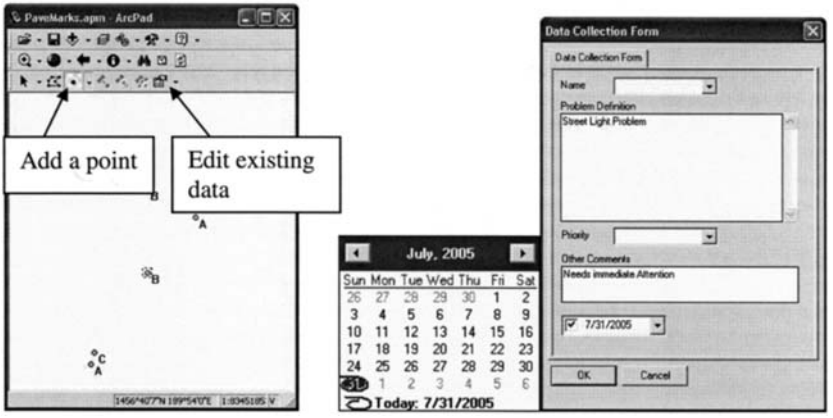
Application for the Field Data Collecting Personnel

In general, field data collection personnel may not be as technically knowledgeable or aware of using technologies like GIS/GPS as might be engineers, planners, and analysts. The development of the GIS tool was done primarily to ease the efforts of the data collection for the field personnel. The application and interface for data collection was developed using ArcPad Application Builder Software. The user needs to enter data such as the name of the individual collecting the data, nature of the problem / problem description, priority and other comments. Details such as problem location (from the GPS) and date and time of data entry are obtained from the settings on the hardware and automatically stored in the database. The option of automatic storage of the time and date of data collection can be changed depending on the need of the user. If the date and time fields do not require the interaction of the individual collecting the data, i.e. the individual should not be allowed to change the date and time, then the fields can be protected in such a way that the user cannot overwrite the default values. Figure 1(a) shows the ArcPad interface that is displayed on the screen of the handheld

instrument to add a new data point or to edit an existing one. Figure 1(b) shows the interface used by the personnel during data collection to enter data attributes for a selected item.

Application for Use by Office-based Staff for Work-order Management

The CCDPW office-based personnel (in addition to the field personnel) interact with internal and external customers to record and track work-orders. Such interactions include recording complaints and service requests – for example, malfunctioning of the streetlights, traffic signals or other traffic control devices. This may be done in several ways such as through the phone, internet, or other media. At present, the staff member fills out forms (paper and pen based system) to record the work order, and forwards it to the corresponding group for necessary actions. The objective of the system developed was to help improve the efficiency and effectiveness of this process, to track the work order, and the assignment of tasks to field crew. Applications were developed for the office-based personnel using Autodesk MapGuide 6.5 and Active Server Pages (ASP). Figures 2 shows the interface developed for recording complaint data by the office personnel by clicking on a point representing the exact location of the reported problem.



1(a) Basic ArcPad Window. 1(b) Interface for Field Data Collectors (New Data).

Figure 1. Application of Collecting Data in the Field.

Once the data have been submitted, it can be viewed over the intranet / internet. A priority tag ("Emergency", "Urgent", "Routine", or "Special") is assigned based on the importance of the work order. The current practice is for the staff to file the complaint paperwork at the end of each day, and forward the same to the corresponding departments whereas the new method can deliver the complaint to the necessary department in near real-time. Thus, the proposed methodology significantly reduces the use of paper while improving the speed, accuracy, efficiency, and effectiveness of the data collection and processing steps.

Summary and Conclusions

The development of spatial data tools to help manage and facilitate transportation system data has been summarized in this paper. Key issues related to the system architecture, software and hardware elements and the use of GIS and GPS based tools were addressed. The system used off-the-shelf software and hardware, and also customized interfaces. A small pilot project focusing on work order management was presented to demonstrate the use of the

system to assimilate, manage, analyze, and communicate transportation infrastructure data. The use of such a system significantly improved the speed, accuracy, efficiency, and effectiveness of the data collection and processing steps.

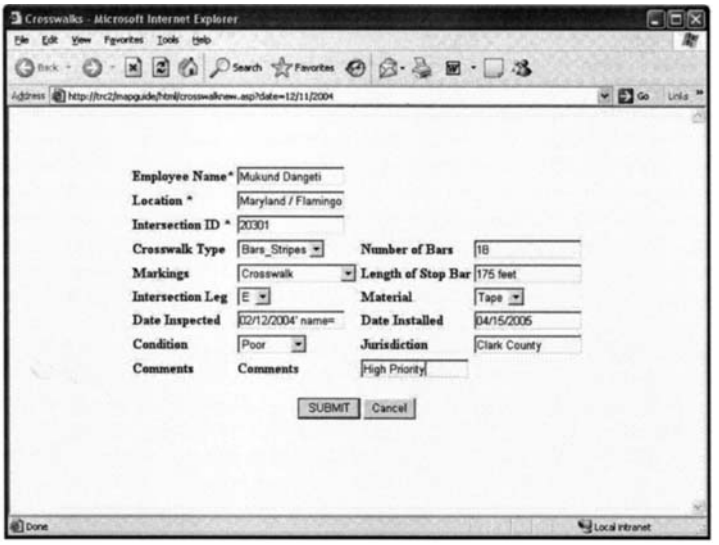


Figure 2. Complaint Recording Form.

Acknowledgments

The authors thank the Clark County Department of Public Works (CCDPW) and the University of Nevada, Las Vegas Applied Research Initiatives Program for their financial support. The authors also acknowledge N. T. Rajah and Tammi Tiger of CCDPW for their technical support and guidance.

References

Autodesk (2005). Autodesk MapGuide Author and Server Software. Website: http://usa.autodesk.com/adsk/servlet/index?siteID=123112&id=3134594, Accessed on 27th December, 2005.

Environmental Systems Research Institute - ESRI (2005). Product Information for ArcGIS and ArcPad Software. Website: http://www.esri.com/products.html, Accessed on 27th December, 2005.

Hu, T.Y. (2002). CORBA-based distributed Geographic Information Systems for Transportation for Emergency Medical Services. Transportation Research Record, 1800 (27-34), Transportation Research Board 81st Annual Meeting, January 13-17, Washington, D.C.

Quiroga, C., Ellis, C.D., and Shin, S.Y (2001) An Integrated GPS/GIS/Internet Platform for Managing Utilities Along Highway Corridors. Transportation Research Board 80th Annual Meeting, January 7-11, Washington, D.C.

Trimble (2005). Product Information for Trimble GeoXT Handheld and Trimble Terrasync Software. , Trimble Website:, http://www.trimble.com/terrasync.shtml, Accessed on 27th December 2005.

An MPEC Model for the Optimal Contraflow Operation Problem with User Equilibrium Constraints

Q. Meng[1], H. L. Khoo[2], Y. K. Huang[3], R. L. Cheu[4]

[1]Department of Civil Engineering, National University of Singapore, Singapore 117576; TEL (65) 6874 5494; FAX (65) 6779 1635; email: cvemq@nus.edu.sg
[2] Department of Civil Engineering, National University of Singapore, Singapore 117576; TEL (65) 6874 5035; FAX (65) 6779 1635; email: g0403653@nus.edu.sg
[3] Department of Civil Engineering, National University of Singapore, Singapore 117576; TEL (65) 6874 5035; FAX (65) 6779 1635; email: g0300289@nus.edu.sg
[4] Department of Civil Engineering, National University of Singapore, Singapore 117576; TEL (65) 6874 2153; FAX (65) 6779 1635; email: cvecrl@nus.edu.sg

Abstract

This paper proposes a mathematical program with equilibrium constraints (MPEC) model for the optimal contraflow operation problem that intends to determine the number of lanes of the candidate links that should be reversed their travel directions during the peak hour, while taking into account the user equilibrium (UE) behavior of drivers in route choice, so as to minimize the total travel time over a study area. The MPEC model includes a parametric variational inequality (VI) that reflects the reaction resulting from drivers with respect to a contraflow operation solution in the case of drivers follow the UE principle in route choice. To solve the MPEC model, this paper develops a genetic algorithm incorporating with a user equilibrium traffic assignment procedure. Finally, a numerical example is carried out to evaluate the proposed model and solution method.

Introduction

Contraflow operation is a practical but a cost effective traffic management scheme in which travel direction on some lanes of a link (i.e., a directed road segment) are reversed during the peak hour in order to alleviate the traffic congestion of some area such as a central business district (CBD). It has already been successfully implemented in the megacities of developing countries including China and Malaysia. This is because during the peak hour, traffic flow on both directions of some links, which actually form the traffic bottlenecks of some congested area, is unbalanced. In other words, one direction of the link tends to have heavy traffic flow while the opposite one shows light traffic demand. Before the implementation, one has to determine which candidate link should be chosen and the number of its lanes that should be reversed travel directions, so as to minimize the total travel time over a study area. For short, this problem is referred to as the optimal contraflow operation problem. Any contraflow operation will surely change the topological structure of a transportation network, and it will result in a redistribution of traffic flow pattern. We are not able to control the drivers' responses to the change of a network because

drivers are independent decision-makers in chosing a route between an origin-destination (O-D). Nevertheless, the user equilibrium (UE) principle can describe the behavior of drivers in route choice (see, Sheffi, 1985; Patriksson, 1994). The purpose of this paper is to study the optimal contraflow operation problem with UE constraints.

Studies on the optimal contraflow operation problem have been received limited attention. Drezner and Wesolowsky (1997) made an initial attempt on the determination of an optimum configuration of one-way and two-way routes in the case of flow-independent link travel time patterns. However, they do not consider the route choice behavior of drivers. Besides, Thedoulou and Wolshon (2004) used a trial-and-error approach based on microscopic traffic simulation software to investigate the contraflow evacuation on the westbound I-10 out of the City of New Orleans. Nevertheless, they did not provide a systematic methodology for solving the optimal contraflow operation problem. Having considered drivers' UE behavior, the contraflow operation problem becomes a new variation of the network design problems (NDPs), however it is quite distinct from the classical discrete network design problem (NDP), continuous network design problem (CNDP) and mixed network design problem (MNDP) (see, Yang and Bell, 1998). To our best knowledge, the optimal contraflow operation problem with UE constraints is a novel research issue in the subject of the transportation network optimization problems with user equilibrium constraints.

This paper will develop an MPEC model for the optimal contraflow operation problem with UE constraints. In the proposed MPEC model, a feasible contraflow operation solution is represented by a vector of the binary (0-1) decision variable indicates whether or not a lane of the candidate link should be reversed the travel direction. The objective is to minimize the total travel time over a study area involving the UE link flow pattern, an implicit function of a contraflow operation solution. One of the constraint of the MPEC model is a parametric VI that characterize the UE conditions induced by a given contraflow operation solution. It should be pointed out that the MPEC as a powerful modeling approach has been used in transportation, economics and management (see, Luo et al., 1996). As the proposed MPEC model is a discrete optimization problem with the time complexity of NP-hard time, this paper will employ the genetic algorithm (GA), one of the well-known meta-heuristic methods, for solving it. The GA will be incorporated with a UE traffic assignment procedure in evaluating the fitness value of the chromosome representing a contraflow operation solution, which is casted as the hybrid GA-UE throughout this paper.

An MPEC Model

Let $G=(N,A)$ denote a transportation network, where $N=\{1,2,\cdots,i,\cdots\}$, the set of nodes, and $A=\{(i,j)|i,j\in N\}$, the set of directed links. In addition, the sets of origins and destinations are denoted by R and S, respectively, and $R,S\subseteq N$.

Let P_{rs} be the set of all the feasible paths between an O-D pair rs, q_{rs} be the number of trips made between the O-D pair during the peak hour, f_k^{rs} be the traffic flow on path $k \in P_{rs}$ and v_{ij} be the traffic flow on link $(i,j) \in A$. Let $\mathbf{v}$ be the vector of all the link flows, i.e., $\mathbf{v} = (\cdots, v_{ij}, \cdots)$ and Ω be the set of all the feasible link flows, we then have

$$\Omega = \left\{ \mathbf{v} \left| \sum_{k \in P_{rs}} f_k^{rs} = q_{rs}, r \in R, s \in S; v_{ij} = \sum_{r \in R} \sum_{s \in S} \sum_{k \in P_{rs}} f_k^{rs} \delta_{ij,k}^{rs}, (i,j) \in A; f_k^{rs} \geq 0, k \in P_{rs} \right. \right\} \quad (1)$$

where $\delta_{ij,k}^{rs}$ is equal to 1 if link (i,j) is on path k between O-D pair rs and 0 otherwise.

Each link $(i,j) \in A$ composes several lanes numbered by integer numbers $1, \cdots, l_{ij}$, where l_{ij} is the number of total lanes in the link. Let c_{ij}^k be the capacity of lane $k \in \{1, \cdots, l_{ij}\}$ of link (i,j), and C_{ij} be the capacity of the link, their relation can be expressed by

$$C_{ij} = \sum_{k=1}^{l_{ij}} c_{ij}^k, (i,j) \in A \quad (2)$$

For the sake of presentation, let $\mathbf{C}$ be the vector of all the link capacities, i.e., $\mathbf{C} = (\cdots, C_{ij}, \cdots)$, where $(i,j) \in A$. With respect to each link (i,j), there is a nonnegative and continuously differentiable link travel time which is a function of link flows on all the links due to link flow interactions, denoted by $t_{ij}(\mathbf{v}, \mathbf{C})$. The vector of all these link flows can be represented by a vector: $\mathbf{t}(\mathbf{v}, \mathbf{C}) = (\cdots, t_{ij}(\mathbf{v}, \mathbf{C}), \cdots)$. It is assumed that vector function $\mathbf{t}(\mathbf{v}, \mathbf{C})$ is strictly monotone on the set Ω (see, Patriksson, 1994).

Let A_1 be the set of all the links in the study area, i.e., $A_1 \subseteq A$, and A_2 be the set of candidate links for implementing the contraflow operation. Note that the candidate links are predetermined by the land transport authority according to the historical traffic conditions. Let $\tilde{A}_2$ be set of those links whose opposite sides are in set A_2, namely,

$$\tilde{A}_2 = \{(i,j) | (j,i) \in A_2\} \quad (3)$$

Without loss of generality, it is further assumed the opposite side of a link in set A_2 is not a candidate link, i.e.,

$$A_2 \cap \tilde{A}_2 = \phi \quad (4)$$

To formulate the preceding contraflow operation problem with UE constraints, define the binary decision variables as follows:

$$x_{ij}^k = \begin{cases} 1, \text{ if lane } k \text{ of link } (i,j) \text{ is chosen to reverse} \\ \quad \text{travel direction} \\ 0, \text{otherwise} \end{cases} \quad ,(i,j) \in A_2 \text{ and } k = 1,2,\cdots,l_{ij} \quad (5)$$

Let $\mathbf{x} = (\cdots, x_{ij}^k, \cdots)$ be the vector of all the above decision variables. The optimal contraflow operation problem with UE constraints can be formulated by the MPEC model:

$$\min F(\mathbf{x}) = \sum_{(i,j)\in A_1} v_{ij}(\mathbf{x}) t_{ij}(\mathbf{v}(\mathbf{x}), \mathbf{C}(\mathbf{x})) \quad (6)$$

subject to

$$\mathbf{t}(\mathbf{v}(\mathbf{x}), \mathbf{C}(\mathbf{x}))(\mathbf{v}(\mathbf{x}) - \mathbf{v})^T \geq 0, \ \forall \mathbf{v} \in \Omega \quad (7)$$

$$x_{ij}^k = 0 \text{ or } 1, (i,j) \in A_2, k = 1, \cdots, l_{ij} \quad (8)$$

where $\mathbf{v}(\mathbf{x})$ and $\mathbf{C}(\mathbf{x})$ are the UE link flow pattern and link flow capacity associated with a contraflow operation solution $\mathbf{x}$. Elements of vector $\mathbf{C}(\mathbf{x})$ have the closed expressions:

$$C_{ij}(\mathbf{x}) = \begin{cases} C_{ij}, (i,j) \in A \setminus (A_2 \cup \tilde{A}_2) \\ \sum_{k=1}^{l_{ij}} c_{ij}^k (1 - x_{ij}^k), (i,j) \in A_2, \quad \forall (i,j) \in A \\ C_{ij} + \sum_{k=1}^{l_{ji}} c_{ji}^k x_{ji}^k, (i,j) \in \tilde{A}_2 \end{cases} \quad (9)$$

The objective function shown in eqn.(6) is the total travel time over the study area, which is a function of UE link flow pattern $\mathbf{v}(\mathbf{x})$ and updated link capacity $\mathbf{C}(\mathbf{x})$ expressed by eqn. (9). The parametric VI formulation (7) is the UE conditions, reflecting the reaction resulting from drivers with respect to the contraflow operation solution $\mathbf{x}$.

A Solution Algorithm

Since the preceding MPEC model is a NP-hard discrete optimization problem, a polynomial-time algorithm for it does not exist. However, the GA is capable of solving such a problem. In view of the inherent feature of the MPEC model (6)-(8), we treat the binary row vector $\mathbf{x}$ and the objective function $F(\mathbf{x})$ shown in eqn. (6) as the chromosome and the fitness function, respectively, for designing the GA. Given a chromosome $\mathbf{x}$, its fitness function value can be evaluated by executing a UE traffic assignment procedure associated with the parametric VI (7). Hence, the hybrid GA-UE can be presented below.

The hybrid GA-UE

Step 0. (Initialization) Randomly generate a population with M chromosomes.

Step 1. (Calculation of the fitness function) For each chromosome **x** in the population, calculate the link capacity according to eqn. (9). The relevant fitness function value $F(x)$ is then evaluated by executing a UE traffic assignment procedure.

Step 2. (Generation of a new population). Create a new population by repeating the following three sub-steps until the new population is completed.

Step 2.1. (Selection) According to the fitness function values evaluated in Step 1, use the rank selection method to choose two parent chromosomes from the population.

Step 2.2. (Crossover) With the crossover probability, p_c, cross over the parents to form new offspring according to one point crossover method . If no crossover is performed, offspring is the exact copy of parents

Step 2.3. (Mutation) With the mutation probability, p_m, mutate new offspring at position in chromosome.

Step 3. (Stopping criterion). If a stopping criterion is fulfilled, then terminate, and output the best solution from the population. Otherwise, go to Step 1.

Note that there are a few efficient and effective algorithms for solving VI (7), i.e., UE traffic assignment. The interested readers can refer to Patriksson (1994).

A Numerical Example

To evaluate the MPEC model and the hybrid GA-UE, the Sioux Falls network is adopted. Note that this network is a benchmark example in studying various NDPs. In view of length limitation, its detailed data, such as the network layout, the link capacity, link travel function and the OD demand is not presented here, however they can be found in Suwansirikul et al. (1987).

It is assumed that the study area comprises of 12 links: 4, 14, 16, 19, 22, 47, 49, 52, 53, 58, 59, 61, i.e., $A_1 = \{4,14,16,19,22,47,49,52,53,58,59,61\}$, among which six links, 4, 16, 22, 49, 53, 59, are the candidate links for implementing the contraflow operation, namely, $A_2 = \{4,16,22,49,53,59\}$. It is further assume that each candidate link has three lanes. Hence, the lane capacity of a candidate link is equal to the relevant link capacity divided by three. Obviously, the length of a chromosome coded in the hybrid GA-UE is equal to 18.

The hybrid GA-UE for this example adopt the following parameters: the population size, M = 50; the crossover probability, p_c =0.6; the mutation probability, p_m=0.03 and the maximum number of generation is equal to 10. As all the link travel time functions in this example are separable functions, the Frank-Wolfe algorithm is used to perform the UE traffic assignment procedure of the hybrid GA-UE.

Figure 1 shows the trend of the hybrid GA-UE convergence and it indicates that the hybrid GA-UE for this example converge after six generations. The hybrid GA-UE also comes out with the optimal contraflow operation solution: one lane of both links 4 and 49, and all the lanes of the other four links should be reversed their travel directions during the peak hour. After implementing the optimal contraflow solution, the total travel time over the study area is 4 veh-hrs. Compared to the total travel time over the study area of the 16.2 veh-hrs of before the contraflow operation, it is clear that the contraflow scheme has reduced the total travel time for about 75%. Note that the overall network performance may not be improved since the contraflow operation may be accompanied by the traffic diversion that worsens other links in the network. However, the methodology is able to improve the performance of the target study area which is at interest.

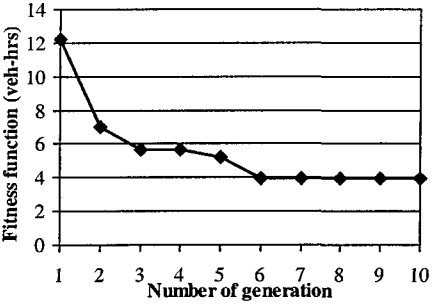


Figure 1. Performace of the hybrid GA-UE

Conclusions

This paper modeled the optimal contraflow operation problem with UE constraints by means of the MPEC approach. A hybrid GA-UE as a solution method was developed accordingly. The numerical example has shown the strength of our proposed methodology.

References

Drezner Z. and Wesolowsky G. (1997) "Selecting an optimum configuration of one-way and two-way routes." *Transportation Science*, 31(4), 386-394.

Luo, Z. Q., Pang, J. S. and Ralph, D. (1996) *Mathematical programs with equilibrium constraints*, Cambridge University Press.

Patriksson, M. (1994) *The traffic assignment problems: models and methods*, VSP, Utrecht, The Netherlands.

Sheffi, Y. (1985) *Urban transportation networks: equilibrium analysis with mathematical programming methods*, Prentice Hall, England Cliffs, NJ.

Suwansirikul, C. and Friesz, T. L. (1987) "Equilibrium decomposed optimization: A heuristic for the continuous equilibrium network design problem." *Transportation Science*, 21, 254-263.

Theodoulou G. and Wolshon B. (2004) "Alternative methods to increase the effectiveness of freeway contraflow evacuation." *Transportation Research Record*, 1865, 48-56.

Yang H. and Bell M.G.H. (1998) "Models and algorithms for road network design: a review and some developments." *Transport Reviews*, 18, pp 257-278.

Selecting Path Flow Pattern for Sensitivity Analysis of Deterministic User Equilibrium

L. Cheng[1], W. Wang[2]

[1,2] Transportation College, Southeast University, Nanjing City, Jiangsu Province, P.R. China; PH 86(25) 8379-5649; FAX 86(25)8379-4102; email: gist@seu.edu.cn

Abstract

The sensitivity analysis is a means for extracting the cause and effect relationship between inputs and outputs in the network model. The paper presents how to find a set of equilibrium path flow pattern by applying the path-based Newton method to deterministic user equilibrium (UE). The solution of path flows becomes candidate flow patterns, from which we can abstract extreme points specially used for sensitivity analysis, by using a linear programming. The numerical example shows that plural extreme patterns exist for sensitivity calculation, but the sensitivity of link flow to the input variables is still unique.

Keywords: sensitivity, network flow, user equilibrium, linear programming

Introduction

The network flow model attempts to duplicate the vehicular flows on the network, by partitioning the origin-destination trip rates between the observed paths. The sensitivity analysis is a method for extracting the cause and effect relationship between the inputs and outputs in the network flow model. The basic idea is that each input channel to the network is offset slightly and the corresponding change in the outputs is reported. For example, link time or origin-destination demand is likely to contain uncertainties because of the difficulty of obtaining accurate data; the input channels that produce high sensitivity values can be considered significant and can most often be paid attention in network design and traffic operation.

The sensitivity analysis of the deterministic user equilibrium (UE) was developed as that of a restricted variational inequality (Tobin and Friesz, 1988). The sensitivity in the context of the logit-based stochastic UE was addressed as a link-based approach (Ying and Miyagi, 2001), implemented in the sophisticated procedure of the Dial's algorithm, and the sensitivity in the probit-based stochastic UE was explored on the basis of the information on the stochastic UE path flows (Clark and Watling, 2002). The sensitivity analysis of deterministic or stochastic UE has been applied for solving congestion pricing and network design problems (Yang 1997; Ying, 2005).

The sensitivity of the deterministic UE requires a previous set of reference path flows, from which an extreme path flow pattern can be abstracted. Unfortunately, they didn't tell us how to obtain the candidate path flow pattern. The traffic condition of deterministic UE is unique with link flows, but not unique with respect to path flows. A link flow pattern may correspond to several path flow patterns (Patriksson, 1994). Even when the link flow pattern is known, finding an equilibrium path flow is still not a trivial task. For a small

network, all the paths could easily be enumerated as the reference set, but for a large network, path enumeration is unimaginable.

The objective of this paper is to demonstrate a comprehensive methodology for investigating factor sensitivities of deterministic UE model over the general road network. We first present how to find a set of equilibrium path flow pattern by using the Newton method, a path-based algorithm, which integrates Newton formula and column generation; second, we illustrate how to abstract the extreme path flow pattern used for flow sensitivity analysis, by using a linear programming. The two problems play a significant role in the applying sensitivity analysis to traffic network practices. A numerical example is finally demonstrated for the deterministic UE models.

Path-Formulated UE Network Model

Given a transport network $G(A, N)$, where A and N are the sets of links and nodes respectively, and each directed link $a \in A$ is associated with a travel time $t_a(x_a)$ as an function of link flow x_a. W is the set of origin destination pairs, and for each pair $w \in W$, there is a given traffic demand q^w. The user equilibrium assignment problem is stated as follows:

$$\text{minimize } z(\mathbf{x}) = \sum_a \int_0^{x_a} t_a(x) \cdot dx \tag{1a}$$

$$\text{subject to } \sum_{k \in K^w} f_k^w = q^w \quad \forall w \in W \quad (\text{or } \mathbf{\Lambda f} = \mathbf{q}) \tag{1b}$$

$$x_a = \sum_w \sum_k f_k^w \cdot \delta_{ak}^w \quad \forall a \in A \quad (\text{or } \mathbf{x} = \mathbf{Af}) \tag{1c}$$

$$f_k^w \geq 0 \qquad \forall k \in K^w, w \in W \tag{1d}$$

where

x_a	= flow on link
$t_a()$	= travel time on link a
δ_{ak}^w	= 1 if link a belongs to path k; otherwise, $\delta_{ak}^w = 0$
f_k^w	= flow on path k within OD pair w
K^w	= set of paths within OD pair w
q^w	= origin-destination demand between pair w

The UE network problem is path-formulated due to the constraints of equations (1*b*) - (1*d*). Alternatively, the same problem can be formulated using link variables indexed by origin or destination (Patriksson, 1994). We are primarily interested in the formulation using path variables in this paper. The following Lagrange function will be used to build the necessary equation at optimality.

$$L[\mathbf{x(f)}, \mathbf{u}] = \sum_a \int_0^{x_a} t_a(x) + \mathbf{u}^T(\mathbf{q} - \mathbf{\Lambda f}) \tag{2a}$$

In terms of path flows, the derivative of the Lagrange function equals zero, so that,

$$\boldsymbol{\tau} - \mathbf{\Lambda}^T \mathbf{u} = 0 \quad \text{(the first-order condition)} \tag{2b}$$

Where, $\mathbf{u}$ is the vector of Lagrange multiplier with elements, $u^w = \min_k \{\sum_a t_a \cdot \delta_{ak}^w\}$, the shortest path travel time within OD pair w. $\boldsymbol{\tau}$ is the time vector corresponding to the used paths, with elements, $\tau_k^w = \sum_a t_a \cdot \delta_{ak}^w$.

The solution of the UE network model is considered to have the characterization of a Wardrop's principle. Either the flow on that path is zero, where the travel time on the path,

τ^w_k, must be greater than or equal to the OD pair specific Lagrange multiplier, u^w, or the flow on the path k is positive, where, $\tau^w_k = u^w$. In any event, the Lagrange multiplier of a given OD pair (1*b*) is less than or equal to the travel time on all paths connecting this pair.

Finding Candidate Path Flow Pattern

The Newton method is adopted for solving the path-formulated network flow model and finding the candidate path flow pattern. Column generation and line search are incorporated into the solution procedure. The column generation allows that the active paths joining each OD pair are generated endogenously only when needed. The line search is implemented to avoid the infeasibility of flows on the used paths and equilibrate the demand among the active paths as well as the newly founded shortest path for every OD pair. The Newton method (Cheng et al, 2003) is reported to be superior to the gradient projection (Bertsekas and Gallager, 1987) in the robustness of convergence.

Different from the optimization of the link-based algorithm, which is performed in the link-flow domain, the Newton method is performed in the path-flow domain. With respect to path flow vector only inclusive of used elements, the proposed method iteratively solves problem (1) from a feasible point to an improved feasible point. Provided a feasible point or say path flow vector $\mathbf{f}^n$, a moving flow $\Delta\mathbf{f}^n$ and a step size λ are determined such that the following two properties are true:

- $\mathbf{f}^n + \lambda\cdot\Delta\mathbf{f}^n$ is feasible
- the value of objective function at $\mathbf{f}^n + \lambda\cdot\Delta\mathbf{f}^n$ is better than that at $\mathbf{f}^n$

This leads to a new point

$$\mathbf{f}^{n+1} = \mathbf{f}^n + \lambda\cdot\Delta\mathbf{f}^n \tag{3}$$

The process above mentioned is repeated until the objective function cannot be further improved. In the space of path flows, the objective function z(**x**) in equation (1a) is approximated by the following second-order Taylor series:

$$z(\mathbf{f}) \cong z(\mathbf{f}^n) + (\mathbf{f}-\mathbf{f}^n)^T\cdot\boldsymbol{\tau}(\mathbf{f}^n) + \tfrac{1}{2}(\mathbf{f}-\mathbf{f}^n)^T\cdot\boldsymbol{\tau}'(\mathbf{f}^n)\cdot(\mathbf{f}-\mathbf{f}^n) \tag{4}$$

where $\boldsymbol{\tau}(\mathbf{f}^n)$, $\boldsymbol{\tau}'(\mathbf{f}^n)$ denote the vectors of path costs and their derivatives, which are corresponding to the first and second order derivatives of the objective function to the respective path flows. $\mathbf{f}^n$ is the vector of current path flow solutions. The minimization is performed at the approximate function. The first derivative of the approximate function is:

$$\boldsymbol{\tau}(\mathbf{f}) = \boldsymbol{\tau}(\mathbf{f}^n) + (\mathbf{f}-\mathbf{f}^n)^T\cdot\boldsymbol{\tau}'(\mathbf{f}^n) = \boldsymbol{\tau}(\mathbf{f}^n) + \Delta\mathbf{f}^n\cdot\boldsymbol{\tau}'(\mathbf{f}^n) \tag{5}$$

The improving incremental along the domain of path flows becomes

$$\Delta\mathbf{f}^n = -\left[\boldsymbol{\tau}(\mathbf{f}^n) - \boldsymbol{\tau}(\mathbf{f})\right]\cdot\left[\boldsymbol{\tau}'(\mathbf{f}^n)\right]^{-1} \tag{6}$$

For the convenience of expression, the active paths are partitioned into the shortest path k^{w*} and the non-shortest paths $k \in K^w$ for any pair $w \in W$. Toward the optimum, flows on all the non-shortest paths will decrease according to the direction expressed by equation (6). The decrease of flows on the non-shortest paths within each OD pair will shift to its corresponding shortest path. The shortest path k^{w*}, if not included in the current path set, will join the path set K^w (*n*), at the next iteration, i.e. $K^w(n+1) := K^w(n) \cup k^{w*}$. On the requirement of the flow conservation of every OD pair, the summation of decreased flows will become the increasing amount of the shortest path, or say, equal the path flow of the shortest path because no flow exists on the shortest path just found. The flow change (6) in used paths can be further expressed by

$$\Delta\mathbf{f}^n = -\left[\boldsymbol{\tau}(\mathbf{f}^n) - \tau(f^w_{k^{w*}})\cdot\mathbf{I}\right]\cdot\left[\boldsymbol{\tau}'(\mathbf{f}^n)\right]^{-1} \tag{7a}$$

where **I** is an vector of appropriate dimension with element one, and $\tau(f_{k^{w*}}^{w})$ represents the cost of shortest path k^{w*} within pair w. Accordingly, the flow or flow change on the shortest path can be calculated by

$$f_{k^{w*}}^{w,\,n+1} = q^{w} - \sum_{k \in K^{w},\, k \neq k^{w*}} f_{k}^{w,\,n+1} \qquad \forall w \in W \tag{7b}$$

The detail of solution procedure might be referred to Cheng et al (2003).

Extracting Extreme Path Flow Pattern

The approach taken here to calculate the derivatives of flows with respect to the perturbation is to select one particular path flow pattern, in particular an extreme point of the feasible region. Tobin and Friesz (1988) point out that the extreme point problem is restricted to only those paths that are positive. The positive path flows are produced endogenously in the column generation of the Newton method. The path-based Newton method provides alternative path flow patterns and might facilitate the sensitivity analysis. If the algorithm generates an extreme point of path flows, it can be used as the path vector, f*, for flow sensitivity; if not, it provides a path set of feasible points, from which the path vector of the extreme points can be extracted.

$$\text{minimize } \mathbf{b}\,\mathbf{f} \tag{8a}$$
$$\text{subject to } \mathbf{f} \in \{\mathbf{f} \mid \text{equation (1b)}\sim\text{(1c)}\} \tag{8b}$$
$$\mathbf{b} \geq 0 \tag{8c}$$

where **b** is a vector of constants of size | **f** |. The exact nature of **b** is not important. For simplicity the vector **b** can be chosen of element one. Note that a column generation is planted in the path-based algorithm, where the shortest path of the given OD-pair is generated. A pick/drop criterion may decide whether the searched paths are saved or not, and is dependent on the specific path-based algorithm. For instance, any shortest paths not inclusive the path set will be preserved and used to augment the path set in the Newton method.

Sensitivity Analysis of Deterministic UE

With respect to the equation (2b), the following differential relation may be created by using chain rule,

$$d\boldsymbol{\tau} = \frac{\partial \boldsymbol{\tau}}{\partial \mathbf{f}} d\mathbf{f} + \frac{\partial \boldsymbol{\tau}}{\partial \mathbf{u}} d\mathbf{u} = \mathbf{D} d\mathbf{f} + \boldsymbol{\Lambda}^{T} d\mathbf{u} \tag{9a}$$

where, $\mathbf{D} = \mathbf{A}^{T} \nabla^{2} z(\mathbf{x}) \mathbf{A} = \nabla^{2} z(\mathbf{f})$. Note that the path flow pattern is here a subset of that of the solution of equilibrium network, although we simply use, **f**, the same notation of path flow pattern.

With respect to the equation (1b), the differential equation exists,

$$d\mathbf{q} = \boldsymbol{\Lambda} d\mathbf{f} \tag{9b}$$

Combined with equations (9a)-(9b), the relationship hold for small changes of the primal and dual variables at optimum,

$$\begin{bmatrix} d\boldsymbol{\tau} \\ d\mathbf{q} \end{bmatrix} = \begin{bmatrix} \mathbf{D} & \boldsymbol{\Lambda}^{T} \\ \boldsymbol{\Lambda} & \mathbf{0} \end{bmatrix} \begin{bmatrix} d\mathbf{f} \\ d\mathbf{u} \end{bmatrix} \tag{10}$$

This matrix is called Jacobian matrix, which builds the effect of perturbations in primal and dual variables on the path times, origin destination demands. The inverse of this Jacobian provides the sensitivities of path flow and related dual variables to those input data, and the relationship is given as:

$$\begin{bmatrix} d\mathbf{f} \\ d\mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{J}_{11} & \mathbf{J}_{12} \\ \mathbf{J}_{21} & \mathbf{J}_{22} \end{bmatrix} \begin{bmatrix} d\boldsymbol{\tau} \\ d\mathbf{q} \end{bmatrix} \tag{11}$$

where, $\mathbf{J}_{**}$ is the block comprised in the inverse of the Jacobian matrix. Except for the origin destination demand, the input data usually used is not the travel times of path but link, and the output of interest is also the traffic conditions of links, in light of link-path incidence relationship, the matrix equation can be reformulated as,

$$\begin{bmatrix} d\mathbf{x} \\ d\mathbf{u} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\mathbf{J}_{11}\mathbf{A}^T & \mathbf{A}\mathbf{J}_{12} \\ \mathbf{J}_{21}\mathbf{A}^T & \mathbf{J}_{22} \end{bmatrix} \begin{bmatrix} d\mathbf{t} \\ d\mathbf{q} \end{bmatrix} \tag{12}$$

Numerical Example

A numerical example is used to illustrate the comprehensive methodology of the deterministic UE sensitivity analysis. The network, shown in Figure 1, consists of 4 links and 3 nodes. The origin destination demand and link performance are listed in Table 1. The standard BPR function is adopted.

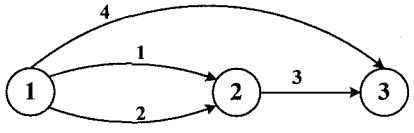


Figure 1. Network

Table 1. Input data

link	t_a^0	c_a	q
1	10	600	$q_{OD(1,2)} = 600$
2	17	500	
3	9	800	$q_{OD(1,3)} = 400$
4	60	400	$q_{OD(2,3)} = 600$

Table 2. Path flow pattern of deterministic user equilibrium

OD	Pattern 1		Pattern 2		Pattern 3		Path Time
	Path Flow	Makeup	Path Flow	Makeup	Path Flow	Makeup	
(1, 2)	$482.11+s$	1	600	1	482.11	1	17.01
	$117.89-s$	2	0	2	117.89	2	17.01
(2, 3)	600	3	600	3	600	3	12.3
(1, 3)	$400-s$	1-3	282.11	1-3	400	1-3	29.3
	s	2-3	117.89	2-3	0	2-3	29.3

When different feasible initial link flows are given, the Newton method can produce diverse path flow vectors of the UE solution. The general form of path flow patterns can be summarized in Pattern 1 of Table 2. Simple calculations verify that every path flow vector for $s \in [0, 117.89]$ observes flow conservation and path travel time comply with the Wardrop principle. The relation of path flows in Table 2 shows that path flows are not capricious, but are subject to some intrinsic relationship. Among node 1 and node 2 there are two paths, we can foretell that the path flow from node 1 to node 2 along link 1 is within the interval [482.11, 541.06], while the path flow along link 2 is within the interval [0, 117.89], etc..

Table 3. Sensitivity of deterministic UE to perturbation from input data

Link Flow	Link Free Time				OD Demand		
OD Time	t_1	t_2	t_3	t_4	q_1	q_2	q_3
x_1	31.1818	-31.1818	0	0	0.0084	0.0084	0
x_2	-31.1818	31.1818	0	0	0.9916	0.9916	-1
x_3	0	0	0	0	0	1	1
x_4	0	0	0	0	0	0	0
u_1	0.0084	0.9916	0	0	-0.0003	-0.0003	0
u_2	0.0084	0.9916	1	0	-0.0003	-0.0135	-0.0132
u_3	0	-1	1	0	0	-0.0132	-0.0132

Although the general pattern in Table 2 can not be directly applied in the sensitivity analysis due to the non-invertibility of the matrix **D**, it provides a candidate path set of feasible points, from which the path vector of the extreme points can be extracted. The solution of the simplex model (8) is the Pattern 3 in Table 2, which is the extracted extreme point suitable for sensitivity analysis. Pattern 3 is one extremity with s=0, while Pattern 2 is another extremity with s=117.89. Pattern 2 is not the result of simplex model (8), however, our numerical experiment shows that Jacobian matrix is invertible in terms of Pattern 2, and that the sensitivity of link flows to the input parameters is the same as that in the use of Pattern 3. It can be shown that there exist plural extreme points suitable for the deterministic UE sensitivity. The sensitivity results of deterministic UE are illustrated in Table 3.

Conclusion

The paper presents a powerful way of how to find a set of equilibrium path flow pattern by applying the path-based Newton method to the deterministic user equilibrium. The solution of path flows becomes candidate flow patterns, from which we can abstract the extreme path flow pattern specially used for flow sensitivity analysis, by using a linear programming. The numerical example shows that the proposed method is efficient for sensitivity and that plural extreme patterns might exist for sensitivity calculation, but the sensitivity of link flow to the input variables is still unique.

References

Bertsekas, D., and Gallager, P. (1987). *Data networks*. New Jersey, Prentice Hall, 365-478.

Cheng, L., Iida Y., Uno N., and Wang W. (2003). Alternative Quasi-Newton methods for capacitated UE assignment. *Transportation Research Record*, 1857, 109-116.

Clark, S. D., and Watling, D. P. (2002). Sensitivity analysis of the probit-based stochastic user equilibrium assignment model. *Transportation Research B*, 36(7), 617-635.

Patriksson, M. (1994). *The traffic assignment problem: models and methods*. VSP, Utrecht.

Tobin, R. L., and Friesz, T. L. (1988). Sensitivity analysis for equilibrium network flow. *Transportation Science*, 22(4), 242-249.

Yang, H. (1997). Sensitivity analysis for the elastic-demand network equilibrium problem with applications. *Transportation Research B*, 31(1), 55-70

Ying, J. Q., and Miyagi, T. (2001). Sensitivity analysis for stochastic user equilibrium network flows - a dual approach. *Transportation Science*, 35(2), 1-10.

Ying, J. Q. (2005). Sensitivity analysis based method for optimal road network pricing. *Annals of Operations Research*, 133, 303-317.

Pricing Incremental Reductions in Vehicular Emissions through Motor Fuel Regulation

F. W. Rusco[†] and W. D. Walls[‡]

[†]Center for Economics, U.S. Government Accountability Office, Washington D.C., U.S.A. Email: RuscoF@gao.gov

[‡]Department of Economics, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4. Email: wdwalls@ucalgary.ca

Abstract

Environmental regulations have increasingly focused on mobile-source pollution and in particular on the automobile itself. More recently, focus has shifted toward the energy source used to propel vehicles. In the United States, in particular, the attainment of emissions standards at the local and regional level has led to a proliferation of special motor vehicle fuels. This has occurred because federal environmental standards must be achieved, but regional authorities are able to choose how to achieve those environmental standards. As a result, regional jurisdictions---sometimes as large as an entire state---have adopted specialized motor vehicle fuels also known as 'boutique' fuels. The heterogeneity in motor fuel markets created by this proliferation of fuel types has stressed an already taxed supply infrastructure and led to generally higher and more volatile gasoline prices. In this paper we provide a brief summary overview of the North American market for motor vehicle fuels with a particular emphasis on gasoline. We execute a statistical analysis of the price effects of the proliferation of unique blends of motor fuels in the achievement of federal environmental standards. We find that the effects of adding more fuels to the supply chain increases overall fuel prices, in particular for those areas using unique gasoline blends. In addition, the proliferation of unique gasoline blends increases the likelihood of supply disruptions, particularly in cities or areas that are isolated by virtue of using an uncommon motor fuel or because they are located far from sources of supply. The higher and more volatile prices of boutique motor fuels must be balanced against the environmental benefits of cleaner burning fuels.

Introduction

The United States Environmental Protection Agency (EPA) and others have developed models that provide detailed emissions estimates for several of the special gasoline blends currently in use. These models have been used to estimate the expected emissions reduction resulting from requiring the use of special gasoline blends as a substitute for conventional gasoline. The models estimate that special gasoline blends will reduce emissions by varying degrees. California's gasoline---the blend formulated to reduce emissions the most---is estimated to provide the greatest level of emissions reductions; about 25-29 percent for VOCs

and about 5.7 percent for NOx. Federal reformulated gasoline (RFG) is estimated to provide about the same level of VOC reduction, a lower NOx reduction of about 0.7 percent, but also a 10-20 percent reduction in CO. The special gasoline blend most commonly used in areas not using conventional gasoline---gasoline with a Reid vaporization pressure (RVP) of 7.8---is estimated to reduce VOC emissions by 12-16 percent and NOx by about 0.7 percent. RFG and California's cleaner burning gasoline also reduce emissions of some other toxic substances such as benzene.

The extent of emissions reductions associated with various gasoline blends remains somewhat uncertain. The GAO (1997), the National Research Council, and others have identified concerns about the overall accuracy of emissions estimates. EPA has addressed some of the concerns about emissions estimates. In one effort to address concerns about the validity of emissions estimates, EPA sponsored a study that compared emissions estimates to measured emission data obtained between 1992 and 2001. The study looked at pollutant concentration data from tunnels and vehicle exhaust data collected from vehicles on roadways using special remote sensing devices at a limited number of sites using a limited range of gasoline blends,. As a result, EPA found that the observed emissions data conflicted with emissions estimates; in some cases the testing data were higher than predicted, while in other cases it was lower.

In addition to these broad concerns, there is also controversy over the emissions benefits associated with special blends containing oxygenates, which were initially added to gasoline to reduce the emissions of carbon monoxide and other pollutants. However, although there appears to be agreement that oxygenated fuels help reduce emissions of CO from older vehicles, recent studies indicate that the emissions benefits for newer vehicles are questionable. For example, AQIRP and the National Science and Technology Council and others have reported that improvements in emissions controls on newer vehicles, such as oxygen sensors and computer-controlled emissions systems, may now automatically reduce emissions of CO and other pollutants and may negate many benefits of adding oxygenates. Further, some experts have concluded that adding oxygenates to gasoline may increase emissions of NOx and VOCs and may contribute to increased levels of ozone. As a result, some states, including California, New York, and Georgia have requested waivers from EPA to allow them to use fuel that does not contain an oxygenate. The state of California stipulated in its waiver application, that its fuel reduces emissions to a greater extent than federal RFG and that the oxygenate requirement has impeded its efforts to reduce ozone. To date, EPA has not granted any of these waivers.

EPA and other experts have concluded that improvements in air quality in some parts of the country are at least partly attributable to the use of special gasoline blends. In 2004, EPA reported that ground-level ozone has decreased over the past 10 to 25 years, and that these reductions resulted, at least in part, from emissions control programs that include requirements to use special gasoline blends. Further, EPA and other experts concluded that special gasoline blends, such as RFG and low-RVP blends, are effective strategies for states to use to reduce ozone pollution. In addition, a research effort funded by AQIRP found that reducing RVP decreased peak ozone in several cities and would continue to provide benefits for years to come. In addition, the National Research Council reviewed EPA data and found that average ozone levels dropped by about 1 percent coincident with reduced emissions of VOCs, NOx and CO from on-road vehicles, which fell by 31 percent, 2 percent, and 20 percent, respectively. Based on these and other data, the Council concluded that improvement in air quality is likely attributable, at least in part, to recent improvements in gasoline properties. Other possible reasons for the improvements include the advent of more

stringent standards for vehicles that gradually replace old vehicles built to more lenient standards than current models; maturation of new-vehicle emissions-control hardware and software as field experience accumulated.

It is typical for areas using special gasoline blends to be surrounded by regions that use conventional gasoline. In some cases, these areas are relatively large, as is the case for the state of California, where nearly all of the state uses the same fuel. In other cases, "islands" of special gasoline use can divide otherwise regional gasoline markets. For example, in the St. Louis metropolitan area, which includes parts of two states (Each state is overseen by a separate EPA regional office; Missouri is overseen by EPA region 7 and Illinois is overseen by EPA region 5.)-Missouri and Illinois-uses three different fuels: one special gasoline blend required on the Missouri side, a different special gasoline blend required on the Illinois side, and conventional gasoline is allowed in the surrounding area. In some cases, special gasoline blends are used in only one area of the country. For example, California CBG, Arizona CBG, and the special blend used in Atlanta, Georgia are not used anywhere else in the United States. Even relatively common special gasoline blends can create isolated markets if they are not used in nearby areas. For example, although 7.8 RVP is a relatively widely used blend, Pittsburgh, Pennsylvania, is the only city in its region that uses it. Similarly, the Chicago/Milwaukee area uses RFG North with ethanol, a gasoline blend used in the Northeast, but not used elsewhere in the Midwest.

Fuel Type	Description
Conventional	The most widely available gasoline; used where air quality is satisfactory; formulated to evaporate more slowly in hot weather with RVP limits; contains detergent additives to reduce engine deposits.
RFG	Reformulate gasoline; mandated in areas where air quality is persistently unsatisfactory; contains oxygenates.
CARB	California Air Resources Board RFG; a different formulation of RFG that burns cleaner than regular RFG. After 2002 MTBE no longer used as oxygenate in CARB RFG due to concerns over MTBE contaminating ground water.
Low Sulfur	Gasoline with a low sulfur content and low RVP. Originally mandated for Atlanta and now used in many Georgia counties.

Table 1. Blends of Gasoline.

Special gasoline blends accounted for more than half the gasoline consumed in the United States during the summer of 2001---the last year for which complete data were available. Of the special fuel blends, RFG and 7.8 RVP blends together accounted for about 33 percent of the national gasoline market. California CBG and Arizona gasoline blends accounted for roughly 13 percent of total U.S. gasoline consumption. The remaining 6 percent of gasoline use was divided among four separate blends.

While there were 11 special blends of gasoline during the summer of 2004, more than 45 gasoline blends were sold in the United States throughout the year. Special winter-only gasoline blends are required to be used in areas of 8 states; these blends contain an oxygenate to address winter carbon monoxide pollution. And because many gasoline stations sell gasoline in three octane grades there is also a doubling of fuels. (Both premium and regular grades are refined and shipped to terminals, where they are blended together to make mid-grade gasoline.) Thus, pipelines, terminals, and retailers carry multiple variations of the

gasoline blends. And these gasoline blends differ regionally and seasonally: differences in outside temperature require different blends to maintain vehicle performance. The primary difference among these blends is RVP. Refiners produce gasoline with higher RVP in cold conditions to allow cars to start, and gasoline with lower RVP during warm conditions to improve vehicle operation, even in areas that use conventional gasoline. As a result of these differences, refiners routinely ship different fuels to different regions and also ship different gasoline blends seasonally, but special blends tend to compound these variations.

Statistical Analysis of Fuel Prices

Wholesale gasoline prices were obtained from the Oil Price Information Service (OPIS) for ninety-nine cities in the United States. The price data are weekly observations on each type of fuel sold at each city over the interval from 7th December 2000 to 28th October 2004. The sample of data includes cites that use conventional gasoline as well as those that use special fuels. We subtracted from each gasoline price the corresponding weekly price of West Texas Intermediate (WTI) crude oil and constructed a single price series for each city that represents the price of the fuel actually used in that city in a particular week: the OPIS data may list numerous fuel prices for a particular city, but the city price we refer to will be for the particular reformulated gasoline mandated for use in that city at that calendar date. Because the fuel required varies across summer and winter seasons---and because the fuel regulations are varying across years in some cities---the price series for each city will typically represent multiple fuels. This variation in fuels with city markets as well as across city markets provides the richness present in our data set and the opportunity to analyze it using statistical methods designed for analyzing time series of cross-sectional observations. Complete descriptive statistics are presented in Ludwigson *et al.* (2005).

The statistical model relates the price level for gasoline in each city i during a particular week t to the price of West Texas intermediate (WTI) crude oil, the distance to closest alternative source of supply for the particular fuel in that week, the attributes of the particular fuel, and a set of city-market-specific effects a_i. Algebraically the regression model can be expressed as

$$\text{Price}_{it}=a_i+b_1\text{WTI}_t+b_2\text{Distance}_{it}+G'\text{Fuel Attributes}_{it}+m_{it}$$

where m_{it} is the random disturbance for city i in week t. The model quantifies the joint distribution of gasoline prices across cities; it is not a demand or supply equation. The purpose of the statistical analysis is to explain the level of fuel prices in relation to city-market-specific effects and fuel-specific effects, and for this purpose a reduced-form model is appropriate.

The columns of Table 2 display the estimates for the fixed-effects and random-effects specifications. In each of these specifications a separate city-market-specific effect is estimated for each city, but the specifications differ in whether the city effect is modeled as being predetermined or random. In the random-effects formulation, statistical inferences that are unconditional with respect to the population of all possible city-specific effects are made, while in the fixed-effects formulation statistical inferences that are conditional on the city-specific effects in the sample are made (Hsiao, 1986). The estimates from fixed-effects and random-effects models can differ significantly in the case where a large number of cross-sectional units are observed over a small number of time periods (Hausman, 1978). Given the context of our analysis, it is sensible for us to make statistical inferences that are not conditional on the city-market-specific effects in the sample; however, for completeness we

investigate the fixed-effects formulation. In any case, because we observe prices across 99 cities for 204 weeks in our sample of data, and because the effect of having a large number of time observations for the cross-sectional units is that the fixed-effects and random-effects estimates are nearly the same, this is not an issue in the interpretation of our statistical results.

The coefficient on WTI indicates that a one cent increase in the price of crude oil increases the price of wholesale gasoline by about 1.17 cents which is statistically different from zero. The coefficient on the distance to the nearest substitute fuel shows that the price of gasoline rises about 4 cents for each additional thousand miles after controlling for attributes of the fuels and city-specific effects.

Variable	Fixed Effects	Random Effects
WTI (¢ per gallon)	1.169201	1.169123
	(0.00421)	(0.00423)
Distance to Substitute	0.003912	0.004283
	(0.00095)	(0.00087)
Low Sulfur	5.071134	4.788197
	(1.37793)	(1.36936)
RVP 7.8	4.470009	3.861721
	(0.36578)	(0.36109)
RVP 7.2	9.674229	9.407312
	(1.40078)	(1.39413)
Ethanol 5–5.7%	1.034584	1.499145
	(1.35499)	(1.33506)
Ethanol 10%	3.622967	4.027807
	(0.63527)	(0.62227)
RFG MTBE RVP 8.2	8.534264	9.119724
	(0.91706)	(0.91720)
RFG MTBE RVP 7.2	6.862264	6.730516
	(0.79989)	(0.80202)
RFG Ethanol RVP 8.2	6.016466	7.581270
	(1.04539)	(0.96414)
CARB RFG Ethanol	16.16030	16.42418
	(1.30387)	(1.30772)
R^2	0.821	0.819

Table 2. Regression Results.

The coefficient on low sulfur fuel is statistically larger from zero with an estimate of about 5 cents per gallon increase in price due to sulfur content regulation. Next, we examine the impact of increasingly stringent Reid vaporization pressure standards. The coefficients on RVP levels of 7.8 and 7.2 are roughly 4 and 9 1/2, respectively, with both being statistically different from zero at the one percent marginal significance level; we expected the coefficients on all the RVP indicator variables to be positive and significant, and we also expected successively lower RVP to result in successively higher prices. The coefficients on the required ethanol content are also positive and increasing in the percentage of ethanol, however only the coefficient on 10% ethanol content differs statistically from zero. The remaining fuel attribute variables are for various types of reformulated gasoline, first those that use MTBE as an oxygenate, where the effect on price is about 6 3/4to 9 cents per gallon. Reformulated gasoline using ethanol increases the expected price by about 6 to 7 ½ cents per gallon, and California's CARB gas using ethanol as an oxygenate increases the expected price by about 16 cents per gallon, with all coefficients statistically different from zero.

The results of the regression equation provide estimates of the reduced-form impact of changes in particular fuel specifications on expected wholesale fuel prices, and these are generally consistent with our expectations that requiring more stringent and costly fuels—those with lower Reid vaporization pressure, higher ethanol content, and the most stringent CARB formulation—results in higher wholesale fuel prices.

The city-market effects reflect the constant terms a_i in our regression equation for each city i after the regression model has explicitly accounted for all of the explanatory variables included in the estimation, such as the price of WTI, the attributes of the fuel used, and the distance to the nearest city with a substitute fuel. The city effects are highest for Anchorage followed by the California, Arizona, and Nevada cities. The city effects are lowest for Beaumont and Meridian, and also low for other cities that are in close proximity to the Gulf Coast refining centers.

Conclusion

In this paper we have provided a brief summary of the North American gasoline market in relation to Federal environmental standards. We provided a statistical analysis of the price effects of the proliferation of unique blends of motor fuels in the achievement of federal environmental standards. We found that the effects of adding more fuels to the supply chain increases overall fuel prices, in particular for those areas using unique gasoline blends. In addition, the proliferation of unique gasoline blends increases the likelihood of supply disruptions, particularly in cities or areas that are isolated by virtue of using an uncommon motor fuel or because they are located far from sources of supply. In applied policy work, it is important that the higher and more volatile prices of boutique motor fuels be balanced against the environmental benefits of using those cleaner burning fuels.

References

Dahl, C. (1986). Gasoline demand survey. *The Energy Journal*, 7(1):67–82.

Dahl, C. and Sterner, T. (1991). Analysing gasoline demand elasticities: A survey. *Energy Economics*, 13(3):203–10.

GAO (1997). Air pollution: Limitations of EPA's motor vehicle emissions model and plans to address them. Technical Report GAO/RCED-97-210, U.S. Government Accountability Office, Washington, D.C.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, 46:1251–1272.

Hsiao, C. (1986). *Analysis of Panel Data*. Number 11 in Econometric Society Monographs. Cambridge University Press, Cambridge.

Ludwigson, J., F. W. Rusco and W. D. Walls. (2005). Environmental regulation and the market for motor vehicle fuels: Unintended consequences of fuel-type proliferation, including higher and more volatile prices. Proceedings of the 25th IAEE North American Conference, Denver, September.

Muehlegger, E. (2002). The role of content regulation on pricing and market power in regional retail and wholesale gasoline markets. Technical Report 008, MIT Center for Energy and Environmental Policy Research, Cambridge.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and direct test for heteroskedasticity. *Econometrica*, 48:817–838.

Modeling Time-dependent Tolls under Transport, Land Use, and Environment Considerations

X. Q. Li[1], W.Y. Szeto[2] and M. O'Mahony[3]

[1]Centre for Transport Research (TRIP), Department of Civil, Structural and Environmental Engineering, The University of Dublin, Trinity College, Dublin 1, Ireland; PH (353) 1-608-2537; FAX (353) 1-677-3072; email: lixq@tcd.ie
[2]Centre for Transport Research (TRIP), Department of Civil, Structural and Environmental Engineering, The University of Dublin, Trinity College, Dublin 1, Ireland; PH (353) 1-608-3646; FAX (353) 1-677-3072; email: szetow@tcd.ie
[3]Centre for Transport Research (TRIP), Department of Civil, Structural and Environmental Engineering, The University of Dublin, Trinity College, Dublin 1, Ireland; PH (353) 1-608-2084; FAX (353) 1-677-3072; email: margaret.omahony@tcd.ie

Abstract

Recently, there has been resurgence of interest in road pricing (Lo and Szeto, 2005). Along with the popularity of the concept of sustainability, transport planners no longer ignore the environmental considerations when analyzing and designing a transportation system for pricing. This paper develops an analytical model to determine the optimal tolls over time to control the traffic emissions while capturing the land use-transport interaction. To illustrate the effect of tolls on the transportation system with land use and environmental considerations, a numerical study is performed. The results show that the implementation of tolls can reduce the overall traffic emissions, generate more transit revenue and alter the travelers' choices of modes, routes, living and working locations.

Introduction

Environmental pollution due to traffic is one of the most critical problems confronted by policy makers as well as the public. Transport planners no longer ignore the environment when designing and evaluating a sustainable transport system, and they do not ignore the dynamic relationships between land use, transport, and environment. One popular method adopted by transport planners to manage the transport system is by charging tolls on roads. In the past, much research has been carried out to address some practical perspectives when designing an optimal road pricing scheme (e.g. Mun et al., 2001; Hyman and Mayhew, 2002; Zhang and Yang, 2004). In particular, Sumalee (2005) proposed an optimization algorithm considering a time factor to design an optimal implementation path for a charging cordon scheme. Nonetheless, he did not address the relationships between land use, transport and environment. To address this, we develop an analytical model to determine the time varying optimal tolls to control the traffic emissions while capturing the land use-transport interaction. This model is new in the sense that there was no analytical model to determine time-varying tolls capturing the land use-transport-environment interaction. We study the effect of tolls on environment, the land use-transport interaction as well as the public transit system. In the following, the second section

depicts the proposed model and its structure. The third section is the numerical study. Finally, the last section gives concluding remarks.

Model and its structure

An analytical model is developed to determine the tolls over time while capturing land use and environmental considerations. The way in which land use interacts with transport and environment over time is outlined in figures 1 and 2 below.

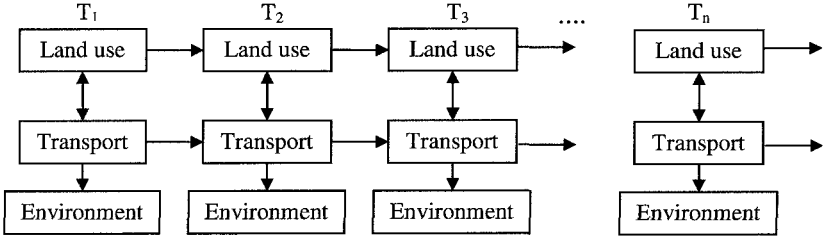


Figure 1. Dynamic relations in the land use and transport systems with environmental consideration.

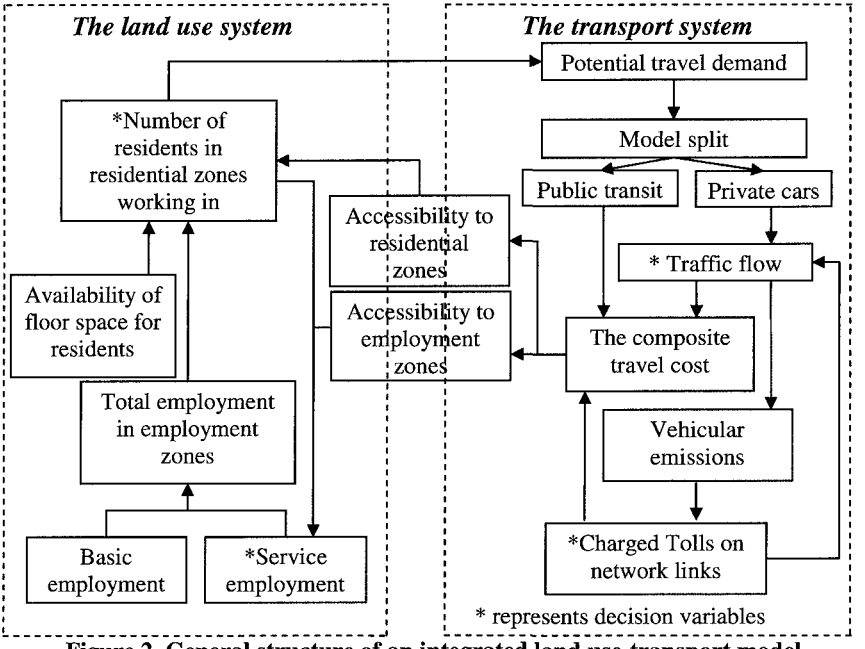


Figure 2. General structure of an integrated land use-transport model with environmental consideration.

To simplify the analysis, the following assumptions are made in this paper. 1. Basic employment, basic employment growth rate over time, and the zonal attractiveness are known; 2. Each zone has only one characteristic, and the residential zone and the

employment zone are not mixed; 3. Traffic assignment follows the user-equilibrium principle; 4. The link cost and travel demand functions are separable; 5. The interest and inflation rates are constant over time for the sake of simplicity. These assumptions are not restrictive from a modelling perspective; most of these assumptions can be relaxed in further studies.

In this paper, the air pollution caused by traffic can be measured by vehicular emissions. The proposed model can be stated as:

$$\mathbf{M\,1} \qquad \min_{E,R,f,\rho} Q = \sum_{\tau}\sum_{a} Q_{a,\tau} = \sum_{\tau}\sum_{a} h^{e}_{a,\tau} f_{a,\tau}\,, \tag{1}$$

subject to Lowry-based land use model constraints, and the annual-based traffic assignment constraints which are referred to by Lo and Szeto (2006). In this model, E,R,f,ρ represent, respectively, the vectors of service employment trips, residential trips, path flows, and tolls. Q is the overall vehicular emissions; $Q_{a,\tau}$ is the vehicular emissions due to traffic on link a in year τ; $h^{e}_{a,\tau}$ is the emission factor on link a in year τ, which is assumed to be given for all links. The variables affecting the value of $h^{e}_{a,\tau}$ are discussed in Nagurney (2000); $f_{a,\tau}$ represents the hourly traffic flow on link a in year τ. According to (1), the vehicular emissions on a particular link is the product of the link flow and the corresponding emission factor, and the overall vehicular emissions are obtained by summing all the vehicular emissions of each link over time.

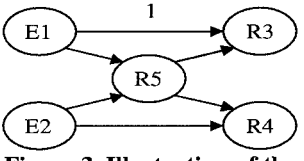


Figure 3. Illustration of the transport network.

Numerical studies

To illustrate the effect of tolls on the vehicular emissions and dynamic relationships between land use and transport systems, a simple scenario is set up. The scenario network is shown in figure 3 with five nodes, six links and six origin-destination (OD) pairs as well as the public transit system operated between each OD pair. The six OD pairs are E1-R3, E1-R4, E1-R5, E2-R3, E2-R4, and E2-R5 respectively. E1 and E2 represent employment zones 1 and 2 respectively. R3, R4, and R5 correspondingly represent residential zones 3, 4, and 5. The parameters adopted in this scenario are shown in table 1 below.

Table 1. Model parameters.

Planning horizon: 3 years.
Transport model parameters: The interest rate=0.03; the inflation rate=0.01; Value of time= 10/hour; The road network: Free flow travel times (mins): T_{13}= T_{24}=20, T_{15}= T_{25}= T_{53}= T_{54}=15; Car-specific constant=0.8; The public transit system: Travel times (mins): T_{13}= T_{24}=12, T_{14}= T_{23}=20, T_{15}= T_{25}=10; Transit fares (euros/person): f_{13}=f_{24}=2, f_{14}= f_{23}=2, f_{15}= f_{25}=1; Transit-specific constants: p_{13}=p_{24}=1.8, p_{14}= p_{23}= p_{15}= p_{25}=1.

Land use model parameters: Basic employment: E1=1000jobs, E2=800jobs; Population to employment ratio=5; Service employment to population ratio=0.1; Parameter for accessibility to the residential zone=0.8; Parameter for accessibility to the employment zone=0.6; parameter for the composite cost=1; Attractiveness: E1=E2=1000jobs; R3=R4=R5=1000houses; population and employment growth rates=0.04.
Environment parameters: Emission factors (liter/veh): e_{13}=0.7, e_{24}=0.6, e_{53}=0.4, e_{54}=0.3, e_{15}= e_{25}=0.5; maximum allowable vehicular emissions (liters/hour/link): S_{13}=S_{24}=1000, S_{53}=S_{54}=S_{15}=S_{25}=900.

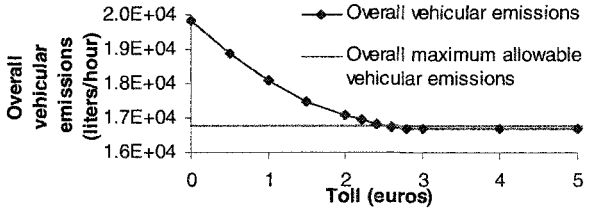


Figure 4. The overall vehicular emissions and the maximum allowable vehicular emissions over the entire planning horizon on the whole network.

In this scenario, we vary the toll level on link 1 from €0 to €5. The overall vehicular emissions on the entire network are plotted in figure 4. From this figure, we can observe that the tolls can be implemented to control the traffic emissions. However, the toll operator has to select the toll level carefully, because not any toll value can be set to achieve the overall vehicular emissions under the maximum allowable vehicular emissions. For example, in the scenario studied, when the toll is greater than €2.4, the overall vehicular emissions is under the maximum allowable vehicular emissions of 1.68E3 liters/hour. In particular, when the toll is greater than €3, the overall vehicular emissions remain at the same level. Thus, this problem raises an issue for the government of how the optimal toll can be chosen when implementing tolls on the transport network. The proposed model can be used for this purpose.

The implementation of tolls on links not only lowers the traffic emissions but also has a significant impact on the public transit system and the land use pattern. Figure 5 depicts the number of passengers on OD pair E1-R3 on the public transit system and the total transit revenue over the entire planning horizon. It is demonstrated from this figure that the toll implementation on link 1 causes travelers between OD

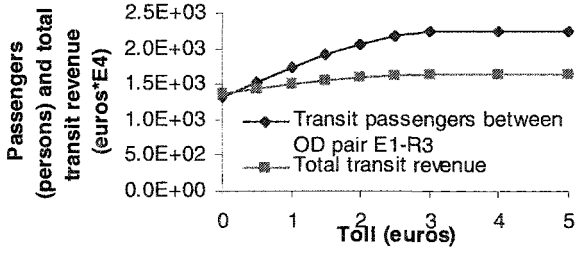


Figure 5. Number of passengers on OD pair 1-3 on the public transit system and the total transit revenue over the entire planning horizon.

pair E1-R3 to change their mode choice and take public transit instead of private cars. When the toll level rises, more passengers are attracted to the public transit system in order to decrease their travel costs. Beyond the toll level of €3, the number of passengers on transit between OD E1-R3 goes stable. Figure 5 also reveals the implementation of tolls can generate more transit revenue. When there is no toll on the transport network, the total transit revenue is only €1.35E7. However, after charging a toll of €3 on link 1, a 21.5% increase up to 1.65E7 on the revenue is obtained.

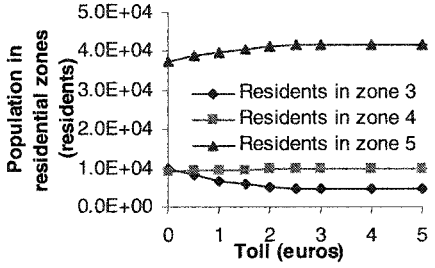


Figure 6. Total number of residents in zones 3, 4 and 5 over the entire planning horizon.

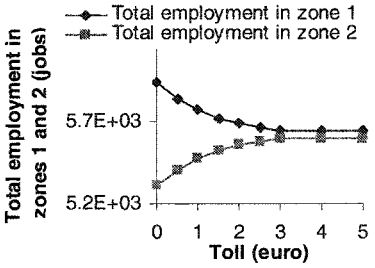


Figure 7. Total employment in zones 1 and 2 over time.

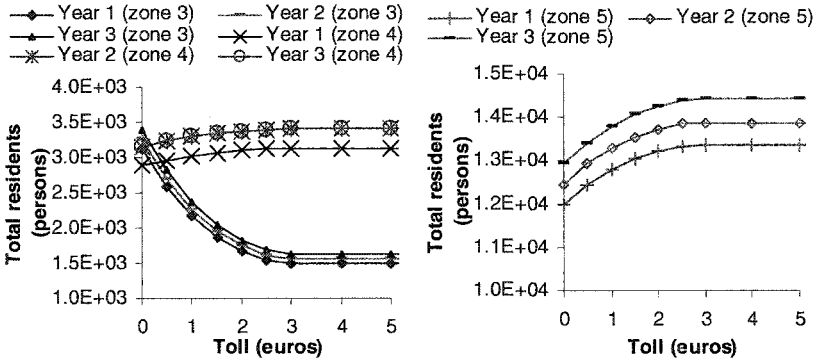


Figure 8. Number of residents in zones 3, 4 and 5 over time.

Figures 6, 7 and 8 illustrate that the implementation of tolls has an impact on the land use pattern. Figure 6 reveals the change of the total number of residents in zones 3, 4 and 5 due to the implementation of tolls over the entire planning horizon. Figure 7 shows the rearrangement of employment due to the change of tolls on the network. Figure 8 indicates the change of the number of residents in residential zones 3, 4 and 5 over time. From figure 8, it is revealed that there are dramatic changes in the numbers of residents in zones 3 and 5 over three years when tolls are ranged from €0-€3. These results demonstrate tolls can alter traveler's choice to choose their working and living locations, which has a strong implication on land owners' profits as the profits depend on the population and the size of the population-serving sector on their land.

Concluding remarks

In this paper, we propose an analytical model to determine the optimal tolls over time to control traffic emissions while capturing the land use-transport interaction. The model is a single level minimization program in which the objective function is to minimize the overall vehicular emissions over the entire modeling horizon. This paper also studies the impacts of the toll implementation on the environment in terms of vehicular emissions, the total transit revenue and the change of the number of residents and employment locations in a simple example. The results demonstrate that: (i) regarding to road pricing policy effectiveness, road pricing is a tool to mitigate the traffic emissions by diverting some travelers to take public transit. However, we need to set an optimal toll; (ii) road pricing contributes to generate more transit revenue; (iii) pricing has a great impact on the distribution of residents and employees in the land use system, which has an implication on land owners' profits. In this case, tolls must be carefully selected, and; (iv) not all toll values can affect the land use pattern and traffic pattern (not shown here due to space limitation). In the future, we can extend this study to consider the impact of road pricing on society such as social welfare and the land use owner's profits. It would also be interesting to model the transport network under the stochastic user equilibrium assumption in order to study the performance of the transport system effectively, as the model with the deterministic user equilibrium assumption does not represent the real world scenario.

Acknowledgment

This research is funded under the Programme for Research in Third-Level Institutions (PRTLI), administered by the Higher Education Authority. The authors are grateful for the constructive comments of the referee(s).

References

Hyman, G. and Mayhew, L. (2002). "Optimizing the benefits of urban road user charging." *Transport Policy,* 9, 189-207.

Lo, H.K. and Szeto, W.Y. (2005). "Road Pricing Modeling for Hyper-congestion." *Transportation Research,* 39A(7-9), 705-722.

Lo, H.K. and Szeto, W.Y. (2006). "Transport Network Improvement and Tolling Strategies: The Issue of Intergeneration Equity." *Transportation Research,* 40A(3), 227-243.

Mun, S., Konishi, K. and Yoshikawa, K. (2001). "Optimal Cordon Pricing." Proceedings of the Ninth World Conference on Transport Research, Seoul, Korea.

Nagurney, A. (2000). *Sustainable Transport Networks*, Edward Elgar Publishing, UK.

Sumalee, A. (2005). "Optimal implementation-path of road pricing schemes with time-dependent model." *Journal of the Eastern Asia Society for Transportation Studies,* 6, 624-639.

Zhang, X. and Yang, H. (2004). "The optimal cordon-based network congestion pricing problem." *Transportation Research,* 38B(6), 517-537.

Multi-system Interconnections Modes of Logistics Public Information Platform

Chen Youlin[1] and Yang Yang[2]

[1]Department of Transportation Engineering, Tongji University, P.O. Box 200092, City, Shanghai PRA; PH (86) 02165982272; email: chenyoulintina@126.com

[2] Department of Transportation Engineering, Kunming University of Science and Technology, City, Kunming PRA; PH (86) 02165982272; email: yytongji@163.com

Abstract: Logistics public information platform involves many aspects, and it is the key of the whole logistics system efficiently that how to ensure the information flow right, just-in-time, efficiently and fluently. Based on the analysis of logistics public information system demand and the function design, this paper analyzes three modes of multi-system interconnections of logistics public information platform and points out the suitable condition of each mode. The author proposes multilevel network system more suitable for the development object and function frame of logistics public information platform in China. At the same time this paper discusses the key technology of multi-system interconnections.

Key Words: Logistics; Public Information Platform; Multi-information System; Network Interconnection;

0. Introduction

Along with the development of computer and communication technology, the traditional management mode of modern enterprise is changing. The former independent, close operation and management mode can't adapt the fierce competition in modern commercial society. Owing to the more and more close cooperation and relating among the enterprises (or other organizations), the cooperative enterprise (or organization) turns to the cooperative group from the independent, close entity, forming a strategic network with common prospect unconsciously. Logistics has already become the new economical point of growth recognized by the various countries. The modern logistics involves the industry and commerce enterprise, the logistics enterprise, the bank, the insurance, the tax affairs, the customs, the examination quarantine, the foreign trade, the transportation, the information industries, the government and so on. Logistics remarkable characteristics are the information integration and the advanced information technology application, and its effective operation cannot leave the support of logistics information platform.

The introduction and the construction of logistics public information platform (LPIP) will be certainly to bring huge influence to each enterprise or the government department at present under the closed condition which has various information systems serving for oneself. Now the logistics characteristic in our country is "separate division" and "each dose things in his own way" in various local, various professions, various departments that lead to logistics information system construction redundantly, data nonstandard unable to exchange and interconnect. Through the public information platform construction, the government will obtain the macroscopic information of logistics development and management, the logistics enterprise to be possible to obtain the logistics demand information and the cargo tracking information, the industry and commerce enterprise to promptly obtain the logistics supplies information and so on. The planning and construction of LPIP has become an important content of modern logistics system construction.

LPIP involves many government management departments, logistics hub, logistics zone, logistics enterprises, the industry and commerce enterprise and so on. How to guarantee the information flow correct, prompt, highly effective, unobstructed is the key factor of the information platform construction. There are some problems that must be solved in information

platform planning. Such as, how to organize and process substantial sharing information, how to provide these information for relative departments to realize the coordinated work, how to provide these information for logistics enterprises to support function realization, and so on.

1. Demand Analysis and Function Design of Logistics Public Information Platform

The main bodies in the platform include the government, the enterprise and the customer and so on. The goal to construct such a platform is to provide information for the enterprise to carry on logistics management and market standardization management; integrate the whole society microscopic logistics resources and provide different information service to become a connection of logistics system as a link; communicate effectively between local professional management department and enterprise group through establishing urban logistics information platform to realize convenient, real-time logistics information exchange.

1.1 Analysis of Information Demand

Through three aspects to investigate and analyze information demand condition: logistics enterprise, the industry and commerce enterprise and the government sector.

(1) Logistics enterprise to logistics information service request: public logistics infrastructure resource information, logistics market demand information resource, logistics service operation information resource, other logistics information resource.

(2) The industry and commerce enterprise to logistics information service request: logistics supplier's material, logistics service transaction management, special and other value-added services and so on.

(3) The government sector to logistics information service request: region logistics operation basic data processing, supporting region logistics resource integration, analyzing region logistics and its plan and so on.

Logistics information demand situation and characteristic have decided the function system of LPIP. Information demand's characteristics are shown as following several aspects:

(1) Information dependence inside and outside logistics system. Logistics enterprise has very big dependence to the extraneous information of public logistics infrastructure and the transportation network. So public information platform existence can enhance to obtain logistics information and reduce information cost.

(2) Difference of logistics information demand. Logistics information demand of logistics enterprise, the customer and the government sector are different. The difference mainly manifests in: time difference, content difference and degree difference.

(3) Complexity of logistics information exchange. The logistics service integration involves many operating bodies including the customer. Each body's economy relations, the technical application, the enterprise culture and the information system module are different that make logistics information exchange complicate. Data exchange is to be done between various systems carrying on different operating modes in different enterprises and different subordination relation management system. Construction of data, the memory form and the interface agreement of each system are dissimilar, and it makes logistics data sharing and logistics resources integrating difficult for the lack of standardization.

(4) Limitation of logistics data sharing. Partial logistics enterprises operate to its specific customer as a closed system, and information sharing inside and outside logistics system is extremely limited.

1.2 Function Design of Information Platform

The LPIP's demand pattern decides the system structure. LPIP's system divides into four levels in vertical acquisition: the state-level, the provincial-level, the city-level and enterprise-level. In horizontal acquisition main points stress on administration and cooperation between same-level governmental agency and operational system. Every government function department in vertical acquisition exchanges and interconnects through logistics public information platform respectively which stresses mainly in the identical function between all levels of government sector and the operational system. Figure 1 gives an overview of the four-level system structure.

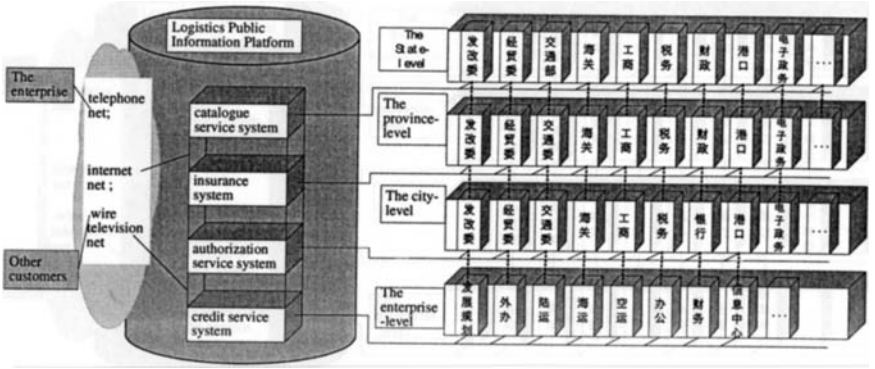


Figure 1. The structure of logistics public information platform

2. Interconnection Mode Analysis of Multi-system Information Platform

In order to achieve system goal of LPIP, it must choose the suitable network connection pattern and the information synthesis processing technology. At present many cities in our country have already started to plan logistics system, and some subsystems have also carried on intellectualization and information. For example some enterprises and government department have established their own logistics information system and management system. However these systems have two characteristics: one is systems between various enterprises and various departments belong to the relative equal subsystem; the other is multiplex management. Every system belongs to different departments involving different professions and enterprises, and each management way and starting point is various, as a result it causes logistics information system construction redundantly, data nonstandard unable to exchange and interconnect.

The overall goal of public information platform is to satisfy the different level participant's function and information demand in logistics system, through collecting and processing public logistics data and exchanging public information to provide information support for the enterprise and the government. At present for the reason that some enterprises and government departments have already constructed logistics information system and management system, how to use the existing resources efficiently through exchanging and interconnecting these systems and integrate platform functions are two important questions which are worth thoroughly studying in logistics plan. Because of the present network construction patterns, multi-system interconnections modes in logistics public information system may divides into three kinds of forms: interconnection architecture, C/S-based (Customer/Server) network architecture and multi-level network architecture.

2.1 Interconnection Architecture

The multi-system interconnection condition shows in Figure 2. Each information system (or platform) in logistics public information platform must exchange information and data, coordinate data pool interface, data type and data transfer agreement separately with other systems (or platforms). This connection characteristic is real-time and high connectivity to obtain, exchange and process information between multi systems. Because logistics information system of each enterprise and government sector has been proposed and constructed from their own condition respectively, each system goal and function has much difference and limitation. Therefore useful information in different information systems will be carried on data fusion and resource sharing, it will lead to a large amount of work such as information coordination and data accessing. Moreover if we use this method to establish a perfect logistics system platform the

overall structure will exist in many problems.

Therefore multi-system interconnection like this kind only suits to carry on between several similar systems. But logistics public information platform involves multitudinous enterprises and different sectors, it is not suitable to take this kind as network construction mode of logistics public information platform.

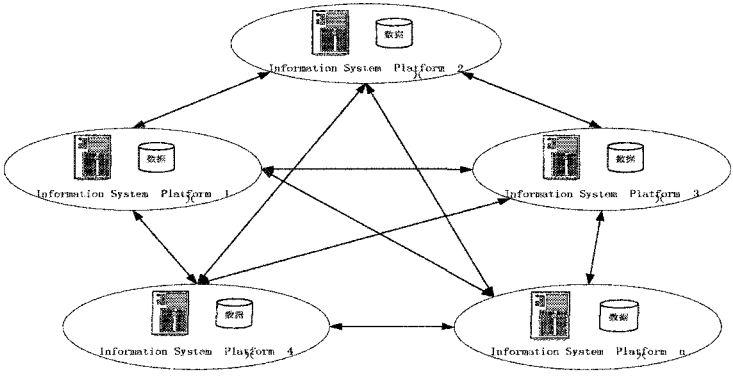


Figure2. The interconnection architecture

2.2 C/S-based (Customer/Server) Network Architecture

C/S-based network multi-system information connection is shown in Figure 3. C/S-based network form shifts coordination between each information system and management system between these systems and information platform. Information exchanging between information system and management system will be realized through logistics public information platform.

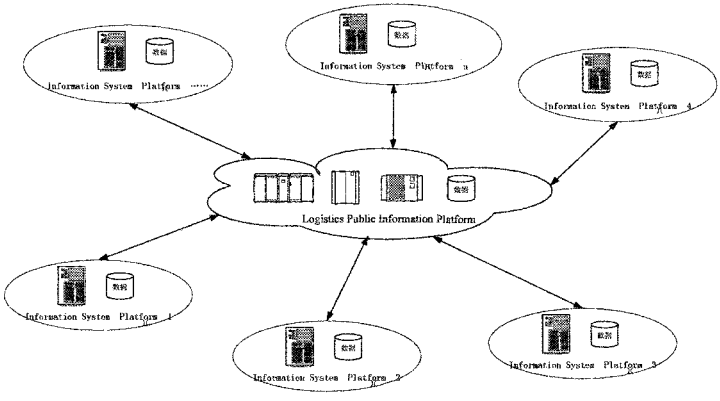


Figure3. The C/S-based (Customer/Server) network architecture

The existing information and management system must coordinate with logistics public information platform which is useful for the uniform plan, comprehensive information, and fully using the existing system. The most important key work needs to construct an open uniform logistics public information platform to integrate existing systems and exchange data and information between different systems.

C/S-based network pattern could simplify the network architecture, be good to use and expand the existing system, save construction cost, and reduce the difficulty of system construction.

2.3 Multi-level Network Architecture

Multi-level network connection form is sketched in Fig 4. The multi-level network connection form is suitable for logistics public information platform system structure in vertical acquisition. In order to reduce logistics public information platform center service system load and adapt logistics public information platform level structure (see Fig 1), strengthen real-time data accessing between each system, it may construct multi-level network structure according to the different regions. The multistage network form is adaptable for state-level logistics public information platform construction in China, but its structure is more complicate and needs to construct logistics information platform from the bottom layer. At the same time it still needs to solve many essential technologies which construct the whole system.

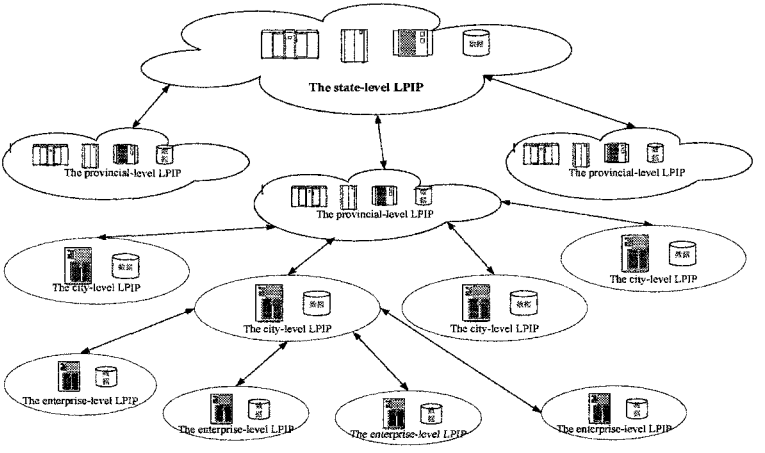


Figure4.The multi-level network architecture

3. The Key Technology Analysis of Multi-system Interconnection

3.1 Network Connection between Each System

Although this paper discusses multi-system interconnection of logistics information platform, it is only limited from the macroscopic point of view in the research. To the microscopic point of view, it is also needed to study the concrete network connection method how to connect enterprise interior network in the information platform, local area network in the electronic commerce system, VPN metropolitan area network, Internet and the enterprise exterior network together and which kind of communicating protocol will be chosen. At the same time logistics public information platform is a network integrative system which provides communication environment for data exchange through interconnecting different kind of form data together.

3.2 Information Source Connection and Information Standardization

The operation of LPIP needs a large number of data and also produces a large number of data. Highly effective data management is an important guarantee of the public information platform. Logistics public information platform itself includes many subsystems, and each subsystem all exceptionally needs to collect and exchange various kinds of correlation data from other subsystems and provide its service function by processing data. Therefore the whole sharing information platform construction should establish on the consolidated information source connection and normative information standard, based on that the comprehensive benefit and the

superiority of this platform can be reflected. Although it can be able to realize resources sharing to use the different connection technologies, but the appropriate connection technology is one of the key factors to guarantee the overall system effectively. At present EDI as a way is often used to transmit commercial certificate on electronic commerce from one application system to other application system between enterprises. But it also exists in some limitation. To choose which kind of transmission standard to communicate data is urgent to solve because it will be a crucial technology in logistics public information platform.

3.3 Data Fusion

Because logistics public information platform involves multitudinous subsystems and needs to integrate information management systems which have been already constructed at present and each subsystem has a set of independent, rounded data structure, therefore it will have to face many problems, such as many different kinds of data and different data structures in order to conformity subsystems. How effectively to organize these data to provide for the public information platform application and sharing is a technology which is urgent to be solved.

3.4Data Warehouse and Data Mining

According to analysis above, logistics public information platform contains many kinds of data and an extraordinary vast number of data. It can't be solved to storage and classify data if only to put these data together simply. Therefore it needs widely to use the technology which combines "data warehouse + on-line analytical processing + data mining". In the generalized concept data warehouse is one kind of systematization solution scenario, it including three aspect contents: Data warehouse, On-line analytical processing and Data mining. There are obviously some relations and complementarities between three kinds of technologies. If we combine them together to form a framework of decision support system, namely DW+OLAP+DM, it may be able to bring fully into play their ability. What is mentioned above is the technical difficult problem which should be urgently solved in logistics public information platform.

4. Conclusion

Logistics public information platform is a multi-system including many industry and commerce enterprises, logistics enterprises and the government sectors. Different participant has different demand for the platform. This paper analyzes three modes of multi-system interconnections of logistics public information platform which are possible to be adopted through demand analysis and functional design, and points out the suitable condition of each mode. Multi-level network system is more suitable for the development object and function frame of logistics public information platform in China. At the same time we discuss the key technology of multi-system interconnections. The development of logistics public information system in China could be able to make considerable progress only if these key technologies could be researched and solved.

References:

[1] Liu Xinjing, Dai He, Yang Dongyuan(2001). "Analysis of logistics public information platform" . *Issues in Traffic and Computer*, 19(1),34～38.

[2] Cui Nanfang, Liu Yinzi, Zhao Zhenfeng(2004). "The system design of regional public logistics information platform" . *Issues in Science Development and Countermeasure*. 22(8),142～144.

[3] Dong Qianli, Yuan Yi(2002). "Research into the functions and structures of regional integrated logistics information platform" . *Issues in Journal of Transportation Systems Engineering and Information Technology*, 2(1),74～78.

This page intentionally left blank

Subject Index

Page number refers to the first page of paper

This page intentionally left blank

Author Index

Page number refers to the first page of paper