Transportation and Development

Innovative Best Practices 2008

Proceedings of the First International Symposium

Edited by
Louis F. Cohn, Ph.D., P.E.

ASCE

T&DI
TRANSPORTATION AND DEVELOPMENT INSTITUTE

TRANSPORTATION AND DEVELOPMENT INNOVATIVE BEST PRACTICES 2008

PROCEEDINGS OF THE FIRST INTERNATIONAL SYMPOSIUM

April 24–26, 2008
Beijing, China

SPONSORED BY
The Transportation & Development Institute (T&DI)
of the American Society of Civil Engineers

China Academy of Transportation Sciences

EDITED BY
Louis F. Cohn, Ph.D., P.E.

Published by the American Society of Civil Engineers

Library of Congress Cataloging-in-Publication Data.

Transportation and development innovative best practices 2008 : proceedings of the first international symposium, April 24–26, 2008. Beijing, China / sponsored by the Transportation & Development Institute (T&DI) of the American Society of Civil Engineers [and] China Academy of Transportation Sciences ; edited by Louis F. Cohn.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-7844-0961-9
1. Transportation engineering--Congresses. 2. Transportation--Technological innovations--Congresses. I. Cohn, Louis F. (Louis Franklin), 1948- II. Transportation & Development Institute (American Society of Civil Engineers) III. China Academy of Transportation Sciences.

TA1005.T69573 2008
388--dc22 2008009030

American Society of Civil Engineers
1801 Alexander Bell Drive
Reston, Virginia, 20191-4400

www.pubs.asce.org

Any statements expressed in these materials are those of the individual authors and do not necessarily represent the views of ASCE, which takes no responsibility for any statement made herein. No reference made in this publication to any specific method, product, process, or service constitutes or implies an endorsement, recommendation, or warranty thereof by ASCE. The materials are for general information only and do not represent a standard of ASCE, nor are they intended as a reference in purchase specifications, contracts, regulations, statutes, or any other legal document. ASCE makes no representation or warranty of any kind, whether express or implied, concerning the accuracy, completeness, suitability, or utility of any information, apparatus, product, or process discussed in this publication, and assumes no liability therefore. This information should not be used without first securing competent advice with respect to its suitability for any general or specific application. Anyone utilizing this information assumes all liability arising from such use, including but not limited to infringement of any patent or patents.

ASCE and American Society of Civil Engineers—Registered in U.S. Patent and Trademark Office.

Photocopies and reprints.
You can obtain instant permission to photocopy ASCE publications by using ASCE's online permission service (http://pubs.asce.org/permissions/requests/). Requests for 100 copies or more should be submitted to the Reprints Department, Publications Division, ASCE, (address above); email: permissions@asce.org. A reprint order form can be found at http://pubs.asce.org/support/reprints/.

Copyright © 2008 by the American Society of Civil Engineers. All Rights Reserved.
ISBN 978-0-7844-0961-9
Manufactured in the United States of America.

Preface

The First International Symposium on Transportation and Development – Innovative Best Practices (TDIBP 2008) was conceived as a forum addressing innovation and practice in transportation planning and development activities, as defined by the American Society of Civil Engineers' Transportation and Development Institute (T&DI) and the China Academy of Transportation Sciences (CATS). The goal of the symposium was to showcase innovative and best practices in transportation and development across three tracks. The tracks were: Improving Integrated Transportation and Development (Track A); Operations and Safety (Track B); and Asset Monetization, Security, Highway Design and Pavements (Track C).

TDIBP 2008 represents an initiative to enhance cooperation between T&DI and its counterparts in China, namely CATS. Such international cooperation is a top priority of both organizations. It is hoped that TDIBP 2008 is only the first in a series of symposia and other cooperative ventures by T&DI and CATS.

The impetus for the symposium came from the T&DI Board of Governors in 2005, under the leadership of then President Eva Lerner-Lam. President Lerner-Lam is the only member to head a major unit of ASCE while residing outside of the United States. She currently resides in Beijing, although she maintains a home in New York. The T&DI "China Initiative", under President Lerner-Lam's leadership, resulted in a Strategic Plan for a series of activities, of which this symposium is a major component.

T&DI is very much indebted to the leadership shown by TTEC (Transportation Technology Exchange Center) of CATS (China Academy of Transportation Sciences). TTEC Director Haiqing Wang and Ms. Juana Zhao were instrumental in making the symposium a success, and demonstrated a great vision regarding future collaboration with T&DI.

In 2005, a Steering Committee for this symposium was organized. The members of this committee are listed below.

Eva Lerner-Lam (Chair)
Marsha Anderson-Bomar
Peter Lai
Rachel Liu
Juana Zhao (CATS)
Kumares C. Sinha (Honorary Chair)
Louis F. Cohn
Jon Esslinger (T&DI Director)
Kam Movassaghi

A call for papers was issued in January, 2006, listing the theme of the conference. Approximately 400 abstracts were submitted for consideration. These abstracts were reviewed by a Program Committee, and ultimately 86 abstracts were accepted as written papers in the proceedings.

Preliminary versions of the papers were submitted to the Program Committee and they were peer reviewed by members of T&DI. In addition, numerous T&DI members volunteered to assist the Chinese authors in bringing the English quality up to ASCE publication standards. These papers and presentations were organized into a formal program by the Steering Committee, assisted by members of the T&DI Board of Governors and ASCE staff.

Louis F. Cohn, Ph.D., P.E., F.ASCE
Professor of Civil and Environmental Engineering
University of Louisville
Member, TDIBP 2008 Steering Committee and
Proceedings Editor

Contents

This page intentionally left blank

Implementing Transportation Infrastructure Improvements to Support Development in Pennsylvania, United States

Larry I. Bankert, Jr., P.E., PTOE[1] and Stephen R. Thompson2

[1] Senior Traffic Engineer at Pennoni Associates Inc., Mechanicsburg, PA (Pennoni Associates Inc. is headquartered in Philadelphia, PA). E-mail: lbankert@pennoni.com.

[2] Senior Designer at Pennoni Associates Inc., Mechanicsburg, PA, and Adjunct Instructor at Messiah College, Grantham, PA. E-mail: sthompson@pennoni.com.

Introduction

Transportation infrastructure is critical to economic and land development in industrialized and industrializing nations. However, dispute over the responsibility for funding and implementation of off-site transportation infrastructure improvements frequently occurs. While public agencies often desire development, they bear the responsibility of ensuring that it does not degrade the mobility and safety of the nearby surface transportation network for the traveling public. Private developers seek adequate access to their property, but at times feel there is unfair burden on the private sector to fund and implement off-site improvements. The purpose of this paper is to examine the process for the implementation of off-site transportation improvements associated with land development in the Commonwealth of Pennsylvania, United States, and compare it to best practices in the United States.

Currently in Pennsylvania there are two regulatory means by which local and state agencies may require private developers to fund off-site improvements. At the state level, the Pennsylvania Department of Transportation (PennDOT), through their permitting process, may require developers to fund and construct improvements as determined from the preparation of a traffic impact study (TIS). At the local level, municipalities may adopt a traffic impact fee ordinance. This ordinance allows municipalities to assess developers a fee on a per vehicle trip basis to construct off-site improvements.

However, few municipalities have enacted traffic impact fee ordinances due to its cumbersome, costly process. In addition, the PennDOT TIS / permitting process is not an exact science, in terms of determining an equitable distribution of improvement implementation responsibility. This paper describes the current process for implementing off-site transportation improvements associated with land development, examines their pros and cons, and compares it with best practices in the United States.

Pennsylvania Approach

Pennsylvania is a commonwealth with 67 counties and approximately 2,600 municipalities. There are generally four types of municipalities, each of which have distinct but similar, governing regulations. The four types are Cities, Boroughs, Townships of the First Class, and Townships of the Second Class. The Commonwealth has a Municipal Planning Code (MPC) that sets the framework for municipalities to establish zoning ordinances and subdivision and land development ordinances (SALDOs).

Municipal engineers and public administration consultants we interviewed estimate that of the approximately 2,600 municipalities in Pennsylvania, 80 to 90 percent do have zoning ordinances or SALDOs. For those municipalities that do not have zoning ordinances or SALDOs, their county's zoning and subdivision and land development requirements govern. However, these county-wide land use requirements and ordinances are often general in scope, as they must accommodate a greater range of land use needs largely because of the area they cover, while local municipal ordinances may be specific to the relatively few land uses within the particular municipality. Unfortunately, this deference to sometimes more ambiguous higher government level control can become an issue for rapidly developing municipalities such as Hamilton Township, Franklin County, Pennsylvania, which will serve as an example for the uses of this paper.

Hamilton Township is located in a rural area of Pennsylvania, adjacent to the small urban Borough (town) of Chambersburg with several larger urbanized areas located within 100 miles of the Township, such as the Washington, D.C. / Baltimore metro area and Harrisburg, Pennsylvania area. There are two major transportation corridors in the vicinity of the Township - US Route 30 and Interstate 81. The Township has a growing population, having increased 16 percent during the decadal period 1990 to 2000. The estimated population for year 2007 is 9,923 persons. Approximately 2,100 new homes are expected to be built in Hamilton Township in the next 10 years, which could add an additional 20 percent to the population. Also, significant commercial development, including big box retail operations such as Wal-Mart, is expected to support the growing population.

Currently, Hamilton Township has no zoning ordinance, meaning developers can build what they want, where they want within the Township. The Township does have a SALDO that sets requirements on the design of the development. The SALDO includes requirements for a TIS. The TIS scope may include off-site intersections or roadway segments. However, Pennsylvania's MPC does not allow a municipality to impose the requirement for off-site improvements where a development meets the zoning and SALDOs. And like most municipalities in PA, Hamilton Township does not have an impact fee ordinance. Therefore, off-site improvements are often negotiated between the developer and the Township.

Municipalities were given the authority to enact Impact Fee Ordinances in December, 1990. However, the municipal engineers, traffic engineers, and public administration consultants we interviewed suggest relatively few municipalities have taken this step. The process can take one to two years to complete and may cost \$200,000 to \$300,000 for a municipality such as Hamilton Township to enact the ordinance. Additionally, there are significant restrictions on applicable infrastructure improvements for use of the funds.

Municipalities (and PennDOT) have tight budgets and difficulty funding improvements themselves. In most cases, improvements can be funded in one of four ways by municipalities using public funds:

- General Fund – General funds are funds available to the municipality through local taxes but are usually needed to provide other necessary services within the municipality.
- Liquid Fuels Funds - Liquid fuels funds are provided by PennDOT (essentially a pass-through of federal and state gas tax monies) that uses a formula based 50 percent on lane miles of roadway and 50 percent on population using the latest decadal census figures. These funds are generally barely enough to maintain the existing transportation infrastructure and rarely provide the luxury to fund capital improvements to support development.
- Transportation Improvement Program (TIP) – The TIP is a four-year planning program administered by the governing Metropolitan Planning Organization (MPO) or Rural Planning Organization (RPO). All projects using state or federal funds must be on the TIP. Competition to be on the TIP is fierce, as there are always more projects needing funding than funds available. Therefore, unless a development under consideration is large with the potential of compelling economic arguments for inclusion on the TIP, such as significant job creation, TIP support of private development is rare. Lobbying efforts are generally needed. While there are other funding programs through which project money is actually distributed, the TIP essentially determines eligibility.
- Earmark – Earmark funds are funds that are set aside by legislation for a specific project. In general, as noted above, the MPO or RPO sets the priority for projects in a fiscally-constrained program. Earmarks provide funds for a specific project above the TIP. In most cases, significant lobbying efforts are required to secure a project earmark.

None of these sources provides enough funding to use as a foundation for the significant capital improvements that may be needed to support development. For example, through the preparation of a Capital Improvement Plan (CIP), Hamilton Township determined the need for a $30 million transportation improvement program to fix existing deficiencies in the system and support the anticipated development which is expected to occur mostly in the next ten years. Looking at its impact in simple terms, this is a $3 million per year program. Examining the budgets of second-class townships around Pennsylvania, we observed that township budgets tend to cluster in three ranges, with rural townships averaging around $5 million per year. The lowest budget we encountered was $2.6 million. For rural municipalities in Pennsylvania similar to Hamilton Township, a $3 million per year transportation improvement program, at worst, can be more than its annual budget, but is more likely to be around half its annual budget – certainly a prohibitive percentage.

Besides the local impact fee ordinance, another means by which developers can be required to make improvements is through PennDOT's Highway Occupancy Permit (HOP) process. HOPs are required by any private or public entity who must work within PennDOT right-of-way. The most common requirement for a HOP is driven by the need for access to a state roadway. In these cases, the developer may be required to prepare a TIS to determine off-site impacts. Pennsylvania has guidelines for the preparation of TISs that include guidance on scoping a TIS. While the TIS does provide a technical basis for required improvements for an individual development, it is lacking from a planning perspective in a number of ways:

- TISs usually contain no or little multi-modal, pedestrian/bike, and public transportation discussion. And even more rarely make an effort to address their future needs.
- TISs are generally a snapshot of a small study area and do not lend themselves to give an assessment of a region and impacts of other future developments. For this reason, they sometimes create 'piecemeal planning.'
- In high growth areas, several TISs may be produced simultaneously, looking at the same intersection(s) with little or no coordination.

In both the local process (when an impact fee ordinance is not in place) and in the HOP process, often the 'last developer in' and sometimes the 'first developer in' pays for improvements in a growing region. For example, each development in an area may use the available capacity at a particular intersection and the developer who comes in when the capacity begins reaching unacceptable levels of service will be burdened with making the necessary improvements. Vice versa, the 'first' developer in a region may be required to make an improvement that adds capacity beyond their needs and effectively creates capacity for future developers in the region.

Practices in Other Regions of the United States

The approach toward local transportation funding tends to be generally similar across regions of the US, and when variations do appear, they appear in the details of implementation. The details of this study have, thus far, focused on tying land use planning and development to transportation planning in Pennsylvania. Pennsylvania is unique among the States in terms of the relative power between its County governments and its municipal governments. Public services and control that tend to remain in the hands of County governments in States south and west of Pennsylvania are often found in municipal governments in Pennsylvania. For comparison to the Pennsylvania experience, transportation infrastructure improvements implementation and funding was examined in the states of Florida in the south, Indiana in the US heartland, and California in the west.

Generally, these three states are similar to each other and to Pennsylvania in that each has enacted legislation guiding the imposition and collection of impact fees for transportation funding by local governments. The details of how these fees are imposed, and by what level of government, do vary, though.

Indiana State Law requires that comprehensive plans contain a transportation element: the result is that transportation issues are necessarily a part of comprehensive planning. An example is offered in the Indianapolis Long Range Transportation Plan Update - Status Report (1993). In Indianapolis/Marion County, Indiana (Indianapolis MPO), comprehensive plan updates include transportation information such as existing and future functional classifications for all thoroughfare plan segments and existing and projected levels of service. Examining the consistency between land use plans and the transportation components of comprehensive plans, then, necessarily focuses on the use of transportation impact studies (The Indianapolis Metropolitan Planning Organization, 1993).

Florida appears to be more rational, or structured, in its collection of impact fees in that Counties in Florida have fee schedules based upon development units, such as residential, commercial floor space, etc. – even fees on holes at a new golf course or rooms at a hotel. For example, Alachua County's (Florida) Transportation Impact Fee Ordinance presents a fee schedule enumerating the fees required based use (residential fees are calculated per 1,000 square feet, motels are calculated by room).

However, a developer may choose to try to circumvent the fee schedule by submitting a transportation engineering study and an economic documentation study to the impact fee administrator, which demonstrates that the transportation impacts generated by that particular development are less than the average impacts quantified in developing the fee schedule. By ordinance, the impact fee administrator "shall" accept the documentation for consideration, but is not required to collect an impact fee based upon the submitted documentation. If not satisfied by the submitted documentation, the impact fee administrator may require the impact fees be paid according to the fee schedule (*Ordinance 06-30 [Transportation Impact Fee]*, 2006).

In the event the documentation submitted by a developer is deemed satisfactory to the impact fee administrator, the impact fees are then calculated using the trips generated, or other impacts, from the accepted study and the formulas set down in the Ordinance. A developer is not free, under this Ordinance, to negotiate the fee, or even to suggest what fee might be justly applied. A developer is able only to document trips generated and associated costs; the fees imposed are still determined by ordinance, albeit less than the fee schedule with accepted documentation.

California's impact fees collection is also well-structured, combining impact fees for many services, such as law enforcement, sewerage, as well as transportation infrastructure - and based upon standard dwelling sizes – resulting in a total of impact fees per house. In 1987, the California Legislature adopted Assembly Bill 1600 - the Mitigation Fee Act - that established a uniform process for formulating, adopting, imposing, collecting, accounting for, and protesting developer impact fees to pay for the expansion of infrastructure and services necessitated by development (Born, 2001). Monterey County developed its fee schedule by polling agencies of other jurisdictions to learn the fees they collected, and produced a comparative fee schedule as part of its General Plan Update process (Born, 2001).

The structured fee schedule approach of Florida and California appears to separate transportation planning from a direct relationship with transportation funding. That is, the planning seems to have occurred at a different governmental level or agency and prior to the presentation of specific plans by developers, who bring the fees based upon predetermined fee schedules. In the case of the Florida county presented above, as well as the Indiana county presented earlier, the planning occurs at the County level as part of periodic comprehensive plan updates. The California county examined appears to have built its impact fees schedule from comparative research. Developers need only calculate their impact fees by multiplying their "units" of development by the appropriate figure in a schedule, potentially limiting variance in impact fee administration.

Conclusions

Implementing infrastructure improvements to support development in Pennsylvania is often lacking due to its disconnection to the land use planning process, the number and levels of governmental bodies, and their disjointed authority to impose improvements. Specifically, there are several issues related to funding transportation infrastructure improvements to support development in Pennsylvania:

- In growing municipalities, the traditional public sources of funding cannot keep pace with the required funds to implement transportation improvements to support development. Therefore, developers or the private sector investment is necessary.
- The shear number of municipalities in Pennsylvania and their flexibility for planning leads to disjointed land use and transportation improvements.
- Relatively few municipalities in Pennsylvania appear to have enacted impact fee ordinances early, possibly due to its cumbersome, costly process and its limitations on the use of the funds.
- TISs, the technical foundation for most development-related improvements, are not comprehensive, do not address all modes of transportation, do not coordinate adjacent development in all cases, and sometimes create inequities in the responsibilities for improvements.
- To overcome these issues, municipalities often negotiate the final improvements to be paid for or constructed by the developer, and the PennDOT HOP process can be contentious because of its inexactitude.

Best practices around the United States show transportation planning being done at the County and/or MPO level, rather than the municipal level, prior to development, and in States like Florida and California, where the imposition and collection of transportation impact fees is more instrumental, the fees are collected and used by the Counties, with the impact fees used as a means to ultimately fund improvements as development occurs.

References

Born, K. (2001). *Development Impact/Capacity Fee Study*. <http://www.co.monterey.ca.us/gpu/Reports/Impact%20Fees%20-%20FINAL.pdf> (Aug. 8, 2007)

Guide for the Preparation of Traffic Impact Studies. (2002). State of California Department of Transportation. <http://www.dot.ca.gov/hq/traffops/developserv/operationalsystems/> (Aug. 8, 2007)

Indianapolis Long Range Transportation Plan Update - Status Report: Review of ISTEA Metropolitan Transportation Planning Factors. (1993). The Indianapolis Metropolitan Planning Organization. <http://ntl.bts.gov/DOCS/dec93.html> (Aug. 10, 2007)

Marion County Transportation Impact Fee Fact Sheet. (2004). Marion County Florida - Board of Commissioners. <http://www.marioncountyfl.org/BU341/BU_Impact_Fees.htm> (Aug. 10, 2007)

Ordinance 06-30 (Transportation Impact Fee). (2006). Alachua County Florida – Board of County Commissioners. <http://growth-management.alachua.fl.us/formsdocs/IF_Transportation.pdf> (Aug. 9, 2007)

Transportation Impact Fees. (2007). Polk County Florida – Board of County Commissioners. <http://www.polkcounty.net/county_offices/bldg_inspections/transFees.aspx> (Aug. 9, 2007)

Transportation Development and Globalization Trends: A Comparative Global Assessment

Manouchehr Vaziri[1] and Ali Rezaee[2]

Abstract

Globalization is shaping a new socio-economic order with profound, and still unfolding, implications for transportation development. A prerequisite for national competitiveness in the global market place is efficacious transportation when its recent technological advances have created an unprecedented rise in mobility and accessibility. Developments of efficient and effective transportation infrastructure and services are among the principal driving forces for the globalization process. Using international databanks, this paper describes an attempt to shed some light on national trends of globalization and transportation, and their relationships. Deploying a comparative macroscopic approach and using relevant time-series statistics, the study provided some quantitative clues of the possible interrelationships between globalization and transportation.

Introduction

Globalization is a term used to describe the acceleration and intensification of economic interactions through increasing internationalization. The process of globalization has led to phenomenal surges of international interactions. Globalization is shaping a new socio-economic order with profound implications for transportation development (Beck, 2000; Vaziri and Rasafi, 2006). A prerequisite for national competitiveness in the global market place is efficacious transportation when its recent technological advances have created an unprecedented rise in mobility and accessibility. Developments of efficient and effective transportation infrastructure and services are among the principal driving forces for the globalization process. In recent decades, the costs of transportation have fallen significantly, which have immensely fueled globalization. The relationships between globalization and transportation, as shown in Figure 1, are two directional even though inadequately understood. The literature reflects some of these relations, but often their characterizations are not based on empirical evidences (Dimitriou, 2005; Nijkamp, 2003; UNESCAP, 2005). This paper describes an attempt to shed some light on national trends of globalization and transportation, and their relationships.

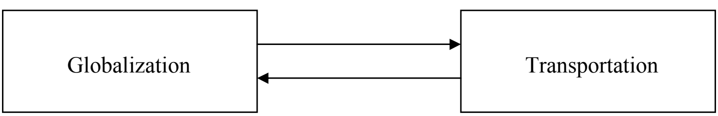


Figure 1 Globalization and transportation interactions

[1]Professor, Civil Engineering Department, Sharif University of Technology, Azadi Avenue, P.O. Box 11365-9313, Tehran, Iran; Tel.: +98-21-22282507, Fax: +98-21-22284311, Email: manouchehrvaziri@yahoo.com.
[2]Graduate student, Civil Engineering Department, Sharif University of Technology, Azadi Avenue, P.O. Box 11365-9313, Tehran, Iran.

Database

The limited study resources confined the data collection to information gathering from the international databanks. The data reliability bore the assumption that for the accessible databases, definitions were similar and comparable through time and among countries. For the period of 1970 to 2000, relevant time-series national transportation and globalization statistics were collected for 207 countries. Initially, information regarding 80 globalization variables and 25 transportation variables was extracted from centralized databases (World Bank, 2004). After several stages of database screening, evaluation and refinement, and due to data incompleteness and missing, a subset of the initial variables was selected for detailed analysis. The 23 finally selected globalization variables reflected cross-border flows, cooperation and exchanges for products, people, resources and information. These were globalization characteristics often used in other studies (Nijkamp, 2003; Vaziri and Rasafi, 2006). For the period of 1970 to 2000, the percent missing cases for the selected globalization variables had a range of 19% to 98%. The 18 finally selected transportation variables reflected national passenger and freight demand and supply statistics for different modes. For the period of 1970 to 2000, the percent missing cases for the selected transportation variables had a range of 31% to 96%. The selected variables were neither standard nor unique; nevertheless they reflected the basic accessible national indicators of globalization and transportation. The univariate statistical analysis for the selected variables shed light on the database cross-sectional and time-series variability. The mean and coefficient of variation for 4 globalization and 8 transportation variables are shown in Table 1. The missing information for year 2000 was around twice as of year 1995.

Table 1 Descriptive analysis for 12 variables

Variable	Description	Mean 1995	COV 1995	Mean 2000	COV 2000
GNIR	International reserve in million $	8077.4	2.5	15450.6	2.7
GITR	International tourism receipts in million $	2168.9	2.9	6483.3	3.2
GRAT	Radio receivers in million	12.11	4.3	23.71	4.0
GTRD	Trade in million $	72340.27	2.8	102100.7	2.4
TTLS	Travel services in million $	2310.1	3.0	2280.2	3.1
TTTS	Transport services in million $	2241.2	2.5	2382.4	2.9
TRPV	Paved roads in km	77020.1	4.3	291248.2	4.0
TRQT	Freight in million ton-km	63898.63	3.0	204816.05	3.2

TPSC	Passenger cars	2874203	4.0	6805921.5	4.1
TADE	Aircraft departure	111852.17	5.5	123570.29	5.4
TATR	Air transport passenger	8147083	5.3	8053154.9	5.2
TATF	Air transport freight in million ton-km	519.56	3.6	672.79	4.4

The univariate statistical analysis shed light on the database cross-sectional and time-series variability. The mean values of the 41 selected globalization and transportation variables showed growing trends during the 30-year period. Nevertheless, for individual years, large cross-sectional variability became evident by COV in a range of 2 to 6 among countries.

Multivariate Analysis

To develop an understanding of the interrelationships among the database variables, as a first step, pair-wise correlation was performed. For each year, the 41x41 correlation matrix revealed many statistically significant cross sectional correlations. Furthermore, the correlation analysis of variables' changes during the 5-year periods was also performed. For selected years and 5-year periods, the results of correlation analysis of globalization with transportation variables are summarized in Table 2. The first row shows the year or the 5-year period. The second row, correlation computed, shows the percent of computable correlations as were restricted by missing information. The third row, significant correlation, shows the percent of positively and statistically significant correlation. The last row shows the percent of positively significant correlation among computed correlations. The pair-wise correlation analysis provided a preliminary understanding of the cross-sectional and time-series relationships between the selected variables. The cross-sectional correlations between globalization and transportation were often more than 60%.

Table 2 Percent correlation between globalization and transportation variables

Date	70	75	75-70	80	80-75	85	85-80	90	90-85	95	95-90	00	00-95
Correlation computed	25	21	22	46	22	44	39	77	43	90	77	63	55
Significant correlation	15	15	11	29	14	29	21	50	26	64	40	38	23
Significant/ Computed	61	68	47	62	65	66	53	65	60	72	52	60	42

The statistically significant correlations suggested the possibility of developing models among globalization and transportation variables. Assuming globalization as dependent to transportation up to n periods time-lag, a general relationship such as Equation 1 can be presented:

$$G_t = f(\mathbf{G}_{t-1}, \ldots, \mathbf{G}_{t-n}, \mathbf{T}_t, \ldots, \mathbf{T}_{t-n}) \quad (1)$$

Where G_t is a globalization variable at time t, f is the function symbol, $\mathbf{G}_{t-1}$ is the vector of globalization variables at time t-1, $\mathbf{G}_{t-n}$ is the vector globalization variables at time t-n, $\mathbf{T}_t$ is the vector of transportation variables at time t, and $\mathbf{T}_{t-n}$

is the vector of transportation variables at time t-n. Using globalization variables as cross-sectional dependent to transportation variables, eleven typical functional forms were evaluated for year 1995 and all years together. For simple models, the functional forms were linear, growth, compound, quadratic, logarithmic, cubic, S-shape, exponential, inverse, power, and logistic. More than five hundred models were developed and evaluated. The statistically significant models, based on f-test for the model, t-test for the coefficients, and at a level of 0.05, for the largest coefficient of determination were identified. Based on the adjusted coefficient of determination, the linear form often was found as good as the other functional forms. Statistically significant relation was not observed for 3 out of 23 globalization variables namely, GNFA, net foreign assets, GFFA, financing from abroad, and GEBG, external balance on goods and services. Four out of 18 transportation variables did not appear in any simple model namely, TDLA, diesel locomotives available, TPPD, pump price for diesel fuel, TPPG, pump price for gasoline, and TRPK, railway passenger kilometer. To enhance modeling, using data for all the years, multiple linear regression analysis was also deployed, and 20 multivariable models explaining globalization variables were developed. Examples of cross-sectional models are Equations 2 and 3, where variables are defined as in Table 1:

$$\text{GITR} = 461.043 + 0.004\ \text{TADE} + 0.532\ \text{TTLS} \qquad (R^2 = 0.76) \qquad (2)$$

$$\text{GTRD} = -\ 2912.13 + 27.87\ \text{TTTS} \qquad (R^2 = 0.81) \qquad (3)$$

Correlation analysis with time-lag showed that the globalization variables often were also significantly correlated with the transportation variables with time-lag. Consequently, based on the results of correlation analysis with time-lag, preliminary time-lag relationships for globalization variables of years 1980 to 2000 were also evaluated. Transportation variables with time-lag of around 10 years were found good candidates to determine globalization variables in linear models. Equation 4 is an example of the developed models, where variables are defined as in Table 1:

$$\text{GTRD}_{1990} = -45265.7 + 1.72\ \text{TRQT}_{1980} + 4.55\ \text{TTLS}_{1980} \quad (R^2 = 0.82) \quad (4)$$

For further exploration of time-series relationships of globalization variables with respect to transportation variables, for individual countries, their arc elasticity's were evaluated. The arc elasticity E of a variable G with respect to a variable T for the period t1-t2 reflects the percent variable G changes with respect to one percent change of the variable T, as is shown by Equation 5:

$$E_{G/T,t1\text{-}t2} = [(G_{t2} - G_{t1})/(G_{t2} + G_{t1})] / [(T_{t2} - T_{t1})/(T_{t2} + T_{t1})] \qquad (5)$$

Where $E_{G/T,t1\text{-}t2}$ is the arc elasticity of variable G with respect to variable T during the period t1 to t2. When the difference between t1 and t2 becomes small, the arc elasticity converges to point elasticity. The arc elasticity of globalization variables G with respect to transportation variables T reflected individual country's trends and progressions with respected to the deployed variables. For selected periods between 1970 to 2000, the 23 globalization variables with respect to 18 transportation variables, based on non-missing values, provided a maximum of 23x18=414 elasticity's. For seven periods, six 5-year periods and one 30-year period covering 1970 to 2000, elasticity's were evaluated. The descriptive analysis of the developed elasticity's showed several interesting results. For individual periods, by observing large COV's, large variability of elasticity's among countries became evident. For selected variables listed in Table 1, for the 5-year period of 1995 to 2000 and the 30-year period of 1970 to 2000, the mean and COV of arc elasticity's are shown in Table 3. As for other variables, the positive mean values for elasticity's reflected that on the average the globalization and transportation trends were parallel in sign. The large COV's reflected large variations among countries, showing diverse inelastic, unit elastic or elastic behaviors. The elasticity's statistics, such as mean and standard deviation, were used as benchmarks for comparative assessment of individual countries. This provided a clue to individual countries status with respect to all 207 countries.

Table 3 Descriptive analyses for selected elasticity's

	1995-2000		1970-2000	
Elasticity	Mean	COV	Mean	COV
GITR/TTTS	1.27	8.26	3.91	5.38
GITR/TRQT	0.29	9.47	1.12	2.09
GITR/TATF	0.80	5.18	0.66	5.97
GTRD/TTTS	0.88	3.03	0.79	3.06
GTRD/TRQT	1.46	5.31	1.55	2.68
GTRD/TATF	0.66	4.97	0.32	6.92

To develop peer groups with respect to globalization and transportation, several classifications were deployed and evaluated, using different combinations of elasticity's. For each pair of deployed globalization and transportation variables, a generic classification consisting of 12 groups, as shown by 12 branches in Figure 2, was suggested and developed.

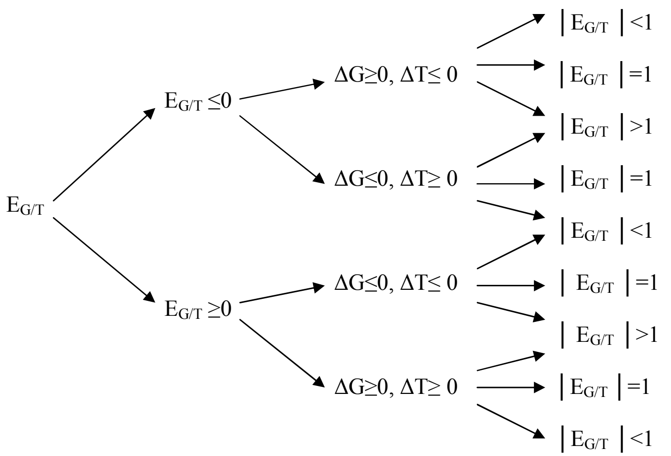


Figure 2 Taxonomy of countries based on sign and value of elasticity

The elasticity's statistics, such as mean and standard deviation, for developed peer groups, were also used as benchmarks for comparative assessment of individual countries. Nevertheless, as each country is unique in many aspects, their comparative assessment needs due considerations of these local factors and issues. The peer comparison was suggested as a complimentary to other types of appraisals.

Conclusions

This study was a preliminary step toward investigating relationships between national transportation and globalization characteristics. The study outcomes were based on limited accessible data and are of more methodological values. The 41 selected globalization and transportation variables were reasonable in reflecting individual country's status during period 1970 to 2000. Their descriptive analysis showed significant cross-sectional and time-series variations among 207 countries around the globe. As expected, globalization and transportation variables were often found positively correlated, and their time-series trends were found to have similar sign. Linear cross-sectional models explaining globalization variables with transportation variables were found statistically significant and reasonable. Transportation variables with time-lag of around 10 years were also found good candidates to determine globalization variables in linear models. The time-series arc elasticity's of globalization variables with respect to transportation variables showed significant variations. Each of the developed elasticity's represented a unique facet hinting on possible relationships among the deployed pair of globalization and transportation variables. Although considerable variations were observed among countries in their trends for globalization and transportation, their comparative assessment provided preliminary clues to their relationships. The elasticity's were used in taxonomy of the countries. Developed peer groups provided sound bases for comparative assessment and showcasing of good practices. Robust management of similar national transportation and globalization variables would facilitate national competitiveness and productivity in our contemporary globalizing world.

Acknowledgement

The authors wish to thank the Office of Research of the Sharif University of Technology for providing partial funding for this study.

References

1. Beck, U. (2000). *What is Globalization?* Cambridge Polity Press.
2. Vaziri, M. and Rasafi A. (2006). "An Appraisal of Globalization and Sustainable Development for European Countries," *Asian Journal of Water, Environment and Pollution,* Vol. 7, No. 1, pp. 147-152.
3. Dimitriou, H.T. (2005). "Globalization, Mega Transport Projects and the Making of Mega Places," *84th TRB Annual Meeting CD-ROM,* Wash. DC.
4. Nijkamp, P. (2003). "Globalization, International Transport and the Global Environment: A Research and Policy Challenge," *Transportation Planning and Technology,* Vol. 26, No. 1, pp. 1-9.

5. United Nations Economic and Social Commission for Asia and Pacific, UNESCAP (2005). *Toward an Asian Integrated Transport Network*, Monograph Series on Managing Globalization, No. 1, UN, New York.
6. World Bank (2004). *World Development Indicators 2003, CD-ROM,* Washington DC.

Are Asian Cities Leading the Way to Sustainable Mobility?

Isabel Cañete-Medina[1]

Abstract
This paper shows how Asian cities are leading the way to sustainable mobility based on a set of selected indicators. The reasoning behind the selection of the indicators is discussed. They focus on measuring the sustainability of personal mobility (versus freight mobility) for land-based transportation. Using the selected indicator set, a comparative analysis is presented showing how Asian cities compare with major metropolitan areas in the United States, Canada, Australia, Europe and other countries. The paper will discuss the areas of sustainable mobility where Asian cities are performing well and which categories they need to improve on. From the analysis, the paper will demonstrate why Asian cities are currently leading the way to sustainable mobility.

The challenge for Asian cities is whether they can maintain their leadership in this area and how they can improve on other issues, in particular air quality and traffic congestion. The paper provides a summary of development best practices based on the experience of developed countries, in order for Asian cities to continue to take the lead in providing sustainable mobility.

Definition of Sustainable Mobility
There is considerable debate on how to define the concept of sustainable mobility. Gudmundsson (2003) presented a discussion of the problems in providing a clear definition of the concept and they involve three issues:

- *Environmental sustainability* and how to define the limits of the sustainable use of the environment;
- *Economic allocation* and how to determine the contributions of each sector of the economy in solving sustainability problems;
- *Social interlinkage* and how to isolate transportation from other types of activities such as location and lifestyle choices.

For this reason, different organizations have proposed varying definitions of sustainable mobility. Such organizations include the Center for Sustainable Transportation (CST), the Organization for Economic Cooperation and Development (OECD), the World Business Council for Sustainable Development (WBCSD), European Union (EU) and Transport Canada – Moving on Sustainable Transportation (MOST). However it is defined, it is generally agreed that sustainable mobility involves the integration of economic, environmental and social concerns into transportation policy and decision-making.

[1] Isabel Cañete-Medina, AICP, Am. ASCE, Principal, cmQue, inc., 909 Foster Street, Suite 100, Evanston, Illinois, USA; T 847.475.8099; F 847.491.0134; E isabel.canete@cmque.com

This paper will not provide a definition of sustainable mobility and will instead focus on sustainability indicators that can be used to measure, evaluate, and compare various cities and regions. Mobility can be divided into two areas: personal mobility (passenger movement) and freight mobility (goods movement). This paper deals mainly with personal mobility using land transportation modes.

Sustainable Mobility Indicators

Because of the varying definitions of sustainable mobility, the types of indicators used to measure progress or performance also differs. Indicators are variables used to monitor conditions. They are selected because they state something significant about a specific policy or issue.

A comparative study and evaluation of sustainable mobility indicators (Gudmundsson 2001, 2003; Bickel, et.al. 2003; Litman 2007) shows the broad range of indicator systems that are currently in use. The indicators used for this study were chosen based on how *often* the measures were included in five selected indicator systems, and whether the *data is readily available*. Table 1 shows the most common indicators among five indicator systems, and whether city or country data are available. The criteria for selecting the five indicator sets to be included in the study are: 1) their specific application to mobility and transportation concerns; and 2) the degree in which they address a comprehensive set of economic, environmental and social issues.

A positive (+) measure for each indicator means more is better and a negative (-) measure means less is better. The indicators selected for the analysis in this paper are highlighted in Table 1.

Comparative Analysis of World Regions

A comparative analysis ranking world regions using the selected sustainability indicators was performed using a global data set compiled and analyzed by Cañete-Medina (2007). For each indicator, the regions were rank-ordered from 1 to 8, based on their median values, with the top region given a rating of 1 and the least sustainable region given a rating of 8.

Countries and their associated cities were divided into the following world regions: North America, South America including Central American and the Caribbean, Western Europe, Eastern Europe, Middle East & N. Africa, Sub-Saharan Africa, Asia (excluding Middle East) and Oceania (Australia, New Zealand and Papua New Guinea).

The comparative analysis is based on the assumption that all the indicators are equally weighted. For the conventional pollutants emissions indicator (CO, NOx, SO2, VOC), the regions where ranked separately for each major pollutant and a combined ranking was given for this indicator.

TABLE 1: Sustainable Mobility Indicators

	Indicator Sets/Frameworks					Data	
Indicators	**STPI(1)**	**WBCSD(2)**	**SUMMA(3)**	**TERM(4)**	**Litman(5)**	**Availability(6)**	**Measure**
Environmental							
Fuel Consumption	X	X	X	X	X	C	-
CO2 Emissions	X	X	X	X	X	C	-
Conventional Pollutants Emissions	X	X	X	X	X	C	-
Air Quality	X	X		X	X	U	-
Noise Pollution	X	X	X	X	X	nrd	-
Water Pollution	X	X	X	X	X	nrd	-
Land Take	X	X	X	X	X	C, U	-
Preservation of Habitat	X	X	X	X	X	nrd	-
Resource Consumption		X	X		X	C	-
Economic							
Commute Travel Time		x		x	x	U	-
Vehicle Miles Traveled	x			x	x	C	-
Land Use & Accessibility	x		x	x	x	U	+
Mode Split	x			x	x		
Percent Car Use						U	-
Percent Public Transportation						U	+
Percent Walking, Biking, Others						U	+
Traffic Congestion Delay	x		x	x	x	nrd	-
Household Travel Costs	x	x	x	x	x	C	-
Facility Costs	x	x	x	x	x	nrd	-
Transport Cost Efficiency		x	x	x	x	nrd	+
Economic Equity/Fair User Price	x		x	x	x	nrd	+
Social							
Safety	x	x	x	x	x	C	-
Community Livability/Social Cohesion		x	x		x	nrd	+
Social Equity/Affordability/Accessibility		x	x		x	nrd	+

Source: Table adapted from Canete-Medina (2007)
(1) Sustainable Transportation Performance Indicators (STPI), Centre for Sustainable Transportation (2002).
(2) Mobility 2030, World Business Council for Sustainable Development (WBCSD) (2001)
(3) Sustainable Mobility, Policy Measures and Assessments (SUMMA), (Ahvenharju, S. et.al. 2004)
(4) TERM 2001, European Environment Agency (2001)
(5) "Well Measured", Litman (2007)
(6) C=Country data; U=City data; nrd=Not Readily Available

The analysis results as presented by Canete-Medina (2007) are summarized in Table 2. The analysis shows that when per capita values are used for some indicators, Africa and Asia have the highest sustainability rating among the group. When total values (versus per capita values) are used for some of the indicators, Sub-Saharan Africa still ranks highest in terms of sustainable mobility, with Eastern Europe and Central & South America next, followed by Asia and the Middle East. The Asia region ranked fourth or fifth, despite having more than fifty percent (50%) of the world's population.

In both per capita and total bases, North America ranked the lowest in terms of sustainability, followed by Oceania and Western Europe.

Table 2: Sustainable Mobility Ranking of World Regions
(Comparative Analysis Using Selected Sustainable Mobility Indicators)

Rank	1	2	3	4	5	6	7	8
Rank Order Using Per-Capita Values	Africa	Asia	Eastern Europe	Middle East	South America	Western Europe	Oceania	North America
Rank-Order Using Total Values	Africa	Eastern Europe	South America	Asia	Middle East	Western Europe	Oceania	North America

Source: Table adapted from Canete-Medina (2007).
Notes: 1 indicates the highest ranking and 8 indicates the lowest ranking among the regions.
South America includes Caribbean and Central America. Middle East include North Africa.

It can be argued that the United States, Canada, Australia, New Zealand and Western Europe have the largest economies in the world and therefore resource consumption and waste is expected to be higher than in smaller economies. More resource inputs are required to produce more (economic) output. Thus, the analysis should be stratified by economic group such as developed countries and developing countries.

The analysis presented here, however, is still useful as it illustrates how world regions compare with each other in terms of the selected sustainable mobility indicators.

Comparative Analysis of Developed Countries

A comparison of developed countries was prepared using the same dataset. The group of developed or industrialized countries included the United States, Canada, Australia, New Zealand, Japan, Hong Kong, Singapore, South Korea and the countries of Western Europe. The average per capita GDP for 2005 of each region are as follows: North America ($31,166); Asia ($27,019); Western Europe ($25,533) and Oceania ($19,169).

The results of the analysis as shown in Table 3 indicate that on a per capita basis, the Asian region ranks highest in sustainability while Western Europe ranks highest when total values are used. Oceania and North America ranks third and fourth, respectively for both per capita and total values.

Table 3: Sustainable Mobility Ranking of World Regions
(Comparative Analysis Using Selected Sustainable Mobility Indicators)

Rank	1	2	3	4
Rank Order Using Per-Capita Values	Asia	Western Europe	Oceania	North America
Rank-Order Using Total Values	Western Europe	Asia	Oceania	North America

Source: Table adapted from Canete-Medina (2007).
Notes: 1 indicates the highest ranking and 4 indicates the lowest ranking among the regions.

Asian Cities and Sustainable Mobility

Newman and Kenworthy (1999, 2000) collected various types of data for 37 cities in the 1990s as part of a World Bank project. Their analysis of the data showed results comparable to the analysis presented here. Based on their data, on a per capita basis, industrialized Asian cities do better than other cities in the developed world in a variety of indicators related to sustainable mobility.

Under which indicators do Asian cities perform well? Asian cities have lower per capita fuel consumption and per capita CO2 and conventional pollutant emissions than most cities. These can be due to a number of factors including low car usage and high public transportation use. The high urban densities and mixed used patterns of development also allow for more accessibility to basic goods and services, which can lead to reduced vehicle-kilometers traveled. Asian cities also have lower transportation costs as a percent of household costs probably due to low car ownership and affordable public transportation. Less land is used for road infrastructure, which reduces resource consumption and waste, preserves habitat and lessens disruption to communities and neighborhoods.

The high urban densities in Asian cities are good for accessibility but it can lead to higher concentrations of pollutants and poor air quality. Noise pollution is also an additional problem. One solution would be to decrease urban densities or apply more stringent vehicle emission standards to further lower air pollutant emissions and reduce noise decibels. Asian cities also have to do more in reducing transportation-related deaths and injuries. Pedestrian and bicycle facilities should be incorporated into roadway design to protect non-motorized users from vehicles. Traffic congestion is also a major problem. Investments in public transit have proven valuable in confronting this issue in high density cities like Hong Kong and Singapore.

Future Development Trends in Asian Cities

Asia is the fastest growing region in the world in terms of rural to urban migration. The growing urbanization in Asia is fueled by population growth and new economic opportunities. What patterns of urban and transportation development should the growing Asian cities adopt? In terms of best practices, the developing Asian cities have a choice of following three development models. The critical relationship between land and transportation development is evident in all three models.

- the *Asian Model* of high density development coupled with massive investments in public transit systems.
- the *European Model* of medium density development with investments in both roadway and public transit systems.
- the *American Model* of low density development with investments primarily in road infrastructure.

The analysis presented here has shown that North American, Australian and New Zealand cities have the least sustainable mobility systems. Studies in the United States have shown that the construction of extensive roadway networks has not lead to reduced congestion (Hansen 1995; STPP 1999) and that it has fueled low-density development and urban sprawl. The economic, environmental and social costs of sprawl have been well documented (RERC 1974; Burchell et.al 1998). There is an ongoing debate in the United States on whether sprawl is good or bad and what should be done about it.

Conclusions

Based on a select set of indicators, mobility in Asian cities is shown to be more sustainable than in other parts of the world. Their high urban densities, low car usage, high public transportation patronage and less consumption of land resources for transportation are providing for a more efficient and sustainable transportation system to support economic growth.

References

Ahvenharju, S., Konnola, T., Hielt, M., Tomassi, R, Arend, M., Bickel, P., de Ceuster, G., van Grol, R., Walker, W., Klautzer, L., Burg, R., and Rohling, W. (2004). *Operationalising Sustainable Transport and Mobility: The System Diagram and*, Sustainable Mobility, Policy Measures and Assessments, Leiden, The Netherlands.

Bickel, P., Ahvenharju, S., Konnola, T., Hielt, M., Tomassi, R, Arend, M., Burg, R., and Rohling, W. (2003). *Setting the Context for Defining Sustainable Transport and Mobility*, Sustainable Mobility, Policy Measures and Assessments, Leiden, The Netherlands.

Canete-Medina (2007). "Data and Analysis of Selected Sustainable Mobility Indicators by World Regions (Working Paper)". Will be available for download at <www.cmque.com> on September 15th, 2007.

European Environment Agency (2001). *TERM 2001 - Indicators Tracking Transport and Environment Integration in the European Union,* European Environment Agency, Copenhagen, Denmark.

Gilbert, R., Irwin, N, Hollingworth, B, and Blais, P. (2002). *Sustainable Transportation Performance Indicators (STPI) Project, Report for Phase 3*, Centre for Sustainable Transportation, Winnipeg, Manitoba, Canada.

Gudmundsson, H. (2001), *Indicators and Performance Measures for Transportation, Environment and Sustainability in North America*, National Environmental Research Institute, Roskilde, Denmark.

Gudmundsson, H. (2003). "Making Concepts Matter: Sustainable Mobility and Indicator Systems in Transport Policy." *International Social Science Journal: Sustainable Mobility*, (176), 199-217.

Litman, T. (2007). "Well-Measured: Developing Indicators for Comprehensive and Sustainable Transport Planning." *Victoria Transport Policy Institute*, <http://www.vtpi.org/wellmeas.pdf> (July 2, 2007).

Newman, P. and Kenworthy, J. (1999). *Sustainability and Cities*, Island Press, Washington, DC.

Newman, P. (2000). "Sustainable Transportation and Global Cities." *Institute for Sustainability and Technology Policy*, http://www.sustainability.murdoch.edu.au/casestudies/Case_Studies_Asia/sustrans/sustrans.htm (June 14, 2007).

World Business Council for Sustainable Development (WBCSD) (2004). *Mobility 2030: Meeting the Challenges to Sustainability,* The Sustainable Mobility Project, World Business Council for Sustainable Development, Washington DC, United States.

BUILD OR NO BUILD?-Debates on Honolulu High-Capacity Transit Project

Honglong Li Ph.D.[1], Jennifer Russell[1], Roger Morton[2], Chris Wellander[1], Mark Scheibe[1], Toru Hamayasu[3], and Wayne Yoshioka[1]

[1] PB Americas, Inc. 1001 Bishop Street Suite 2400, Honolulu, HI 96813. PH:1-808-531-7094; FAX:1-808-528-2368; Email: lih@pbworld.com
[2] O'ahu Transit Services, Inc, 811 Middle Street, Honolulu, HI 96819. PH: 1-808-848-4508 Email: rmorton@thebus.org
[3] Department of Transportation Services, City and County of Honolulu, 650 S. King Street, Honolulu, HI 96813. PH: 1-808-527-6978 Email: thamayasu@honolulu.gov

ABSTRACT In view of the greater attention being paid today to such problems as pollution, the traffic congestion that contributes mightily to it, and the ever-increasing cost of gasoline, a recent planning process in Honolulu offers a timely look at how the decisions regarding a new mass-transit system were made and why. Honolulu's linear city layout, mountainous topography, and high density land-use make it ideal for a grade-separated transit system. Multiple efforts to plan a high-capacity mass transit system for Honolulu have occurred over the past several decades, and have been aborted at least three times since the 1970's. Through these floundered processes, it was learned that building a fixed guideway transit system takes more than sophisticated planning and engineering, it takes political will, public support, and a dedicated, predictable source of funding as well. This paper explains the most recent planning process for a new mass-transit system in Honolulu. It summarizes the debating points persistent since the beginning of the previous planning efforts. It was through these healthy debates that a broad consensus was reached on exactly what alternative best met locally defined goals and objectives for the specified corridor.

KEY WORDS: Mass Transit, Rail, Planning.

INTRODUCTION

Growing urban traffic congestion on existing transportation infrastructure has pushed jurisdictions around the world to aggressively seek mass-transit solutions. The City and County of Honolulu, Hawai'i, USA on the island of O'ahu is among these jurisdictions. Multiple efforts to plan a high-capacity mass transit system for Honolulu have occurred over the past several decades, and have been aborted at least three times since the 1970's. Through these floundered processes, it was learned that building a fixed guideway transit system takes more than sophisticated planning and engineering, it takes political will, public support, and a dedicated, predictable source of funding as well. In 1992, the City Council voted 5-4 to reject raising taxes to fund a new transit system. The rejection turned away more than $600 million in federal funds authorized for Honolulu's fixed guideway system. Ten years later Honolulu proposed Bus Rapid Transit (BRT) as an affordable alternative but a lack of dedicated right-of-way made it less attractive so the project was later terminated.

The most recent effort to resurrect a high-capacity mass-transit system was initiated in 2005. This time the planning came with a tangible local financial commitment from the state legislature – a half percent State General Excise and Use Tax (GET) surcharge was designated for building and operating a high-capacity transit system on O'ahu. Immediately after this most recent planning effort was announced, organized support and opposition mobilized and the project stepped into its familiar territory. Will it survive this time?

The paper summarizes Honolulu's so-far successful planning of a new mass transit system. The paper has five sections: 1) Introduction; 2) Planning Procedure; 3) Debating Points; 4) Selection of Locally Preferred Alternative; and 5) Conclusion.

PLANNING PROCEDURE

The planning and project development process follows the requirements of the U.S. Department of Transportation, Federal Transit Administration (FTA) because FTA New Starts funds will be used for this project.

This paper covers the very first step, the Alternatives Analysis (AA) which includes:

- Identifying specific transportation problems in an area, or "corridor" being studied;
- Defining reasonable alternative strategies to address these problems;
- Forecasting potential environmental, transportation, and financial impacts of these alternatives; and
- Evaluating how each alternative effectively addresses the transportation needs, goals, and objectives for the corridor.

The primary result of the AA is selection of the Locally Preferred Alternative (LPA). The AA Report provides decision makers (who in this case are the nine members of the Honolulu City Council) with enough information to select a specific project design concept, and determine the scope of the project. After completion of the AA, the City Council has the information needed to select an LPA. With this information, the project can advance to preliminary engineering and the final phases of environmental review, design and eventually project construction.

Alternatives Considered

Four alternatives were considered through the project scoping process (City and County of Honolulu, 2006) including:

- No Build. The No Build Alternative includes existing transit and highway facilities and committed transportation projects anticipated to be operational by 2030. This alternative shows what the transportation system would be like if there were no additional new changes made to the system.
- Transportation System Management (TSM). The TSM Alternative would provide an enhanced bus system based on a streamlined route network, expansion of the present morning peak-hour-only reversible lane to both a morning and afternoon peak-hour operation, and relatively low-cost capital improvements on selected roadway facilities to give priority to buses.
- Managed Lane. The Managed Lane Alternative would include construction of a two-lane grade-separated facility for use by buses and other priority vehicles. The lanes would be managed to maintain free-flow speeds for buses,

while simultaneously allowing High-Occupancy Vehicles (HOVs) and variable pricing for toll-paying single-occupant vehicles.

- Fixed Guideway. The centerpiece of the Fixed Guideway Alternative is a mostly elevated fixed guideway system integrated with the bus, walking and bicycling networks for access and egress, as well as with automobile access with park-and-ride and passenger drop-off facilities at appropriate stations. Multiple alignment options were studied and potential station locations were identified.

Public Outreach

Public outreach was a critical part of the AA process because the decision makers must consider public acceptance when selecting the LPA. Conveying timely and accurate information is essential to ensuring full public awareness of the facts surrounding the project and its issues. Honolulu embraced a proactive methodology for raising community awareness and understanding by implementing a "Speakers Bureau" which presented information on the project to any group who requested it. Through Speakers Bureau meetings, citizens engaged directly with project planners and engineers regarding the project issues relevant to them. The intent was to engage the public through technical discussions at the meetings and offer the opportunity to address any constituents' questions directly. In less than one year, over 200 Speakers Bureau engagements were fulfilled. The effect was an increase in a focused understanding of the project and a deeper comprehension of the process. The deeper understanding is evident in news broadcasts, newspaper letters to the editor and in the thoughtfulness and detail of questions received from the general public regarding the project. The Speakers Bureau engagements have proven to be an effective method of reaching the deeper community roots and activating them to become involved in the process.

Yet the proactive approach did not quiet the well-organized, vocal opposition who, in the past, successfully presented highly contentious and controversial aspects of the proposed mass-transit solution. During this AA process discussions and arguments, sometimes emotional and unsubstantiated, were carried out at public meetings, Speakers Bureau presentations, in the newspaper, internet forums, and on radio and TV. It was through these challenging, yet healthy debates that the public was given more avenues to thoroughly understand the alternatives and decide which alternative was the best fit for their communities.

DEBATING POINTS

1 Does Honolulu have the population to support the new mass transit system?

Opponents of the new mass transit system claimed that Honolulu's less than 1 million in population can not provide adequate ridership. Population is an important consideration in planning for any mode of travel including transit or highways. But the population within a transportation corridor is more important than overall population. The primary transportation corridor in Honolulu extends from Kapolei in the 'Ewa District to the University of Hawai'i at Manoa and Waikiki in the east. The east/west length of the corridor is approximately 25 miles. The north/south width is a maximum of four miles, bounded by the Koolau Mountain Range and the coastline.

This corridor encompasses 60% of the island’s population currently and will encompass close to 70% in 2030. In addition, 93% of population growth and 95% of employment growth will occur in this corridor by 2030.

Population density is actually a better determinant of the potential for a new mass transit system ridership than pure population. The population density for the entire island of O’ahu was 1461 people per square mile making O’ahu the nation’s 16th most dense metropolitan area (U.S. Census Bureau, 2000). The population density of the primary transportation corridor is over 50% higher than the O’ahu average and is higher than all but four metropolitan areas on the mainland. Thus, the potential of the new mass transit system for Honolulu is higher than many other successful existing fixed guideway systems on the Mainland.

<u>2 Why should we support transit when most people use cars to get to work?</u>

In the Honolulu District, the most recent data indicates that about 8% of the work force used transit to get to their jobs while 83% used automobiles. Islandwide transit use is about 11% of all work trips. This ranks O’ahu in the top twenty U.S. metropolitan areas for transit use. High trip attraction and production centers such as Waikiki, Downtown, the University area, and areas immediately west of Downtown have a transit modal split as high as 36%.

Transit is not for everyone for all purposes. Some people need a car for their work. For other people, transit does not go where they want to go with a sufficient level of service. But providing an effective transit system gives people a choice, and many people will choose transit over automobiles if the level of service is high enough and the trip is convenient. Today’s massive traffic congestion in Honolulu coupled with the fact that the current transit system operates in the same right-of-way has made transit less competitive than it used to be. That’s precisely why the proposed fixed guideway will have its own right-of-way to bypass daily traffic congestion.

<u>3 Will the whole Island benefit from it?</u>

The reactions from individual neighborhoods to the proposed mass-transit alignment have been distinct: generally, neighborhoods in the vicinity of the alignment are more supportive than the ones that will not be served directly. However, island-wide planning on O’ahu has focused future growth into Central O’ahu and the ‘Ewa area, rather than East Honolulu and windward. There was an explicit decision to preserve other areas of the island at close to current development size and density. The planned growth into Central O’ahu and the ‘Ewa area will relieve growth pressure from the rest of the island. Therefore, there is a political sense of responsibility to support the areas designated for future growth by providing the necessary transportation infrastructure.

If a fixed guideway system is implemented, bus resources will be reassigned to the underserved areas of the island and will enhance the feeder bus network for the fixed guideway.

System benefits will be experienced island-wide. Commuters will be able to get to their destinations more reliably. People who drive will benefit from less congested roads. Businesses will require less parking for employees and customers.

4 How does Fixed Guideway perform better than a Managed Lane Option?

Neither the Fixed Guideway nor Managed Lane alternative is expected to reduce future congestion to levels less than today, but future island-wide hours of traffic delay would be 20% greater with the Managed Lane Alternative than with the Fixed Guideway Alternative. The Managed Lane actually attracts more cars into the network systemwide. The system increase in automobiles results in a net increase in system delay. And, in the case of the Managed Lane Alternative, transit riders would be subjected to the same delay as automobile drivers through critical portions of the corridor (e.g. downtown).

Bottleneck conditions will exist as drivers attempt to access and exit the Managed Lane facility. The number of automobiles per hour conveyed by the Managed Lane facility would put tremendous pressure on the existing downtown roadway network where considerable improvement is often impossible due to scarce land availability.

Fixed guideway has a lower unit length cost than the managed lane. The managed lane structure is identical to a two-lane viaduct and is typically more than 45 feet wide, while the fixed guideway could be as narrow as 25 feet wide because a median or shoulder lanes are not needed. Either structure will be supported by columns about 30 feet tall.

5 Can Honolulu afford a Fixed Guideway System?

Opponents of a fixed guideway system have claimed that Honolulu does not have the population base to finance it. Paying for such a large project does require collective effort. However, financial estimates show that a GET surcharge increase could raise most of the local funding needed.

Local money must be provided to meet FTA matching funding requirements. The GET provides a steady cash flow because it is collected for all sales transactions. It was estimated that 30 - 36% of the GET tax burden will be borne by tourists. After years of discussion and analysis, the majority of political members agreed that the GET is the best mechanism to raise the funds necessary to finance a transit project. The Hawai'i state legislature passed Act 247, authorizing the county to levy a tax surcharge to construct and operate a mass transit project serving Honolulu. In August, 2005, the Honolulu City Council subsequently adopted ordinance 05-027 to levy a 0.5% general excise tax surcharge to fund public transportation. It was estimated that about 80% of total project cost will be paid by GET surcharge revenue and 20% will be paid by FTA New Starts Funds.

SELECTION OF LOCALLY PREFERRED ALTERNATIVE

The City Council examined the findings of the analysis by having a transit advisory task force review the methodologies documented in the AA report. The task force found that the numbers were reasonable and ended up supporting the report's overall findings.

The Fixed Guideway Alternative was chosen as the LPA. Fundamental to the council's decision were ensuring maximum ridership in densely populated areas and encouraging development in the designated areas of O'ahu. Because the amount of available and expected funding is a concern, the Council requested that transportation officials determine a minimum operable segment (MOS) which would be a section of the LPA serving a significant portion of the transit corridor.

Technologies retained for future study include light rail, people mover, monorail, magnetic levitation, and rapid rail. Honolulu intends to select the technology through a process that includes considerations for cost and performance criteria.

CONCLUSION

Honolulu's most recent mass transit planning considered four alternatives: No Build; Transportation System Management; Managed Lane; and Fixed Guideway. The Alternatives Analysis Report and supporting materials provided a substantial comparison of the transportation, environmental, and financial costs and benefits between the various alternatives. The proactive public outreach entailed many community and regional scoping meetings to allow the public to comment and offer suggestions on the various ways of providing alternative transportation solutions.

Debates between fixed guideway supporters and opponents were healthy for the community. It helped to reach a broad consensus on exactly what type of improvement best meet locally defined goals and objectives. The debates positively answered the following questions:

1. Does Honolulu have the population to support the new mass transit system?
2. Why should we support transit when most people use cars to get to work?
3. Will the whole Island benefit from it?
4. How does Fixed Guideway perform better than a Managed Lane option?
5. Can Honolulu afford a Fixed Guideway System?

The discussion on debating points is not limited to Honolulu and it can be applicable to other areas that are considering mass transit system.

The Honolulu City Council chose Fixed Guideway as Locally Preferred Alternative. The development of the mass transit system is advancing to reality 40 years after it was first recommended as a necessary component of the future transportation system.

REFERENCES

PB Americas, Inc. (2006) Honolulu High-Capacity Transit Corridor Project Alternatives Analysis Report.

U.S. Census Bureau (2000) Census 2000 Summary File 1; http://factfinder.census.gov/ .

ACKNOWLEDGEMENT

Development of materials used in the preparation of this paper was financed in part through grants from the US Department of Transportation, Federal Transit Administration and the City and County of Honolulu. The contents of this document do not necessarily reflect the official views of the US DOT and the City and County of Honolulu.

Transportation Investment in Rapidly Urbanizing China: Best Practices for Supporting Balanced Regional Economic Returns

Miles Cheang[1], Xiaobo Liu[1], Minghua Huang[2]
1 Jacobs Edwards and Kelcey, Department of Planning, Morristown, NJ, USA
mjcheang@gmail.com; Xiaobo.liu@jacobs.com, (973)267-8830
2 Chengdu Municipal Committee of Communication, Department of Science and Technology, Chengdu, China, huangminghua742300@sina.com, (28)8556-0606

To support balanced and sustained economic returns from transportation investment, the convergence of development strategies on a regional scale is crucial. Recognizing that regional economies are shaped by unique geostrategic conditions, effective transportation development must be planned in the context of regional urbanization dynamics. This paper examines the development patterns of the Pearl River Delta region in Guangdong province, serving as a case study for exploring broad strategies that best promote balanced regional growth through transportation investment and development. Based on a methodical assessment of strengths and weaknesses in the region, the rural periphery was identified as an underutilized asset crucial to supporting balanced regional growth. Recommended best practice strategies emphasize the role of policy-making and transport infrastructure as allocative mechanisms in a regional economy.

Introduction

While ample evidence exists in scholarly literature and real-world indicators documenting a casual relationship between transportation investment and economic growth, a detailed look reveals many contingent influences of which transport infrastructure is only one (Banister and Berechman 2000). It is more apt to subsume infrastructure provision as a crucial component of economic growth and accord equal emphasis to other influences, such as urbanization dynamics and social capital. To be effective on a regional scale, transportation investment must be cognizant of these cohort factors that together impact regional development and prospects for growth.

Since transport infrastructure serves as the backbone of an economy, it is irrefutable that transportation development contributes to regional growth. A widely held explanation posits regional economic growth as the natural result of increased productivity in local firms. In this view, regional growth stems from individual firms making strategic decisions aimed at optimizing efficiency based on perceived transport benefits, such as investment that result in reduced travel times, distance costs, and congestion. However, without taking into consideration the broader impact of transport infrastructure on regional dynamics, this approach may only result in local economic growth. In the worse case, local economic growth may be exclusionary and take place at the expense of regional development. For instance, the persistent urban-rural wealth disparity in the Pearl River Delta region (PRD) of Guangdong province has been linked to the uneven provision of transport infrastructure and access to economic opportunity (Fan 1996, Lin 2001). While this suggests the allocative potential of infrastructure on regional economies, the pervasiveness of wealth disparity throughout and between many different regions in

China suggests a more systemic issue may be at work; transportation infrastructure is but one contingent factor of economic growth.

A three-step planning process based on the perceived strengths and weaknesses in a regional economy must take place in order to support balanced returns on transportation investment. First, issues must be identified in the context of existing linkages between functional places that comprise a regional economy. Second, any solutions devised must take into account the allocative potential of transport infrastructure on regional dynamics. Third, a broad framework of place-based policies must be identified to support new regional dynamics that may result from implementing solutions identified in step two. As a brief exercise in this approach, this paper methodically assesses the prospects for balanced regional growth through transportation investment in the PRD. In accordance with step one, the following section briefly examines dynamics that have traditionally defined the economic vitality of the PRD region in order to identify potential weaknesses in present day. The concluding section addresses steps two and three by asserting both transportation investment and policy-making as indispensable tools in supporting balanced economic development in the PRD region.

Regional Development Dynamics in the Pearl River Delta

The regional development pattern in the PRD has been shaped by a strong export-oriented economy. Place-based policies in the early 1980's have evolved to foster regional linkages between places of complementary economic functions, enabling the creation of "metropolitan interlocking regions" (MIR). A MIR is characterized by a metropolitan population living in multiple large cites within a region connected by a comprehensive transportation system, as well as close social and economic ties between functional areas (Zhou 2002). In the PRD MIR, a region-wide export economy consisting of numerous industrial clusters manufacturing labor-intensive products accounts for the majority of productive activities. Its spatial structure is defined by a transportation system linking economic activities between cities within the MIR. As such, each city within the MIR can be broadly categorized by the particular segment or segments of the manufacturing value chain that it occupies. Based on a comparative analysis of economic and demographic characteristics, as well as land utilization, in thirty-one cities and counties in the PRD, Lin (2001) identifies two distinctive zones radiating outward from the urban centers of Guangzhou, Foshan, and Jiangmen that stratify economic functions within the MIR: the peri-urban zone and the rural periphery.

In the PRD MIR's export-oriented economy, the main economic beneficiaries beyond urban centers have been peri-urban localities that enjoy close ties with both central cities and rural areas (Lin 2001, Xu and Tan 2002). Generally located between urban centers and the rural periphery, activities within the peri-urban zone lend crucial support to both areas through dynamic urban-rural interactions between agricultural and non-agricultural sectors. Farmers within and beyond the peri-urban zone benefit from producing farm commodities for a more affluent market, as well as by joining the urban workforce during non-harvest periods. Urban industries gain from access to cheaper land and a large surplus rural labor pool for manufacturing activities, fueling and sustaining a robust comparative advantage in the global economic system through low-cost manufacturing. These dynamics have made the peri-urban zone an ideal destination for foreign investment and migrant labor, potentially offering multinational companies a business-friendly regulatory environment with substantial fiscal incentives as well as cheap land and labor.

The development of transport and information technology (IT) infrastructure has allowed economic activities to be outsourced and coordinated over vast regions, enabling companies to achieve efficient industrial structuring through vertical disintegration of production management. This regional development trend has greatly favored peri-urban cities with access to well-developed transport corridors, as seen in cities such as Dongguan. The city's illustrious position is due in large part to its strategic location along the major highways and railway between Guangzhou and the Hong Kong metropolitan area. Rapid population growth, rising land leasing costs, and a thorough transition from an agriculture-based to a manufacturing-dominant economy have accompanied peri-urban Dongguan's development into a major industrial center of the PRD (Webster and Muller 2002). In contrast, the rural periphery has been left behind by the PRD's urban-centric and export-oriented growth. Accounting for over 60% of the total land area and 38% of the total population of the PRD, this zone contributed only 14% of industrial and agricultural output and received 8% of foreign investment (Lin 2001). As such, the rural periphery remains underdeveloped and contributes primarily out-migrant labor to service urban manufacturing activities in the PRD MIR.

The regional dynamics described above have made Guangdong province the traditional hot-spot for foreign investment and in-migrant flow. However, despite the vast reserve of surplus workers in rural China, Guangdong province now reportedly suffers from a labor shortage that threatens the region's traditional comparative advantage. A detailed investigation of the same dynamics that transformed Guangdong into a leading global exporter reveals structural issues deeply ingrained in the PRD's regional economy that may explain the labor shortage.

Rising production costs (e.g., land, labor), inflationary pressures, and competition from other industrializing regions have put considerable strain on typically low profit-margin urban manufacturing enterprises in the PRD. Due to the fact that many of these enterprises occupy the least-profitable segment of the production value chain, it may not be economically feasible to upgrade aging facilities nor raise wage levels sufficiently to remain competitive with newer MIRs. For instance, in the Yangtze delta region, manufacturing enterprises are increasingly offering potential migrants higher wages, more modern working facilities and amenities, as well as a comparatively better civic environment (e.g., schools, lower crime) (Chen 2007). As workers in the underdeveloped rural periphery and inland provinces migrate to urban areas seeking the best wages and livelihood, the PRD will continue to lose its share of out-migrant laborers to newer MIRs along its present development track.

The labor shortage in the PRD has called into question the sustainability of the region's traditional growth model of high-volume, low-value, and low-cost manufacturing built on strong urban centers and a weak rural periphery. Viewed in this light, the solution can be seen as two-fold: first, the PRD must devise strategies to retain rural laborers both within and beyond Guangdong and second, manufacturing industries within the PRD must find a way to maintain low production costs in order to remain competitive in the global economy.

Transportation Investment and Policy - Regional Restructuring in the PRD

In a dynamic economy, structural issues can be viewed as opportunities for growth. A potential solution to the question of sustainability facing the PRD lies in the physical integration of the rural periphery into the regional economy, which amounts to a proposal for a regional restructuring of economic functions within the MIR. On the one hand, urban industries that move out of cities benefit by escaping the rising cost of production in urban centers and peri-urban areas, as well as directly accessing the abundance of cheaper land and latent labor in the rural periphery. Alternatively, these enterprises may outsource a portion of production to the rural periphery by investing in the manufacturing capacity of existing small-scale rural township, village, and household enterprises (Xu and Tan 2002), thereby creating new backward supplier links to relieve production demand on city factories. On the other hand, rural areas benefit from a local higher-wage alternative to agriculture production and, in many cases, subsistence farming. This approach serves to increase the competitiveness of the PRD by retaining rural workers who may otherwise migrate to distant MIRs seeking higher wages and superior amenities.

Much as it was a crucial component of economic growth in peri-urban cities, the provision of transport infrastructure is fundamental to integrating the rural periphery with the PRD economy. While investment toward improving transport infrastructure in urban areas will benefit certain industries, it would likely serve to increase land value and exacerbate the difficulties facing low-margin manufacturing enterprises. Due to this, transportation investment should be increasingly balanced between urban areas and the rural periphery. Extending the transport network to rural areas reduces distance costs associated with the movement of freight, making feasible vertical disintegration dynamics in production that have enriched the peri-urban zone of the MIR. An extended transport network is also crucial to the creation of new backward supplier links and, as the rural periphery develops, presents the possibility of future forward distribution links to emerging regional markets (Chen 2007).

A crucial aspect of regional economic restructuring is the willing introduction of peri-urban dynamics into the rural periphery, similar to those discussed above that have enriched cities such as Dongguan. Although this process is continuously underway via an organic process of urban expansion and subsequent peri-urbanization of previously rural land (Webster and Muller 2002), it may be willingly accelerated by premeditated transport investment toward effecting that end. If successfully implemented, induced growth in the periphery will contribute to mitigating the regional urban-rural wealth disparity by providing the latter with a reliable stream of revenue for local capacity-building and social development. There is evidence of this already happening in peri-urban, or previously rural areas. Lin (2001) cites a case study of peri-urban Nanhai (a district of Foshan) wherein rural workers' increased personal savings contributed significantly to capital investment in rural industries. Among other priorities, fostering peri-urban dynamics is crucial to integrating the rural periphery into the regional economy.

The role of accompanying policies is equally important in the regional restructuring process. Local governments must support the regional development of transport infrastructure out toward rural areas, similar in concept to the central government's official Great Western Development Strategy for inducing growth in inland provinces. This requires the willing adoption of 'push' dynamics in which government policy and investment are geared toward the development of areas away from the city proper, notably in the provision of transport infrastructure. Thus far, local development policies have generally favored 'pull' dynamics in generating economic growth. This refers to the creation of county and township industrial estates, and Economic and Technological Development Zones that seek to agglomerate both urban and rural industries to achieve economies of scale. Since municipalities are typically the developers of these profitable initiatives, these zones are often located close to the city (Wu 1998). As successful as 'pull' factors have been, these dynamics increase the dependency of urban areas on contingent rural in-migrant labor, encourages fierce intra-urban competition between low-margin manufacturing enterprises, intensifies 'overheating' of the urban core areas, and exacerbates the urban-rural wealth disparity. In this light, shifting local policy and regional development priority toward a 'push' orientation is crucial to sustainable and balanced economic growth in the PRD.

A second dimension in the role of policy in regional restructuring involves the necessary development of social and human capitals in both urban and rural areas. The creation of new urban-rural business networks and local personnel in the rural periphery capable of facilitating the introduction of modern industry is crucial to sustainable development (Copus 2001). Meanwhile, as urban manufacturing enterprises outsource or relocate production to the rural periphery, economic activity in urban areas must upgrade to higher value-added functions in the manufacturing production chain to replace lost employment and maintain economic vitality (Chen 2007). Specifically, activities such as research and development, marketing, product design, and logistics management require highly skilled professionals. Local governments must encourage these pursuits by enhancing the education system, supporting training programs, and attracting talent from other regions and abroad.

Conclusion and Future Study

This paper has exposed inherent weaknesses stemming from the unbalanced regional development patterns in the PRD as a legacy of the region's traditional economic strengths. Underdevelopment of the rural periphery and urban-centric development have together created a substantial floating population seeking compensation beyond the means of many enterprises within the PRD. To a large extent, this uneven development can be attributed to the geographic provision of transport infrastructure. While the primary benefits of transportation investment are traditionally seen as efficiency-based (e.g., distance costs), the allocative potential of infrastructure on regional economic activities cannot be overstated. Transport infrastructure must be conceptualized as an intermediate good (Preston 2001), i.e. an input and/or output of economic activity, and be utilized as such in planning regional development priorities. In the context of regional restructuring in the PRD, this approach implies that transportation investment as an allocative mechanism must be accompanied by economic and socioeconomic policies that generate derived demand (i.e., demand for an item not for its own sake but for use in the production of goods and services) for productive activities in the underdeveloped rural periphery.

Thus far, the fact that the PRD continues to register consistently high rates of economic growth suggests that sufficient industrial upgrading has taken place to offset the present shortage of labor. However, as domestic competition from newer MIRs such as the Yangtze delta become increasingly fierce, the sustained development of the PRD along its present track is far from certain. This challenge requires policy makers to devise pragmatic solutions that build upon existing strengths of a region through examining and capitalizing on the full range of available assets. From this perspective, the rural periphery in the PRD can be seen as a crucial asset that has yet been utilized to full potential. Not only would successful regional restructuring of economic activities toward the periphery support future growth in the PRD, it would also relieve in-migrant pressure on urban systems, aid in narrowing the urban-rural wealth disparity, and spur the first steps toward developing a balanced domestic economy.

Advanced information technologies and dematerialization of economic functions in today's evolving economy present the opportunity for progressive industrial structuring (e.g. flexible specialization). Reflecting this trend, recent literature in transport geography have highlighted the potential for spatial benefits such as the physical agglomeration and clustering of firms to be distributed aspatially in disregard of the traditional core-periphery dichotomy (Copus 2001). Research toward a new "transpatial" perspective on core-periphery economic relationships within a MIR would contribute greatly to balanced growth in the PRD region.

References

Banister, D., Berechman, J. (2000). "Transportation Investment and Economic Development." *UCL Press*. London.

Chen, X. (2007). "A tale of two regions in China - rapid economic development and slow industrial upgrading in the Pearl River and the Yangtze River Deltas." *International Journal of Comparative sociology*, 48(167).

Copus, A. K. (2001). "From core-periphery to polycentric development: concepts of spatial and aspatial peripherality." *European Planning Studies*, 9(4), 539-552

Fan, C. C. (1996). "Economic opportunities and internal migration: a case study of Guangdong province, China." *The Professional Geographer*, 48(1), 28-45

Lin, George, C. S. (2001). "Evolving spatial form of urban-rural interaction in the Pearl River Delta, China." *The Professional Geographer*, 53(1), 56-70

Preston, J. (2001). "Integrating transport with socio-economic activity – a research agenda for the new millennium." *Journal of Transport Geography*,9(1), 13-24

Webster, D., Muller, L. (2002). "Challenges of peri-urbanization in the Lower Yangtze region - The case of the Hangzhou-Ningbo Corridor." *Asia/Pacific Research Center, Stanford, CA*.

Wu, F. (1998). "The new structure of building provision and the transformation of the urban landscape in Guangzhou, China." *Urban Studies*, 35(2), 259-283

Xu, W., Tan, K. C. (2002). "Impact of reform and economic restructuring on rural systems in China." *Journal of Rural Studies*, 18(1), 65-81

Zhou, Yi-Xing. (2002). The prospect of international cities in China. In: Logan J.R., editor. New Chinese city: Globalization and market reform (pp. 59-73). Oxford (UK): Blackwell Publishers.

Infrastructure Innovation – Turning Public Works into Private Ventures

Stephan Ellsworth Butler, E.I.T., AAAS Congressional S&T Policy Fellow[1]

Abstract: The U.S. government allowed the Interstate System to become compromised when it failed for decades to properly adjust for inflation the gasoline excise tax used to fund the system and increasingly diverted federal fuel-tax revenues to purposes other than roads. In addition, during the same time period, increases in the cost of road construction significantly outpaced inflation. Privatization, on the other hand, will bring about a series of benefits: it will provide new, much-needed, private capital; facilities will be built more quickly and more efficiently and operated more cost-effectively; execution and pricing risk will be transferred from the public sector to private investors, which can bear more risk and thus may develop more innovative solutions; and new sources of tax revenue will be created.

The Rise and Fall of the U.S. Interstate Highway System

In 1952, Dwight D. Eisenhower, who believed that completion of a "National System of Interstate and Defense Highways" was essential to the national interest, was elected President of the United States.[2] Under his leadership the question of how to fund the Interstate Highway System was resolved with enactment of the Federal-Aid Highway Act of 1956.[3] The program was proposed as a $50 billion program to be completed in 10 years, but it became a $450 billion, 50-year project, with the funding provided by gasoline excise taxes.[4]

However, due to a multitude of reasons discussed later in this paper, the level of funding provided by the excise tax has not kept up with inflation. The federal excise tax rate would now need to be $0.22 per gallon, not $0.184, to remain at 1956 levels, given inflation; the real excise tax rate is now 20% lower than it was in 1956.[5] But this is only part of the story. Between 1972 and 1993, the excise tax rate was an average of 55% per year lower in real terms than it was in 1956. Not surprisingly, during this period the United States saw its biggest negative change in the condition of its infrastructure.

Similarly at the state level, the volume-weighted average motor vehicle gas excise tax rate per gallon has not kept up with inflation and as a result is significantly below where it was in 1956 in real terms. In today's dollars, the average 1956 state excise tax rate on motor vehicle gas of $0.059 would be $0.44, which is 55% higher than the current average rate of $0.284 per gallon. As with the federal excise tax, this is only part of the story. Over the period from 1972 to 2006, the average state excise tax rate was 60% lower in real terms than the 1956 baseline.

Moreover, the problem of federal and state motor vehicle excise taxes not keeping up with inflation is only one of the factors contributing to the deteriorating conditions of the USA's roadway infrastructure system. Another factor is the growing diversion of federal fuel-tax revenues to purposes other than roads. In 1982, for example, a mass transit account was created within the highway trust fund, and today as much as 18 percent of trust-fund revenues paid by motorists are reserved for transit programs that benefit only a tiny fraction of commuters – currently about 5 percent.[6]

Not only have federal and state motor vehicle excise taxes not kept up with inflation and been siphoned off for other purposes, but demand for roads has increased. Since 1956, the USA's population has grown from 170 million to 300 million and vehicle miles traveled have increased from 650 billion to 3.1 trillion. In addition, the cost of construction for both rural and urban highways has significantly outpaced inflation.

As the world's economy becomes more integrated, there will be an ever-increasing economic premium placed on rapid, reliable transportation for goods and passengers. Accordingly, the ability to compete in the globalized economy will require well-connected, nationwide, high-capacity systems capable of high speeds and reliability. Notwithstanding the lack of U.S. governmental response, this problem is fully recognized by transportation experts. For instance, on October 30, 2006, the Board of Directors of the American Association of State Highway and Transportation Officials approved a series of policy recommendations for a Congressional Commission that in part read as follows: "Our generation inherited the world's best transportation system made possible by the commitment of the past two generations to invest in the country's future. We have spent that inheritance."[7]

Moreover, the pace of development in China underscores the challenge for America's transportation networks, as economic analysts suggests that China, armed with excellent transportation, a low-cost and enormous labor pool, and an aggressive policy for economic growth will supersede the Unites States as the dominant trading nation perhaps as early as 2020.[8] In particular, the rate of increase in the volume of containers being shipped from China through U.S. ports threatens to overwhelm the capacity of the U.S. system, and China's appetite for materials such as steel and concrete to fuel its massive construction program has put considerable upward pressure on the cost of construction in the United States.[9]

When considering both the importance of transportation infrastructure to the USA's economy (and hence the standard of living of the country's citizens) and the fact that the total increase to gasoline taxes required to restore the U.S. highway system would be in the neighborhood of $0.25 per gallon, it is bewildering that the country's elected officials have not passed appropriate tax increases, instead allowing the system to go from a catalyst of economic supremacy to a drag.

Earmarks and the Absence of Market Based Decision Making

The failure of the USA's elected officials to maintain and improve upon the system inherited from the prior generation is due to the use of earmarks, a lack of political will, the influence of special interest groups, and the absence of a market-based decision making process. In the book *The Broken Branch*, authors Thomas E. Mann and Norman J. Ornstein specifically call out the earmarking of the transportation highway bills since 1970 as an example of the systemic failure of government to take into account costs and benefits in its decision-making process. A particularly egregious example of earmarking occurred when the "bridge to nowhere," a $233 million allocation for a bridge from an Alaskan town of 14,000 to an island with a population of 50, was proposed as an earmark in the 2004 transportation appropriation bill.[10] Congress publicly purported to rescind that earmark as a result of "a major and sustained public outcry" but provided the exact same amount of money to Alaska, whose state officials made it clear that the bridge would be a priority item for them.[11]

Under this system of earmarks, every municipality has had its hand out, and the government has obliged. Under market conditions, each piece of infrastructure would have been subjected to marginal analysis, to confirm that the cost of building and maintaining it was outweighed by the benefit to be derived from its use. Under earmarks, unsustainable building policies and practices and dubious business models have ruled the day. Because it did not adequately model the infrastructure supply and demand requirements of the market place, the USA built an infrastructure system that it cannot afford to keep up, rebuild or add capacity to as needed. The USA misallocated its resources by building infrastructure projects that are not economically justified while failing to maintain those that are.

The rise and fall of the USA's infrastructure system is breathtaking. In the span of one century the USA built over 4,000,000 miles of roads and 590,000 bridges, which have now reached such a state of disrepair that dire headline-grabbing predictions warning of severe consequences are becoming the norm. Prophetically the warnings were eventually proven true when on August 1, 2007, the I-35 bridge, a structurally deficient , eight-lane, 1907 feet (581 m) steel truss bridge, which spans the Mississippi River in Minnesota, collapsed killing thirteen people and injuring one hundred. Even a natural disaster of epic proportions cannot make the system respond. As Christopher Cooper noted in a New York Times article on January 27, 2007, "In August, 2005 Hurricane Katrina flattened two bridges – one for cars, one for trains – that span the two miles of water separating New Orleans from the town of Pass Christian. Sixteen months later, the automobile bridge remains little more than pilings. The railroad bridge is busy with trains. The difference? The still-wrecked bridge is owned by the U.S. government, whereas the other is owned by railroad giant CSX Corp. of Jacksonville, Florida."

At the crux of the problem lies the USA's present infrastructure procurement system, whereby the government decides, independent of competition and market forces, how much money is to be raised through taxes, as well as what type of projects are to built and when and where to build them. A monopoly has effectively been created in favor of public officials, who have abused the system in pursuit of re-election as badly as any private entity without oversight would do in pursuit of profits.

Solving the underlying problem that has prevented the correct level and allocation of investment will require the root causes of the problem to be recognized by those in power. Finding enough politicians who will agree that the failure is in the political processes and policies that created the infrastructure system in the first place will be difficult, but not impossible. As citizens are educated about this problem by the experts – and get increasingly fed up by experiencing it first hand – they will cause their elected officials to listen to alternative solutions, or replace them with other candidates for public office who will.

While the USA quibbles internally about who should get what instead of making those decisions systematically based on the costs and benefits to society, and spends its limited resources on bridges to nowhere, nearly all the major players on the world stage are investing aggressively in their transportation infrastructure. China is building a 53,000-mile national expressway system and Europe is spending hundreds of billions of euros on a network of highways, bridges, tunnels, ports and a rail line.[12]

Privatization of Infrastructure

The present condition of the USA's infrastructure is of no surprise to a believer in competition and free markets. Government services are often costly and poor, not because the people who work in government are any less competent than those who work in the private sector, but rather because the state has no competition and is insulated from the market discipline required of private companies, which live under the threats of bankruptcy, having their management removed, or being taken over by competitors. Also, nationalized industries are prone to interference from politicians for political or populist reasons.

Pricewaterhouse Coopers' Privatization Infrastructure Finance Group contends that privatization of infrastructure will bring about a series of benefits: privatization will provide new private capital, and facilities will be built more quickly and more efficiently and operated more cost-effectively. New sources of tax revenue will be created, and greater customer satisfaction is likely to result, due to the financial interest of the private operator.[13] Privatization is fundamentally a device for capturing the value of infrastructure for the public and thus a means for redeployment of public capital from a place where it is no longer needed to other places where it is needed.

From the public's perspective the removal of non-technical personnel from the decision making process should be seen as a positive, as projects will be built on their merit and not though earmarks. Opponents of privatization argue that if there is so much value locked up in an existing facility, the government should re-finance the project and reap the benefits instead of investors. The Achilles heel of this argument is the completely unrealistic assumption, which this paper bears out, that the public sector could keep toll rates at market levels over 50, 75 or 99 years, thereby creating the value in the first place. Both political theory and real-world experience argue otherwise. For example an Indiana Toll Road called the Skyway, which was recently sold to private investors for $1.8 billion, had not increased it tolls for more than 20 years. Given the reality of inflation over these years, that means the real toll rates had been decreasing all those years.[14]

In order to bring rational decision making and competition back into the equation, while wrestling away some control of the roadway infrastructure sector from the federal and state governments, I propose that anywhere market forces are willing to design, build, construct, finance and maintain an infrastructure facility, the government should not. Simply put, the government should stay out of the way in areas of the country where private enterprise would be willing to invest in infrastructure projects if the proper regulatory framework was in place, subject to exceptions where critical to the national defense.

In order to take advantage of the capital markets' interest in investing in roadway infrastructure, the government should encourage and facilitate private enterprise's assumption of the work ordinarily done by state and federal authorities by creating a new agency, which for the sake of this paper will be called the Public-Private Partnership Administration (PPPA). The PPPA would function as the public's agent, making sure that deals get done while preventing abuses. The agency would retain the power of eminent domain in lieu of giving that right to private entities, and it would make sure that environmental impact statements and laws are fully vetted and reviewed. Also, since a natural monopoly would be created when a private enterprise is allowed to build a roadway concession, the PPPA would be armed with anti-trust legislation and bodies to deal with anti-competitive behavior in these concessions.

Private industry, with the proper regulatory framework in place and armed with the knowledge that government is willing to concede construction of roadways in areas where traffic congestion allows for facilities to be built that could collect tolls and reduce congestion, would jump at the opportunity to provide these services. The indirect benefit of this is that it would optimize the government's regular program of road building, by allowing it to allocate limited public resources to help fund upgrades and repairs in areas where private users cannot afford themselves to pay for necessary facilities. Public funding of these roads should continue, since these roads are essential for national defense, security and the enjoyment of the natural world, but it should not be increased without serious deliberation.

If the PPPA is successful it is expected that an entire new business sector will be created which will do nothing but analyze markets for their ability to support toll roads, which will compel them to approach the government and let them know of their interest in investing. As such it can be anticipated that private enterprises will let the government know where and when new markets have become viable for private investment with unsolicited proposals, which will put the government on notice that the public is no longer required to invest capital in those areas.

Conclusion

Given government's track record, it is unrealistic to assume that the US's elected officials could find the discipline to raise tolls or excise taxes to the levels necessary to fund the system over the long term. Even setting that aside, however, a bigger issue remains: The USA's present infrastructure procurement system, where the government decides, independent of competition and market forces, how much money is to be raised through taxes, and where, when, and what type of projects to build, is inefficient. Without market pressures, government officials have more incentive to allocate resources to their own constituents back home, so that they will vote for such politician's re-election, than to come up with an allocation that is efficient or in the national interest.

Privatization, on the other hand, will bring about a series of benefits: it will provide new, much-needed, private capital; facilities will be built more quickly and more efficiently and operated more cost-effectively; execution and pricing risk will be transferred from the public sector to private investors, which can bear more risk and thus may develop more innovative solutions; and new sources of tax revenue will be created. However, the precise economic incentives and regulatory controls that would lead to the best results need to be developed through additional research.

[1] Office of US Senator Mary L. Landrieu (D-LA), 724 Hart Bldg., Washington, D.C. 20510; (202) 224-0084; email: Stephan_Butler@Landrieu.Senate.GOV

[2] Richard F. Weingroff, "The Greatest Decade 1956-1966," FHWA, www.fhwa.dot.gov/infrastructure/50interstate.cfm [accessed September 2007].

[3] It is assumed that the initial excise tax ($0.03 per gallon) was established after careful analysis. Therefore this paper uses 1956 as a baseline for comparisons because that was the year of enactment of the Highway Revenue Act.

[4] Richard Prentice, "America's Highways," PE, November 2006, 8.

[5] Inflation adjusted results were calculated by taking into account changes in the Consumer Price Index as published by the Bureau of Labor Statistics.

[6] Ronald D. Utt, Ph.D, "Our Roads Suffer Because Federal Program Ships State Funds North," Heritage Foundation, www.heritage.org [accessed September 2006].

[7] AASHTO, Transportation – Invest in our Future: Future Needs of the U.S. Surface Transportation System [Washington, D.C.: AASHTO, February, 2007], 15.

[8] Ibid, 27.

[9] Ibid.

[10] Thomas E. Mann and Norman J. Ornstein, The Broken Branch [New York: Oxford University Press, 2006], 218.

[11] Ibid.

[12] AASHTO, Transportation – Invest in our Future: Future Needs of the U.S. Surface Transportation System [Washington, D.C.: AASHTO, February, 2007], iii.

[13] Savas, E. S., ed. 1992. Privatization for New York: Competing for a Better Future. The Lauder Report. New York.

[14] Poole Jr., Robert W., "A Response to Critics of Toll Road Leasing," Public Works Financing, no. 205, May 2006, 3.

Innovative Cost and Schedule Risk Assessment for Large Transportation Projects

Charles W. Lockhart, P.E., CCM.,[1] Travis C. McGrath, Ph.D., P.E.,[1] and James P. Rost[2]

[1] Golder Associates Inc., 18300 NE Union Hill Road, Redmond, WA 98052; (425) 883-0777; email: clockhart@golder.com; tmcgrath@golder.com
[2] Iowa Department of Transportation, 800 Lincoln Way, Ames, IA 50010; (515) 239-1726; email: jim.rost@dot.iowa.gov

Abstract: Accurate project cost and schedule estimates are a critical component of transportation project decisions. While many owner/agency organizations have had successful estimating experiences, worldwide data suggest that engineers and planners have systematically under-estimated the ultimate costs and schedules for large transportation infrastructure projects; one study suggests up to 90% of estimates for transportation infrastructure projects have been low and, on average, cost estimates for transportation projects have been 20% short of final costs (Flyvbjerg, 2002). This paper briefly summarizes an innovative and flexible probabilistic, risk-based approach to addressing these problems in a practical and cost-effective way.

Introduction

Even when a project is well planned, project conditions can change over time and unanticipated events can occur. These changes and problems, compounded by inevitable estimating biases (ranging from pessimistic to optimistic), can result in significant, undesirable consequences, including cost and schedule over-runs or under-runs, resource competition among projects, negative media attention, and public mistrust. These changes and problems are difficult to predict in advance because they are influenced by a number of variables, including nature (e.g., ground conditions, weather), technology (e.g., design, methods, equipment, materials), and human (e.g., labor, public, political, regulatory agencies, funding, insurance, bonding agencies, and market conditions).

Many government agencies, as well as private owners, now recognize these issues and are beginning to require risk-based cost and schedule estimating, as well as formal risk management, to address these problems. Risk-based estimating explicitly identifies and quantifies uncertainty in the variables affecting project cost and schedule, and therefore quantifies uncertainty and confidence in the ultimate project cost and schedule.

The probabilistic, risk-based approach summarized in this paper is described more fully by Roberds and McGrath (2006). For convenience, the philosophy of the approach and the steps in the approach are summarized below.

Risk-Based Approach versus Traditional Approach

Risk-based methods such as the approach outlined in this paper offer a more explicit treatment of uncertainty than more-traditional estimating methods which use contingencies to account for uncertainties. Important differences include:

- Probabilistic, risk-based methods quantify costs with probability distributions (ranges of possible values, each with a relative likelihood), which quantify confidence in any particular value or range of values. Traditional methods usually produce a single value with unknown confidence.

- Risk-based methods identify and characterize distinct risk and opportunity events and other uncertainties by source and/or type, whereas traditional methods typically combine these together into a single contingency value.

- Risk management can be ad-hoc for traditional methods, but is explicit with a risk-based approach because the risks and opportunities are individually characterized and the cost-benefit of any proposed activities can be determined.

Despite these differences, traditional and risk-based methods share one important trait: both rely on the judgment of experts. Because few projects share the same scope, strategy, and conditions, data on future project performance are typically limited. Therefore, the opinions of experts are typically the best source of information for estimates. Risk-based approaches make the most of that information by quantifying the expert opinions and related uncertainties through carefully elicited subjective assessments.

"Base + Risk"

The approach quantifies uncertainty in project cost and schedule using a "base + risk" approach. The base + risk approach essentially replaces the more traditional estimating approach (cost plus contingency) with an unbiased "base" component, and replaces contingency from the traditional approach with a "risk" component (Figure 1). The approach integrates cost and schedule.

The base represents the complete project as planned (i.e., the assumptions made for the estimate are correct), which generally means without contingency, conservatism (to the extent possible), and float. Significant uncertainties within the base assumptions are included, as are correlations among uncertain base activity costs and durations ("base factors").

Risk and opportunity events represent potential deviations from the base assumptions (i.e., that the project may not go as planned). Risk is defined as probable loss, in terms of the combination of additional costs and/or durations to affected activities and the corresponding likelihood of occurrence. Conversely, opportunity is defined as probable benefit, in terms of reduced cost and/or duration and the corresponding likelihood of occurrence. Significant correlations among risk and opportunity events are included as appropriate. The base is combined with risk and opportunity through Monte Carlo simulation to quantify uncertainty in cost and schedule.

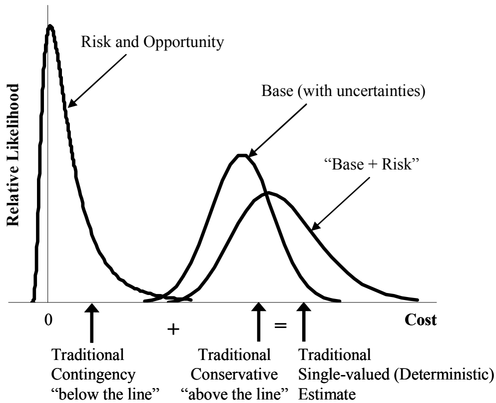


Figure 1. "Base + Risk" versus Traditional Approach.

Steps in the Approach

The basis for the risk assessment is a collaborative effort carried out in a group format which consists of the project team, outside subject matter experts (SMEs), and trained risk elicitors. The typical steps in the approach are as follows.

1. Structure the Risk Assessment. The project team presents the project scope, status, key assumptions, and strategy for delivering the planned project. The group then develops the sequence of major project elements, including activities, milestones, and decisions. This sequence is documented in a project "flow chart," which forms the basis for the integrated cost and schedule model (Figure 2).

2. Review Cost and Schedule Estimates. The group reviews the project's "traditional" cost and schedule estimates through a collaborative effort between the project team and independent SMEs. The review leverages the knowledge and experience of each group member to substantiate or clarify project assumptions, confirm the scope, schedule and cost estimates, and confirm that the design is reasonable. Often, ranges are used for cost elements to reflect cost uncertainties. In the process, the group identifies and removes any explicit or implicit contingencies, which are instead handled separately as part of the risk and opportunity review. This becomes the project "base" cost and schedule.

3. Develop the Risk Register and Assess the Risk Factors: Once the base is defined, the risk elicitor works with the group to identify a comprehensive set of risks and opportunities (i.e., events) which are documented in a "risk register." The elicitor then leads the group through characterization of each significant event in terms of its "risk factors." Risk factors for a particular risk include the consequences of occurrence (in terms of cost changes and/or duration changes to specific activities on the flow chart if the event occurs) and the corresponding likelihood of

occurrence. Any significant relationships or correlation among risk events (either the occurrences or the uncertain consequences) are also quantified as appropriate. The elicitor ensures that accurate and defensible assessments are developed for the consequences and likelihoods by identifying and mitigating biases. To the extent possible, the goal is to achieve consensus from the collective group of project team members and independent subject matter experts.

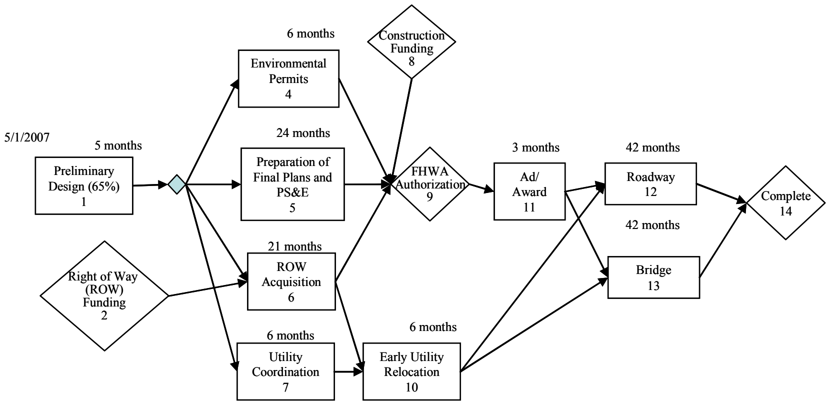


Figure 2. Example Project Flow Chart with Activities

4. Develop the Cost and Schedule Uncertainty Model. The authors recommend an integrated cost and schedule model for infrastructure risk assessment. Several commercial software packages are available for project scheduling, and at least one Monte Carlo simulation add-in is available for each; however, not all of the commercially-available packages can conduct true probabilistic, risk-based, integrated cost and schedule modeling, and those few that can are typically cumbersome. Model flexibility is important to meet varying owner needs, so the authors prefer spreadsheet-based models with a suitable Monte Carlo add-in.

5. Evaluate the Results. Typically, the results are probability distributions (PMF and/or CDF) for project cost and schedule and ranked lists highlighting the critical risks and key opportunities with respect to project cost and schedule. This facilitates the development of a risk management strategy in order to mitigate critical risks or exploit key opportunities.

Case Study of a Major Bridge Project

A recent risk analysis was completed for a proposed bridge replacement project. The existing paired 2-lane bridges carry a major interstate highway over a large river in the mid-western United States. The structures will be replaced with paired 3-lane structures in order to increase capacity, safety, and reliability of the highway. Roadway improvements will include widening, some structure replacement at both approaches, and demolition of the existing river bridges. The agency wanted to evaluate and compare uncertainty in project cost and schedule for three fully-funded alternatives, where each funding alternative reflected a different strategy for when the project could be funded.

The results of the risk assessment for the "baseline" funding alternative are presented in Figures 3 (cost) and 4 (schedule). The probability distributions can be presented as both probability mass functions (PDF) and cumulative distribution functions (CDF). A CDF for cost shows the probability (or confidence or percentile) of not exceeding a given cost. For example, there is an 80% likelihood that the ultimate project cost will be less than about US$ 1,250 million (Figure 3), and will be completed by February 2023 or sooner (Figure 4).

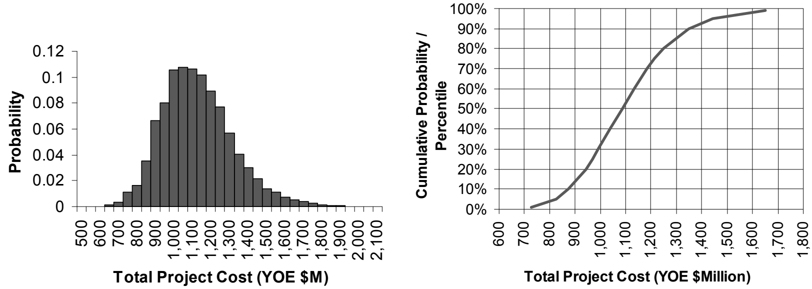


Figure 3. Probability Distribution for the Project Cost in Year-of-Expenditure dollars, in terms of probability mass function (PMF) and cumulative distribution function (CDF).

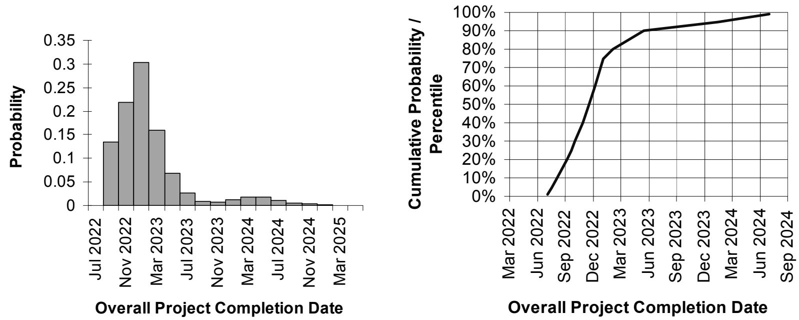
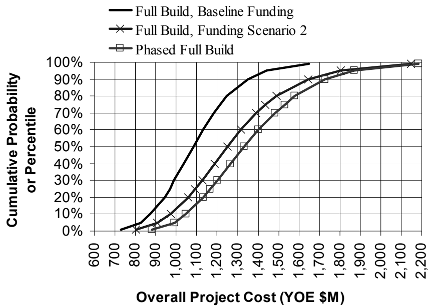


Figure 4. Probability Distribution for Date the Overall Project Completion Date, in terms of probability mass function (PMF) cumulative distribution function (CDF).

Figure 5 presents comparisons of cost and schedule probability distributions for the three funding scenarios evaluated in the risk assessment. For example, the ultimate project cost at the 80% confidence level ranges from about $1,250 million (unconstrained funding) to $1,600 million with an associated schedule increase from year 2023 to 2030. This comparison allowed the agency to directly compare the costs and associated project delivery dates, and to evaluate how the project would fit into their overall capital program planning.



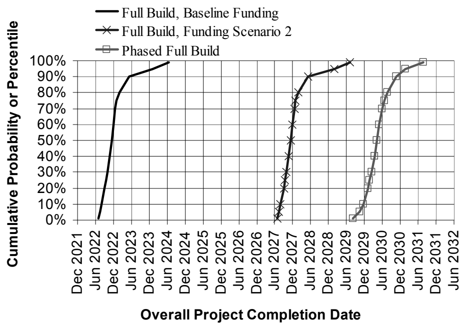


Figure 5. Comparison of Project Cost (in Year-of-Expenditure dollars) and Schedule for the Three Funding Scenarios (in the form of Cumulative Distribution Functions)

Conclusion

This paper has presented the framework, methodology, and an example application of a probabilistic, risk-based cost and schedule assessment and management approach for large transportation projects. The authors have applied this approach successfully to a large number of infrastructure projects worldwide. The approach is based on established fundamentals, but continues to evolve to better serve the needs of project owners for transportation and other capital projects. As projects are completed, and final costs are determined, comparisons between actual costs and the predicted ranges will demonstrate the validity of the process.

References

Flyvbjerg, B., Holm, M.S., and Buhl, S. (2002), "Underestimating Costs in Public Works Projects: Error or Lie?" Journal of the American Planning Association. Vol. 68, Issue 3.

Roberds, W. J., and McGrath, T. C. (2006) "Quantitative Cost and Schedule Risk Assessment and Risk Management for Large Infrastructure Projects," *Proceedings*, The Project Management Institute College of Scheduling Conference, Orlando, Florida.

Integrated Modeling for Regional Transport Planning: Salt Lake Region, Utah

Andy Li[1], Ned Hacker[1], Chuck Chappell[1], John Lobb[2]

[1]Wasatch Front Regional Council, 295 N. Jimmy Doolittle Rd. Salt Lake City, UT 84116, USA; PH (801) 363-4250; FAX (801) 363-4230; email: andyli@wfrc.org
[2]Resource Systems Group, Inc., 55 Railroad Row, White River Junction, VT 05001, USA; PH (802) 295-4999; FAX (802) 295-1006; email: jlobb@rsginc.com

ABSTRACT

The Wasatch Front Regional Council has developed and tested the UrbanSim land use modeling system for the Greater Salt Lake Region of Utah. The UrbanSim model is integrated with the regional travel models, and this integrated model system is an advanced analytical framework used to help develop long-range land use forecasts and to evaluate land-use and transportation scenarios in the Regional Transportation Planning process. This paper will provide an overview of the WFRC modeling process, and describe the recent application of the modeling system during the Wasatch Choices 2040 Visioning effort.

BACKGROUND

The Wasatch Front Regional Council (WFRC) is the designated Metropolitan Planning Organization (MPO) for the Salt Lake City and Ogden Urbanized Areas of Utah and is responsible for transportation planning in the region. Prior to UrbanSim being implemented by WFRC, land-use growth forecasts in the region were not explicitly influenced by or tied to where transportation infrastructure improvements are planned. The previous land use forecasting procedure is based on a trend-based model to allocate new households and jobs and this forecasting process relies on considerable review and adjustment by staff and by planners in the region.

UrbanSim is designed to support metropolitan planning and policy analysis in a more scientifically rigorous manner than the land-use model previously used by WFRC, and one important advantage is that land-use forecasts are influenced by the proposed transportation system. By integrating UrbanSim with the regional travel models, a range of land use and transportation policy interventions can be combined into policy scenarios, and the systematic effects of these scenarios can be explored on urban development outcomes and the quality of the transportation system.

UrbanSim is a complex model system that is challenging to implement. WFRC staff have been working to implement and refine UrbanSim for several years. Recently, WFRC staff used the integrated model system during a regional visioning study, in which the results from the model system were used to analyze the impacts of alternative long-range land-use and transportation scenarios. Additionally, the integrated model system was used to develop official land-use projections consistent with the consensus vision for the Regional Transportation Plan (RTP).

LAND USE AND TRAVEL MODEL OVERVIEW

WFRC's integrated model combines land-use, transportation, and air quality models designed to perform a wide range of analyses. The model system includes several advanced features that place it on the cutting edge of improved modeling methods. The addition of the land use model in theory enables development and analysis of consistent and realistic land-use and transportation scenarios. WFRC's integrated model has a feedback loop whereby accessibility affects future land development, which in turn affects travel patterns and accessibility.

UrbanSim (Waddell, 2004) is an agent-based microsimulation model designed to model the real estate market in a behaviorally robust way. UrbanSim models the decisions of developers, households and businesses, in response to user-specified inputs representing government actions with respect to land and infrastructure. Developers use land to construct housing and nonresidential floor space that are demanded by households and businesses, which are also interacting in the labor market and the markets for goods and services. Governments provide infrastructure and services, while regulating and in some cases altering prices for the use of land and infrastructure. This general framework provides a point of departure for considering the effects of alternative governmental policies and investments. More complete UrbanSim documentation can be found at www.urbansim.org.

IMPLEMENTATION CHALLENGES

In practice, there are a limited number of MPOs in the U.S. that have brought agent-based, microsimulation land-use models into extensive operational use. To date, MPOs for several large cities are in varying stages of developing an UrbanSim implementation, including, but not limited to: Honolulu, HI; Houston, TX; Seattle, WA; Phoenix, AZ; and Salt Lake City, UT. The Salt Lake City and Houston MPOs have used results from UrbanSim in the planning process.

UrbanSim is a complex forecasting tool and there are several challenges to implementation, such as data needs, choice model development, the feeding back of transportation accessibility, and model validation. We discuss each of these in turn.

Challenge 1: Data needs - UrbanSim attempts to mimic the complexity of urban land markets, and the data requirements are significant. The necessary data include detailed descriptions of current land-use, characteristics of households and businesses, future land-use policies, and environmental constraints. Our version of UrbanSim operates at a 5.5-acre resolution, which means that all the above data must be accurate and consistent at this scale. A typical database development effort involves combining data from tax assessor parcel files (current land use), with data from various political and environmental GIS layers (zoning, slopes, water, etc.). Once the existing real estate is described, the buildings must be "populated" with specific households and businesses. Business data is typically available in employer databases and the households are synthesized to be consistent with census data.

In our experience, 3-4 staff worked for over one year to develop a reasonable database, and the most troublesome aspects were the lack of consistency among the different tax assessor databases in the region, the lack of critical real estate data in some counties (year built and building size), the difficulty reconciling business addresses with business locations in the parcel file, and the need to compile and standardize the future land-use policies of over 65 cities and towns into a GIS.

Challenge 2: Model Estimation - UrbanSim is parameterized via a local model estimation effort. While the overall model design can vary considerably from one implementation to the next, a typical UrbanSim modeling system includes a few dozen estimated models to predict land values and the actions of developers, households and businesses. A typical UrbanSim modeling system includes many different types of developers, households and employers, all with unique model coefficients. Our initial model estimation effort took over 1 year, after the base year data was compiled, and refinements to the models continue.

Challenge 3: Transportation Accessibility Feedback - The rationale behind integrated land-use and transportation modeling is that it is behaviorally realistic and allows a model to represent both the effect of land development on transportation and the effect of improving the transportation system on land development. UrbanSim is sensitive to the quality of the transportation system and the WFRC UrbanSim models capture this sensitivity via two types of transportation related variables, *accessibility to activities* and *proximity to transportation facilities*.

The first type of transportation variable is a function of the logsum from the mode choice model, which is a measure of how easy it is to travel by all modes. The logsum is interacted with the spatial distribution of jobs and housing to derive summary measures of accessibility to activities. The second type variable used in the WFRC application of UrbanSim measures the distance to the nearest arterial and the nearest freeway segment. As the transportation network expands, it is reasonable to expect that development will occur to some degree along with the road network.

The transportation system helps shape an urban area, but to have confidence in an integrated model it is necessary to have confidence in the sensitivity of the land-use model to changes in the transportation system, which is challenging to measure and validate. While UrbanSim is flexible and offers theoretically elegant methods for incorporating transportation measures into land-use forecasting, in our case the sensitivity of the land-use forecast to the transportation measures needs to be analyzed further so that the sensitivities are understood and appear reasonable.

Challenge 4: Model Validation - The best way to demonstrate the reasonableness of a model is to compare estimate outcomes from the model to observed outcomes. For land-use models this might entail forecasting land development from 1990 to 2005, for example, and comparing the predicted outcome to what was observed. However, this is difficult to do with UrbanSim because of the complexity of the base year database; the historical data would be very difficult if not impossible to obtain. In our case we did not have these data available.

The validation of our UrbanSim modeling system is an on-going process. The land price models are relatively easy to validate against current data. The development and location choice models are more difficult to validate because the models are probabilistic choice models. For these models we have compared the estimated probabilities to actual locations where households and businesses have located, looking for high probabilities in places where each type of household and business locates today. To validate and evaluate the overall model system we are comparing observed historical trends against the future projections to be able to compare the future estimated growth against the context of growth over the last 20 years. This process requires a lot of judgment but is a useful exercise.

USE OF MODEL FOR LONG RANGE PLANNING

Every city in the region has a comprehensive plan and a future land use zoning plan, which can be combined to create a picture of how growth in the region might look in the future. But the region might not look exactly like the growth planned by these cities. There are many different factors influencing growth and as a result many plausible development outcomes are imaginable. Theoretically, each land use scenario has its own most efficient transportation system. One transportation system cannot serve every land use scenario equally efficiently.

The land use pattern and the corresponding transportation system both play a critical role in determining the livability and sustainability of urban areas. It is important to model them in an integrated way to reflect the strong interaction between land use and transportation. For long-range planning purposes the model input databases can be changed to test the usefulness of different scenarios by varying the land-use and/or transportation policies.

WFRC's ***Wasatch Choices 2040*** planning process (Envision Utah, 2006) was an approach to creating a regional transportation plan which involved many technical planning and public involvement tasks. Based on the output from workshops, four land use and transportation scenarios were used to test various growth and transportation ideas:

- **Scenario A** – "Business as Usual", is based on the existing, adopted city, county and regional land-use and transportation plans.
- **Scenario B** – "Transit Station Villages", emphasized urban development in transit station villages and significantly increased the amount of rail transit.
- **Scenario C** – "Interconnected Network of Complete Streets", intensified mixed-use development along interconnected and walkable boulevards.
- **Scenario D** – "Centers of Employment", envisioned more suburban centers of employment in closer proximity to more diverse housing areas. The construction of new interstates and major roads is emphasized.

The output from each scenario came primarily from WFRC's integrated model system. The model system produced estimation of future land-use and transportation system performance. The examination of the scenarios and evaluation criteria resulted in some interesting observations. For instance, different patterns of development can improve vehicle accessibility but exacerbate mounting transportation challenges. Secondly, some of the proposed future development patterns can help mitigate transportation challenges and reduce the high cost of providing transportation infrastructure. The following are more specific observations:

- **Mixed-use development reduces driving distances and congestion.** The distance traveled to work, shopping, schools, or parks is largely a function of the distance between these destinations and residences. If commercial destinations, like an office or restaurant, are very close to each other and are set in a pedestrian-friendly setting, many people will choose to walk between them, rather than drive their vehicle. People will walk or use bicycles if the trip is short and the design (for pedestrians and cyclists) is convenient. The distance traveled per person directly affects the collective time it takes people to travel and the traffic congestion they are part of. Scenario C significantly reduced diving distances, reduced traffic congestion and improved air quality.
- **Urban growth near transit opportunities encouraged people to ride transit.** Scenario B shows that if transit stations or stops are within walking distance of homes or businesses, more people find transit to be convenient.
- **Transit-oriented development is a key strategy to increase redevelopment in existing built areas.** Scenario B's emphasis on high capacity transit coupled with transit villages created more opportunities for reuse of land.
- **Transportation systems help to determine where growth will occur and how much land area will be developed.** New transportation facilities, increase accessibility, which in turn attracts growth. As planners and decision-makers consider where they should invest transportation dollars, they should ask the question, "Where do we want to encourage new growth: on under-utilized industrial and other urbanized properties, on vacant land near existing communities, or in new undeveloped areas?"
- **Interconnected streets help keep short trips off highways and reduce congestion.** Interconnected streets facilitate smooth traffic flow and the use of more direct routes. They also promote neighborhood cohesion. The length of time it takes to reach destinations is a function of distance as well as congestion. Shorter driving trips and less congestion mean that if the region develops using the strategies embodied in Scenario C, there will be more time available for people to pursue individual choices and less time in congestion.
- **Strategic changes make a big difference.** Surprisingly, the benefits of Scenarios B, C, and D, when compared to the "business as usual" scenario, are the result of relatively minor, but strategic, changes to the density of the region's housing and land use. The strategic placement of this type of development in walkable and mixed use settings adjacent to transit is largely responsible for the advantages that Scenario C anticipated, including almost a 10 percent reduction in congestion. Strategic changes throughout the region can vastly improve the individual quality of life.

CONCLUSION

Based on our successful experience, it is concluded that the integrated modeling can be a useful analytical process for the improvement of land use forecasting, testing of transportation strategies in relation to land use and growth projections, and development of various and matching land use scenarios for different transportation scenarios in the Regional Transportation Planning process. Our land use forecasting platform, UrbanSim, is recommended as an effective and efficient analytical tool for the integration of land use and transportation.

REFERENCES

Envision Utah. (2006) "Wasatch Choices 2040: A Four County Land-Use and Transportation Vision"

Paul Waddell and Gudmundur F. Ulfarsson. (2004) "Introduction to Urban Simulation: Design and Development of Operational Models." In *Handbook in Transport, Volume 5: Transport Geography and Spatial Systems*, Stopher, Button, Kingsley, Hensher eds. Pergamon Press, 2004, pages 203-236.

Integration of Land Use and Transportation
- Development around Transit Systems

Peter Lai
PB Americas
Princeton, New Jersey

Abstract

Urban cities around the world are busy finding ways to manage traffic congestion that increasingly reduces the quality of life of the citizens and the productivity of the city. Many governments of developed countries had already learned that building your way out of congestion is totally cost ineffective and counterproductive, and is certainly not a recommended long-term strategy for congestion management. Instead, transportation professionals are calling the attention of the policymakers to focus on integration of land use and transportation as an effective mean to manage congestion.

While coordinated land use planning can certainly help to enhance the movements of people and goods for the two primary modes of transportation, automobiles and public transport, the presentation will concentrate on the public transport sector instead of the highway sector. It will introduce the concept of Transit Oriented Development including background and principles, and the discussion will touch on of the following technical elements:

- Factors that favoring and hindering development around transit
- The elements of Transit-Oriented Planning
- Accommodating the terminal/station function
- Principles for successful development around stations/terminals

Introduction

In order to gain a better understanding on the concept in integrating land use and transportation, it would be helpful to begin with an overview of the big picture that describes the relationship among the five major elements in urban development and transportation.

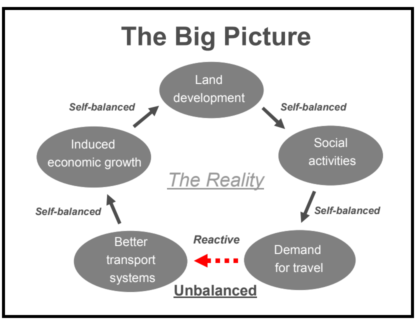


Figure 1.

As illustrated in Figure 1, new land developments will attract and/or produce new social activities to the location, and these extra activities will generate additional travel demand; as travel demand increase so will the needs for more transportation systems and services to accommodate the demand. With the support and implementation of new or improved transportation systems and services, additional economic growth will be induced, which in turn, would simulate the market for more land development.

Under an ideal condition, the above hypothesis would interact perfectly as the development cycle continues its course, however in reality; the outcomes are not exactly as expected because not all of the relationships between these elements are self-balanced and self-adjusted. For examples, the interaction between land development and social activities is self-balanced such that the amounts of new activities is a function of the new developments added; the same also can be expected between social activities and travel demand, between transport systems and economic growth.

However, it will not be the case for travel demand and transportation systems because the increasing in travel demand will not automatically triggered the capacity increasing in transportation systems and services. In fact, the relationship between these two elements is always in an unbalanced and reactive nature. This condition helps to explain the reason why congestion continues to come back not matter how much transportation investments are committed.

In order to balance the transportation demand and supply relationship, considerable amounts of investments will be necessary, while these investments constitute in various types and forms, there are certain areas of commitment that are very important, they include:

- Planning visions – how well we anticipate the future and how appropriate we prepare for them
- Land and right-of-way – this is the basic requirement for any kind of transportation system expansions or in some cases for service enhancements as well
- Financial resources – stable financial commitment would be required for planning, design ,implementation and operation of the needed projects
- Political determination – ability to make the appropriate political decisions and the commitments to standby the transportation policies are very critical
- Public acceptance – to obtain the support from the general public on transportation policies and projects is perhaps the most challenging and crucial parts of the mission

Another critical element is the development of an integrated planning approach to ensure that the transportation investments are spent effectively. Traditionally, the planning approach taken was single-mode and single focus in nature, and it typically produced individual stand-alone plans and reports, such as land use plans, highway development plans, public transportation plans, freight plans, etc.

As the nature of the transportation market changes over time, a new innovative planning approach has quickly earned the recognition as the better and more efficient approach under the modern day environment. In contrast to the traditional approach, the innovative approach is multiple-focus and produces integrated and coordinated multi-modal plans. Some example products include integrated land use and transportation plans, regional intermodal plans, and corridor management plans.

One of the most challenging elements in taking the new approach is the requirement of significant cooperation and partnership from all stakeholders, in particular, the public agencies that responsible for various components of the transportation systems.

The concept of Transit-Oriented Development (TOD)

There are two categories of potential opportunity for developments at transit. One category is Joint Development, which involves developments that build on public owned land typically atop of rail stations, many developments along the Hong Kong MTRC metro lines are excellent examples of joint development. The other category is Transit-Oriented Development where it is more than just developments above the rail stations; it goes beyond the station property onto the immediate surrounding area typically within a comfortable walking distance away (5-10 minutes).

Transit-Oriented Development or TOD in short, is one of the new innovative strategies that attempt to balance the relationship between land use and transportation. The concept encourages the forming and shaping of land developments that enhance potential market for transit services; it also helps to build better community by creating a sense of identify for the area.

There are several important factors that help to set a favorable environment for the growing popularity of TOD concept. They include:

1. Increasing concerns on the harmful effects created by automobile travel in terms of congestions and pollutions
2. The increasing difficulty in adding road capacity to keep pace with new developments
3. The growing popularity and success of the Smart Growth & Place-Making concepts
4. The emphasizing of high density developments also helps to provide better environment for transit services
5. More important, the TOD concept is very flexible because it can apply to many forms of public transport services such as local and express buses; BRT (bus rapid transit), light rail, rapid rail and commuter rail.

The TOD concept builds on several important planning principles. The most important one is that the quality of pedestrian connections between the stations and the origins or destinations is critical to the land use and public transportation relationship. Focusing on pedestrians will not only help to enhance transit access and improve its market, but also helps to create an attractive urban place, therefore within walking distances from the stations, the design and mix of land uses, must strongly promote walking. On the other hands, good physical design is important to the stations, but so is the appropriate mix of activity to complement the stations.

In short, the important TOD guiding principles include: (a) to provide a walk-able environment, (b) to provide appropriate mixture of land use in creating a defined center with 18-hour of activity, and (c) to plan and design for community instead of individual projects.

Past experiences from a variety of TOD projects also indicated that different locations may require different variations of similar principle, and a list of ingredients needed for successful developments around transit is recommended. The list includes:

1. To take a long-term view and do the appropriate planning at the beginning of the project
2. To create a flexible and realistic vision but must focus on its practicality and implementation
3. To build communities not projects, a series of community-supported projects will make the TOD concept more marketable. However, for this principle to be effective, multi-agency cooperation & coordination is absolutely necessary.
4. To transform the station areas into complete, walkable 18-hour activity center by establishing the right mix and combinations of developments including residential, commercial and whatever make sense for the communities by mixing different land use types.
5. To design the facility and operation for the customers – the pedestrians and the transit passengers
6. To establish public/private working partnerships with the communities and the local authorities in developing strategies and implement changes.
7. To provide efficient circulations for internal and external traffic including parking
8. To be flexible in both planning and design stages in order to incorporate internal and external inputs and feedbacks

In addition to the above, leadership and commitment are critical in order to keep TOD projects moving toward their final goals. Political champion and leadership will be needed to assemble resources, to build coalitions and to resolve disputes and conflicts.

One of the major challenges to be successful in TOD project is to be flexible and do not afraid of changing the traditional practices and standards, in particular, implementing policies and processes that are not commonly done before. In short, commitment to changes should take place in following four (4) areas.

- Land use - the mixed use concepts would require an open mindset in creating the appropriate mix and combination to serve the TOD core areas. Very often, modifications and changes on existing zoning laws and regulations are required. In such cases, high level communications and cooperation will be required from all levels of local authorities.
- Orientation - the most sensitive design element is the building setbacks. Because in TOD concept, its planning and design must focus on pedestrian, and the appropriate building setbacks would sometimes make a big difference. Here are the values for US cities only where pedestrian volumes are usually very low, for China cities, higher values would be required
- Scale - the development size in particularly building heights within the TOD core area would have an impact on the visualization and urban design of the entire project. Unfortunately, there are no success formulas or rules because the design theme and emphasis are different from project to project. And again, modifications to existing regulations may be required in many cases.
- Access – the key is to create a pedestrian-friendly environment and the road network in the entire TOD area must be designed to meet the needs for both pedestrians and vehicles. Different types and classifications of roads may need to be rearranged and integrated to work together, and there will be time where modifications to existing regulations and practices are needed.

Past experiences had indicated that TOD will not only generate considerable interests from land developers, it would also enhances the value of public investment in many ways - including added value for transit agencies, the cities, communities and their citizens; It helps to finance the features that make TOD works; and also helps to lower the transit cost serving the TOD areas. As the results of the above, TOD will help to improve the capture market of the public transit systems.

One of the excellent examples is the Rosslyn Ballston TOD corridor in Virginia, which is the most experienced TOD project in US. The TOD Plan was adopted in 1974 and the station opened in 1979; and since 1980, approximately 2.5 million square feet of office and 14,400 residential units had been developed. At once time, the corridor captured 25% of the county's housing supply and 37% of the county's job market. Furthermore, while it consists only 8% of the county's total land, the TOD corridor collected 33% of the county's property taxes with a market value of $9 billion dollars.

Even more significant is the fact that while typical mode split of suburban corridor is 15% by walk and 58% by automobile, the Ballston TOD Corridor showed 73% of the Metro accesses are made by walk and only 13% made by automobile. This TOD project has demonstrated the fact that well integrated land use and transportation development can indeed help to change the travel behavior of the general public.

Summary

In United States, TOD will continue be a growing market supported by both land use and transportation professionals, however, for every opportunity, there will be local issues and concerns that must be proper addressed. Similarly, China is still riding a massive development and redevelopment cycle, and is emphasizing the importance of public transportation; therefore the concept and applicability of TOD will prove to be very attractive in China. Consider the development environment and business practice in China, it is recommended that attention to be placed on the following areas:

1. Define a vision for the community and link all the goals and objectives toward the defined version
2. Conduct all the planning elements up front and get them done right
3. Identify all the stakeholders and understand the approval process required. If possible, streamlining the local review and coordination process
4. Understand the demand for TOD market and be realistic about the financing aspects of the entire project
5. Must educate and provide guidance to the developers, the local leaders and all the decision-makers on land use and transportation.

Study of Coordination Between Urban Transportation and Land Use

LUO Ming[1] ZHAO Yan-feng[2] CHEN Yan-yan[1] LIU Xiao-ming[1]
(1 Beijing University of Technology, Beijing Key Lab of Traffic Engineering, Beijing, China, 100022, Email: luoming@emails.bjut.edu.cn
2 China Academy of Urban Planning and Design, Beijing, China, 100037, Email: data_rover@126.com)

Abstract: Urban transportation planning in Beijing has not been well integrated with land use strategy. Limited land supply cannot meet the demands placed on the transportation system, resulting in newly constructed roadways becoming congestion after a short period of time. Using the method of data envelopment analysis (DEA) and the idea of membership degree of fuzzy mathematics, the evaluation model of a composite system was proposed. State and dynamic degrees of coordination displayed the different temporal concepts. Indexes were then selected and the degree of coordination for Beijing was calculated. From this analysis the conclusion can be drawn that development and transportation in Beijing have not been coordinated in the past several years. This conclusion is also supported by objective data. The example verifies the viability of the model.

Key words: transportation; land-use; coordination degree; composite system; data Envelopment Analysis

1 Introduction

Traffic demand is a product of land-use, and the orientation of the transportation system ultimately impacts land-use. The two systems interact and restrict one another. The composite system, consisting of the two individual systems, plays an active role in urban development. The most desirable outcome for each individual system is the coordinated development of the composite system.

2 Modeling Coordination Degree of the Composite System

The basic theory of coordination degree emanates from Synergistic theory. Synergistic theory was proposed by H. Haken in 1971. Based on this theory, a coordinated system moves toward order or structure when it is in the critical state. Coordination degree refers to the amount of harmony between the systems, or the system elements, in the development process. The composite system is dependent on the harmonious coexistence between its various subsystem components, internal self-organization, and external management activities in order to achieve complete system harmony.

2.1 Relational Evaluation of the Composite Subsystem

Most composite systems are unstable due to changes in elements of the system. Whether or not long-term interaction and linkages in the elements of the subsystem exist, it is assumed that the subsystems exist in a stable relationship. Co-integration analysis is used to evaluate the long-term relationships between subsystems and assess the essential factors of the composite system. Due to the higher reliability of the multivariable Vector Auto-Regressive (VAR) models over single variable models, and the instability of the factors in the composite system, this analysis chooses the multivariable VAR system for identifying and analyzing the relationship between the elements.

2.2 Coordination Modeling

If the factors exist in long-term equilibrium, the two subsystems can be used as inputs and outputs to each other. For example, when the transportation subsystem is used as input, land-use will be used as output and vice versa. Using the method of data envelopment analysis (DEA) and the idea of membership degree of fuzzy mathematics, an evaluation model of the composite system was developed. Steady state and dynamic coordination degrees show the different temporal concepts.

Transportation and land-use systems are the two subsystems analyzed. The composite system, *S*, can be expressed as:

$$S = f(T, L)$$

where *f* is the composite function; *T* is the transportation system, and *L* is the land-use system. If there are *n* decision making units (DMU), they are written by DMU_i (i = 1 ,2 ,…,n) where there are *m* inputs and *k* outputs. Assuming the input vector is X = (x1 , x2 , …, x*m* $)^T$, and the output vector is Y= (y1 , y2 , …, y*k* $)^T$, then the input vector represents the transportation or land-use subsystem, and correspondingly, the output represents the land-use or transportation subsystem. For each decision making unit, DMU_j has the corresponding efficiency evaluation index (where *u* and *v* are the weight values):

$$h_j = \frac{u^T y_i}{v^T x_j} = \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{mn} v_i x_{ij}}, j = 1, 2, \ldots, n \tag{1}$$

Generally speaking for DMU_{j0}, the grater the value of h_{j0}, the less input that is required to achieve more output from DMU_{j0}. By evaluating DMU_{j0}, the maximum value of h_{j0} as allowed by the varying weights (u,v) can be determined. Once the efficiency evaluation index has been determined for DMU_{j0}, the Charnes-Cooper-Rhodes (CCR) model can be developed:

$$\begin{cases} \max h_{jo} = \dfrac{\sum_{r=1}^{s} u_r y_{rjo}}{\sum_{i=1}^{m} v_i x_{ijo}} \\ s.t. \dfrac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1, j = 1, 2, \ldots n \\ u \geq 0, v \geq 0 \end{cases} \tag{2}$$

This model can be solved by the Charnes-Cooper transformation which transforms fractional programming to a linear system.

Based on the above theory, the transportation-land use composite coordination degree model is constructed. Subsystem T serves as the input and subsystem L becomes the output. Using the CCR model, the input surplus of subsystem T and the output deficit of subsystem L are determined. The scale value Σλ, and DEA effectiveness value (written as α) are also determined.

Using the same methodology, subsystem L is used as the input and subsystem T becomes output. Using the CCR model, the input surplus of subsystem L, and output deficit of subsystem T, and scale value Σλ, and DEA effectiveness value (written as β) are once again determined.

Using the fuzzy mathematics idea of membership degree, the coordination degree model of the composite system is proposed. The membership function is defined as $\mu(\theta)=\theta$, where θ is DEA effectiveness value. The steady state coordination degree of T to L is defined as μ_T, and $\mu_T=\alpha$; the steady state coordination degree of L to T is μ_T, and $\mu_T=\beta$; the coordination degree between two subsystems is μ_{TL} where:

$$\mu_{TL} = \frac{\min(\mu_T, \mu_L)}{\max(\mu_T, \mu_L)} \tag{3}$$

Equation (3) shows, the more closely μ_T and μ_L are related the higher the coordination degree of the composite system. When $\mu_T=\mu_L$, the coordination is at its highest. The coordinated development of the two subsystems is an interactive process, and the degree is a function of time. The coordination degree at time t is defined as $\mu(\theta, t)$. The greater the value of $\mu(\theta, t)$ the higher the coordination of composite system. If $\mu(\theta, t) = 1$, the system is completely coordinated; if q1$\leq \mu(\theta,t) < 1$, the system is coordinated; if q2 $\leq \mu(\theta, t) <$ q1, the system is essentially coordinated; if q3 $\leq \mu(\theta, t) <$ q2, the system is essentially uncoordinated; and if $\mu(\theta, t) <$q3, the system is complete uncoordinated. Based on existing research, the threshold values have been identified as q1=0.8, q2=0.6, q3=0.3.

The coordination degree is then defined and calculated. Assuming $\mu(\theta_0, t_0)$, $\mu(\theta_1, t_1)$ to $\mu(\theta_n, t_n)$ is the composite system coordination degree at every moment from $[t_0, t_n]$, where t_0 is the initial moment. The dynamic coordination degree is:

$$\mu(t_0,t_n)=\mu\left(\theta_0,\cdots,\theta_n;t_0,\cdots t_n\right)=\frac{1}{t_n-t_0+1}\sum_{i=0}^{n}\mu\left(\theta_i,t_i\right) \tag{4}$$

If t_{n1} and t_{n2} are two arbitrary moments, and $t_{n1} < t_{n2}$, then if $\mu(t_0, t_{n1}) \le \mu(t_0, t_{n2})$, it is said the two subsystems are gradually becoming coordinated, and if $\mu(t_0, t_{n1}) > \mu(t_0, t_{n2})$, it is said the two subsystems are gradually becoming uncoordinated.

3 Selection of Indices

It is essential to the analysis to select appropriate indices. Selection of indices follows two basic principles: scientific principle and practical principle.

The transportation subsystem indexes selected for this analysis are: transportation area; rail length; public transportation utilization in the peak hour; the number of vehicles and the average network saturation.

The land-use subsystem indexes selected for this analysis are: land area; population density and land use. The land use indices include a diversity index, dominance index and uniformity index.

4 Application of the Model

Beijing is utilized as the case study for this analysis. The study area includes 1088 square kilometers and encompasses 8 districts. Index values for the study area are presented as Table 1. The results of the model calculations are presented as Table 2 and Table 3, and the coordination degree values are presented as Table 4.

As shown in Table 4 in 1994 and 2003, the steady state coordination degrees in Beijing are calculated as 0.552 and 0.526. This shows that the transportation-land use composite system were essentially uncoordinated. In 1994, the population primarily lived in the center of city and land uses were primarily residential and industrial. The land reserved for transportation was inadequate. In 2003, the land use dedicated to roadways increased rapidly. However, congestion became more severe, and land-use became chaotic resulting in a coordination degree that is even lower than 1994. The model results coincide with observed data. The model shows that after implementation of proposed improvements to the transportation and land use systems, the coordination degree will improve to 0.929 by 2020.

5 Conclusions

A coordination model of the transportation-land use composite system based on DEA and fuzzy mathematics was developed. A coordination degree index system was also developed. The theoretical models were applied to the land use and transportation systems in Beijing. Results of the model concurred with observed data, demonstrating that land use and transportation systems in Beijing have developed in an uncoordinated manner since 1994. The model also showed that proposed improvements to the systems will yield coordinated development by the year 2020.

Table 1 Land-use and Transportation Indexes in Beijing

System\Indexes		Year 1994	Year 2003	Year 2020
Land-use Subsystem	**Land Area（km^2）**			
	Commerce	8	18	24
	Industry	141	60	33
	Transportation	28	104	192
	Green	37	59	123
	Residential	486	161	143
	Population density			
	Entire City	625	928	850
	Center Region	589	650	540
	Inner City	172	147	90
	Land use layout			
	Diversity Index	1.35	1.43	1.58
	Dominance Index	0.73	0.65	0.50
	Uniformity Index	0.65	0.69	0.76
Transportation Subsystem	Transportation Area (km^2)	28	104	192
	Rail Length (km^2)	54	114	570
	Public Transportation Utilization in Peak Hour (%)	28	29	50
	Vehicle Number	66	200	500
	Network Average Saturation	*0.86/#0.67	0.9	-

Note: * is the value inside the 2nd ring road, and # is the value outside of the 2nd ring road.

Table 2 DEA Results (Transportation as Input and Land-use as Output)

DMU		1994	2003	2020
Input Surplus	s^-_1	0	7.54	0
	s^-_2	0	0	0
	s^-_3	3.98	0	0
	s^-_4	0	9.56	0.75
Output Deficit	s^+_1	0	0	0
	s^+_2	8.79	0	4.21
	s^+_3	2.1/0/3.67	3.5/4.32/1.09	0/0/0
$\Sigma\lambda$		0.854	1.218	1.004
DEA Effectiveness Value		0.596	0.453	0.964

Table 3 DEA Results (Land-use as Input and Transportation as Output)

DMU		1994	2003	2020
Input Surplus	s^{+}_{1}	13.22	0	4.70
	s^{+}_{2}	9.36	4.38	0
	s^{+}_{3}	1.02/24.21/7.66	0/0/0	1.32/0/0.88
Output Deficit	s^{-}_{1}	3.57	2.78	0
	s^{-}_{2}	0.98	0	0
	s^{-}_{3}	1.22	0	4.29
	s^{-}_{4}	1.40	7.85	0
$\Sigma\lambda$		4.765	1.221	1.745
DEA Effectiveness Value		0.329	0.861	0.896

Table 4 Result of Coordination Degree Analysis

Year	1994	2003	2020
Steady State Coordination Degree	0.552	0.526	0.929
Dynamic Coordination Degree	0.552	0.539	0.669

References

[1]Englerf, Gran Gercwj. Co-integration and error correction: representation, estimation and testing. [J]. Econometrica, 1987, 55 :251-276.

[2]Fan Hua, Tao Xue-yu. Model of Composite System Coordination Degree and It's Application. [J] Journal of China University of Mining & Technology, Vol. 35 No. 4, Jul. 2006:515-520

[4] Johansens, Juselies K. Maximum likelihood estimation and inference on co-integration with application to the demand for money. [J]. Oxford of Bulletin Economics and Statistics, 1990, 52 :169-210.

[5]Meng Qing-song, Han Wen-xiu. Study of the coordination measurement model with respect to composite system. [J] Journal Of Tianjin University. Vol l. 33, No. 4, Jul 2000:444-446

[6] Miller E., Kriger D.S., Hunt J.D. Integrated urban models for simulation of transit and land-use policies. [C]. Final Report for World Bank Project H12, TCRP Web Document 9, 1998

[7] Mills D. The timing of urban residential land development. Research in Urban Economics. 1983[3]: 37-57

[8] Mills ES. An aggregative model of resource allocation in a metropolitan area. American Economic Reviews. 1967[57]:197-210

[9] Ou Xiong , Feng Chang-chun ,Shen Qing-yun. Application of Synergistic Model in Urban Land-Use Potential Appraisal. [J] Geography and Geo-Information Science. Vol. 23 No. 1, January 2007:42-45

[10] Zhao Yan-feng, Chen Yan-yan, Luo Ming. A Coordination Degree Model for Land Use and Urban Transportation. [J], Road Traffic and Safety. Vol. 6 No. 4, April, 2006:31-34

[11] Zhao Yan-feng. Study of land use and urban transportation harmonious development. Master degree paper, Beijing University of Technology.2006

The Motorist as a Customer: Making Customer Satisfaction Important in Operating a Major Tollway

Panos Prevedouros[1], Bill Halkias,[2] Dimitrios Mandalozis,[2] Anna Politou[2]
([1]) University of Hawaii at Manoa, 2540 Dole Street, 383, Honolulu, HI 96822, USA
([2]) Attikes Diadromes S.A., 41.9 Km of Attiki Odos, GR 19002 Peania, Athens, Greece

ABSTRACT
The collection of feedback and adjustments to operations and services in response to motorists' needs and perceptions is presented. Surveys for commercial customers, manual toll payers and electronic toll payers are described, along with main results. New products are toll subscription packages that respond to varied needs and frequencies of travel. Services range from the location where electronic transponders are sold to response times and types of aid offered for incidents on the freeway.

THE ATTICA TOLLWAY

The 65.2 centerline km long Attica Tollway was developed between 1997 and 2004 and was the first urban freeway in Greece. It operates under a flat-fee, open tolling system with manual and electronic lanes (*ePass*). The freeway has exceeded demand forecasts and has produced significant improvements to traffic conditions in the region. Demand was forecast to be at the level of 240,000 vehicle entries in 2010, but in early 2007 entering traffic on weekdays often exceeded 300,000 vehicles per day.

The rapid design and development of Attica Tollway provides valuable know how in implementing a complex freeway in a densely populated metropolis. Attica Tollway also was the road transport backbone for the 2004 Olympic Games in Athens that took place in August 2004 since it provides connection between all the major transportation infrastructure of the Attica region such as other freeways, major arterials, ports, rail stations and the international airport.

MOTORIST SURVEYS

The primary concern of Attica Tollway management is to ensure the smooth and continuous operation of the freeway and provide high quality services to motorists. In order to establish the level of perceived quality and other elements of customer satisfaction, extensive road side surveys (RSS) were conducted.

The management of Attica Tollway realized the importance of motorist surveys from the start of operations. RSS were done in December 2002 at 11 interchanges, April 2003 at 17, February 2004 at 36, December 2004 at 38, June 2005 at 38, April 2006 at 37 and March 2007 at 38 interchanges. The main objectives of the RSS are to: (1) obtain an understanding of customers' travel characteristics such as origin-destination, time of day and purpose of trips and their travel needs; (2) receive information for adjusting the existing or designing new products and services that better respond to user needs; (3) establish a communication channel with the customer; and, (4) collect rating to evaluate freeway performance and services offered.

The survey was done at toll plazas and its target population was the users of freeway using the manual toll lanes. A telephone survey targeted the electronic toll subscribers who cannot be interviewed at the toll stations. These are called the electronic toll collection, or ETC surveys. For both surveys, the sample selection was random. RSS questionnaires addressed to trucks and taxis were analyzed separately under the label commercial. The number of interviews is proportional to the number of transactions served at each toll station as well as proportional to the number of transactions per hour. The interviews were carried from Tuesday to Thursday, between 7:00 AM and 7:00 PM.

UNDERSTANDING CUSTOMERS' TRAVEL CHARACTERISTICS

The main results of the panel surveys are summarized in Table 1. The majority of the questions aim to understand customer's travel characteristics. The relevant results may be summarized as follows. Sample sizes average 8.384, 760 and 1,483 for RSS, ETC and commercial surveys, respectively. About 82% of surveyed motorists are in cars and similar small vehicles. A significant increase is observed in the share of trucks: In 2002 the share of trucks was 5.9% and in 2007 it was12.2%. Vehicle occupancy fluctuates within a 5% range. It is lowest for cash paying motorists with 70% using the car solo and 23% with 2 or 3 occupants. For ETC motorists the respective averages are 67% and 24%, and for commercial vehicles including taxis the percentages are 64% and 27%.

In the RSS, the average percentage of male drivers is 81% and only 19% are female. Female drivers are much higher at 25% in the ETC and commercial samples. Apparently driving is an area where female equality has not been realized in Greece. No significant changes have been observed throughout the years regarding driver age. Around 78% of the users are over 35 years old.

Based on the RSS, during the period 2002-2007 there is a gradual reduction in the percentage of trips which are business or commuting related from 79% to 66%. There is an increase in the percentage of leisure or personal trips, from 21% to 34%. Evidently, Attica Tollway attracts an increasing number of trips which are not business related and it facilitates a wide variety of trip purposes in the region. A similar trend is observed for electronic lane users. In 2003, about 65% of the rips were commuting and in 2007 commuting share reduced to 39.1%. All trips have stable distributions in the three years, 2005-2007 when commercial customer surveys were taken.

In the RSS, the frequency of trips is displaying an increasing trend for low frequency trips. Less than one trip per week increased from 17.3% in 2002 to 24.1 % in 2007 and one trip per week increased by 3.7% for the same period (from 7.9% to 11.6%). On the other hand, high frequency trips decreased significantly. Trips with frequency of 5 times per week decreased from 40.1% in 2002 to 29.1% in 2007. Similarly trips with frequency 6 times per week decreased from 9.4% to 4.5% for the same period, a decrease of over 50%. This large decrease of the high frequency trips is explained by the increase of the number of subscribers, given that Attica Tollway offers attractive subscription programs for the frequent users. In other words, high frequency users switched from cash to electronic payment.

On average, 45% of ETC motorists have a 5 or 6 trip weekly frequency, whereas 36% of RSS motorists have a 5 or 6 trip weekly frequency. A large increase in the share of subscribers using the tollway less than once per week is observed in the ETC surveys: 4.3% in 2003 and 20% in 2007. This increase is due to the promotion of electronic transponders even to users with low trip frequency.

Origins and destinations are major products of motorist surveys and aid in understanding the true movement of flows (and demand) on a freeway facility.

The surveys include detailed questions which were aggregated in six regions for the purposes of this presentation, as follows.

- Center of Athens is a densely populated urban area with medium incomes.
- Eastern suburbs (Mesogeia region between Airport and Sounion in Fig. 1) are a rapidly developing area with a large mix of activities from agriculture to medium-heavy industry and incomes ranging the entire gamut.
- Northern suburbs primarily contain a stable and high income population.
- Western suburbs by Elefsina (Fig. 1) are blue collar, lower income areas.
- Piraeus is the largest port of Greece and one the largest passenger ports in Europe; it has a stable, high density population with a wide range of incomes that are typically higher closer to the sea.
- South suburbs are older, affluent areas with medium to high incomes.

The vast majority of the trips (more than 86%) are home-based (HB). In 2003, HB trips reached 86%, an increase was observed in 2005 where HB trips reached 93% and in 2007 the percentage returned to 86%.

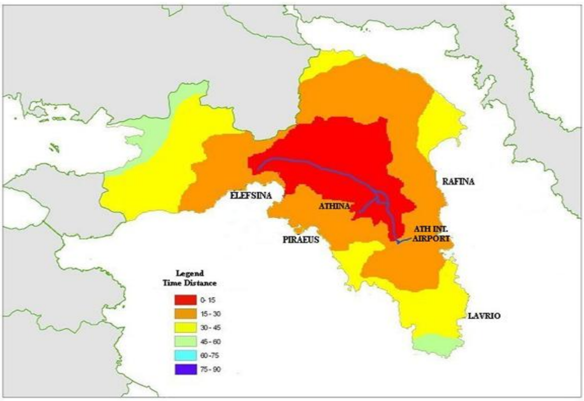


Figure 1. Isochronous contours around Attica Tollway in greater Athens area.

A substantial decrease is observed in the percentages of trip origins from the northern suburbs, from 44% in 2002 to 28% in 2007. Trips originating from western suburbs almost quadrupled from 5% in 2002 to 19% in 2007. Also from Mesogeia and Piraeus region trip shares grew from 7% to 17%, and from 6% to 13%. The decreases in shares mask that actual trips increased from these areas. However, actual trips increased much more from high growth areas. Overall, traffic demand is growing on Attica Tollway and the pace of growth has been about 8% per year in the last three years. In the first half of 2007, the average number of weekday vehicle entries through the toll plazas was about 325,000.

Regarding the destination of trips a decrease is observed to the percentages of northern suburbs (from 45% to 37%), southern suburbs (from 14% to 11%) and Piraeus (from 7% to 5%) between 2002 and 2007 while Mesogeia region and westerns suburbs indicate an 8% and a 2% increase, respectively. It is observed that in both surveys that were carried out (manual and electronic lanes) Mesogeia are represented by an increasing percentage of trips as an origin and a destination region. It was the construction of Attica Tollway that connected this region with the main Athens metropolitan area providing a huge capacity increase and substantial time savings. We look at time savings below. As the traffic volume on Attica tollway increase, travel times begin to elongate at peak times and, over time, the perceived time saving change. For example, 17% of the users perceived time savings of over 60 minutes in April 2007, whereas the percentage for this group in February 2004 went down to 5%. The group with perceived time savings between 31 to 60 minutes was 18% in April 2007, but in February 2004 the same group was 37%. The dominant group has perceived time savings between 16 and 30 minutes and this held constant around 50% for all years.

The ETC customer results show that in general between 2003 and 2005 there is a reduction the perceived time savings. An increase was observed only for time savings below 15 minutes where in 2003, 25.5% of the users perceived that they have time saving until 15 minutes and in 2005 this percentage was 46.5%. In all other groups of perceived time savings reductions were observed. The greatest reduction was observed to the perceived time savings between 45 and 60 minutes where in 2002 this percentage was 9.9% and in 2007 was only 1.6%. The weighted average perceived travel time savings are 27, 21 and 23 minutes for RSS, ETC and commercial motorists, respectively.

DESIGNING NEW PRODUCTS

Several questions were developed to receive information for adjusting the existing or designing new products and services that better respond to user needs. These questions relate to the preferred time of day of motorists (which can be used to create programs in order to switch some trips off the peak periods), who pays the toll, toll increases, awareness of the electronic tolls and the various transponder subscription programs, and evaluation of freeway performance and services in an attempt to gauge potential weaknesses and respond with solutions.

For the last two years (2005-2007) a small reduction in the portion of ETC subscribers entering the tollway during the morning peak period (7:00-10:00 AM) has taken place, from 31.4% in 2005 to 28.6% in 2007. This reduction is mainly affected by the increasing volumes in the morning peak and the ability of non-commute or business users to switch theirs trips to other times. Nearly 40% use the motorway during the extended afternoon peak between 1:00 and 7:00 PM.

The vast majority pays the toll on their own (83%), and no substantial variation in the percentage was observed for the six years covered by the surveys. However, a 17% share of payment by companies is substantial. The introduction of more attractive subscriptions for companies is an active goal of Attica Tollway.

What if the toll is higher? This is a critical question with significant implications to both revenue and level of service. Two scenarios were offered: (a) a small increase of about 10%, and (b) a larger increase of about 20%. No trends or conclusions can be drawn for the answers because there were different toll rates for different years (four different toll rates affecting seven surveys). Interestingly, in the latest survey, over than 50% of the users intend to use Attica Tollway with the same frequency even if the toll price is increased by 20%. However, in April 2006 the percentage was 33%. In general users are more flexible to possible toll increase because (a) it keeps pace with the general cost of living increases and therefore the costs of operating the tollway, (b) motorist service is outstanding, (c) travel times are short and generally superior to the parallel arterial street network during busy parts of the day. With many Greeks working more than one job, the willingness to pay in order to save time is relatively high.

Questions on motorist awareness of *ePass* existence were relevant up to the end of 2004. Since 2003 substantial campaigns were carried out during last years. This was in response to early findings that a large majority of motorists were not aware of *ePass* which is an electronic transaction system on dedicated electronic toll collection, ETC, lanes. The large increase of both the level of awareness and of subscriptions showed the success of the annual campaigns

Subscribers to *ePass* programs are treated to various discounts that are not available for cash payers. The question asked therefore was … *What would be the frequency of the ePass usage if there was no discount offered by the subscription program you are subscribed to?* About 54% of the users surveyed in 2006 would have used Attica Tollway with the same frequency even without a discount; this increased to 55% in 2007. About 32% of the users would have used Attica Tollway less and around 12% would have stopped using the facility completely. Experience in similar questions has shown that the answers are not corresponding to the actual intentions of the users when the increases do take place.

The last section of the survey focuses on an evaluation of freeway performance and services offered. A major issue in the opening years were the signage along the tollway which is the only urban freeway in Greece. Motorists had some difficulty in navigating to their intended paths in the first few years but the opinion about traffic signs along Attica Tollway improved between years 2003 and 2005. In 2003 80% of the users were either satisfied or very satisfied, and in 2005 this percentage increased to 93%.

Around 97% of the users are either very satisfied or satisfied from the customer service provided by the Attica Tollway. A part of this remarkably high percentage is attributable to the high level of safety provided by the facility. In 2003, the percentage of high and very high satisfaction was 86% and in 2007 the percentage reached 96%. Attica Tollway received the 2005 first prize for road safety of the International Road Federation (IRF) for its safety measures and for having among the lowest crash rates among European freeways. Fatalities per million kilometers of travel on Attica Tollway were 0.6 for 2004 and 2006, and 0.5 for 2005, or about six times lower than the rate on other freeways in Greece.

A gradual reduction in the percentage of the users who are not aware of the existence of a telephone customer's service is observed. Between 2005 and 2007, a 5% reduction took place. Only half of the users who are aware of the existence of the telephone service have used it. Also there is increase of the share of users satisfied by the level of service at toll stations, from 95.7% in 2006 to 98.3% in 2007. There are draft plans for expanding Attica Tollway to reach other regions in the metropolitan area. About 94% believe that the construction of the extension of Attica Tollway is important.

CLOSING REMARKS

A large number of cities may benefit by looking into Attica Tollway. Seventy five metro areas in the world have a population between 3 and 6 million people. Of these, 29 have freeway facilities and 46 do not. Selected cities without freeways and with population and density similar to that of Athens include Alexandria (Egypt), Algiers (Algeria), Amman (Jordan), Ankara (Turkey), Baghdad (Iraq), Busan (South Korea), Caracas (Venezuela), Chengtu (China), Kabul (Afghanistan), Khartoum (Sudan), Kiev (Ukraine), Nanking (China), Riyadh (Saudi Arabia), Santiago (Chile) and St. Petersburg (Russia). Attica Tollway provides valuable know-how in implementing a complex freeway in a densely and fully populated metropolis, and in specialized arenas such as public-private partnerships, concession financing, build-operate-transfer projects, and integrated electronic tolling and intelligent transportation systems.

Table 1: Surveys and Main Results

	Date	Dec 2002	Apr 2003		Feb 2004	Dec 2004	Jun 2005			Apr 2006			Mar 2007		
Motorist survey questions	**Type**	RSS	RSS	ETC	RSS	RSS	RSS	ETC	Commercial	RSS	ETC	Commercial	RSS	ETC	Commercial
	Sample size	8010	9043	550	11263	11014	6670	500	1136	6340	991	1599	6333	1000	
	Willingness to answer	94%	92%	87%	93%	90%	93%	85%	93%	94%	72%	94%	96%	65%	
Vehicle type	car, van	82%	81%	-	84%	84%	-	-	-	-	-	-	78%	-	
	bus, truck	6%	7%	-	7%	7%	-	-	-	-	-	-	12%	-	
Vehicle occupancy	1	73%	74%	73%	70%	69%	66%	65%	61%	68%	65%	60%	72%	66%	71%
	2 or 3	21%	20%	19%	23%	24%	25%	25%	27%	25%	27%	30%	22%	25%	25%
Gender of driver	male	81%	80%	77%	82%	81%	81%	77%	77%	79%	74%	71%	85%	74%	78%
	female	19%	20%	23%	18%	19%	19%	23%	23%	21%	26%	29%	15%	27%	22%
Trip Purpose	commute or business	79%	77%	88%	73%	67%	69%	73%	70%	60%	73%	66%	66%	67%	69%
	leisure, other	21%	23%	12%	27%	33%	31%	27%	30%	40%	26%	32%	34%	33%	32%
Weekly frequency of subject trip	<1	17%	15%	4%	19%	24%	17%	12%	-	25%	17%	-	24%	20%	-
	1	8%	9%	5%	9%	13%	10%	13%	-	12%	12%	-	12%	12%	-
	5 times	40%	34%	53%	33%	26%	27%	38%	-	25%	32%	-	29%	29%	-
	6 times	9%	5%	15%	4%	4%	5%	5%	-	4%	6%	-	5%	4%	-
Origin	North Athens	44%	37%	46%	31%	33%	32%	39%	-	28%	43%	-	28%	42%	-
	South Athens	10%	12%	13%	6%	8%	8%	13%	-	8%	11%	-	8%	11%	-
	Airport	5%	4%	4%	4%	4%	4%	2%	-	3%	3%	-	4%	3%	-
	Mesogeia	7%	12%	16%	17%	18%	18%	22%	-	17%	22%	-	17%	22%	-
	West suburbs	5%	5%	7%	6%	5%	5%	10%	-	8%	10%	-	19%	9%	-
	Piraeus	6%	7%	7%	5%	5%	4%	5%	-	8%	6%	-	13%	4%	-
Destination	North Athens	41%	37%	45%	31%	28%	26%	37%	-	30%	38%	-	31%	37%	-
	South Athens	9%	11%	14%	8%	7%	9%	13%	-	7%	11%	-	8%	11%	-
	Airport	7%	6%	4%	4%	6%	5%	4%	-	7%	3%	-	8%	4%	-
	Mesogeia	7%	12%	16%	18%	19%	19%	22%	-	17%	20%	-	17%	22%	-
	West suburbs	11%	8%	7%	15%	15%	15%	10%	-	16%	11%	-	17%	9%	-
	Piraeus	8%	7%	7%	5%	4%	3%	5%	-	3%	6%	-	5%	5%	-
Time savings in minutes	>60	-	3%	0%	5%	3%	4%	0%	4%	2%	-	3%	17%	-	5%
	31-60	-	26%	24%	37%	28%	37%	8%	32%	18%	-	14%	18%	-	16%
	16-30	-	48%	50%	42%	52%	44%	46%	42%	59%	-	44%	47%	-	55%
	<15	-	15%	26%	8%	8%	7%	47%	6%	13%	-	11%	9%	-	18%

Analysis of Parking Fee Effect on Travel Behavior in a Downtown District

Wei Li[1], Xiuyuan Zhang[1] and C. S. Papacostas[2]

Abstract: This paper reports the results of a stated-preference survey of mode choice under different parking fees in downtown Beijing during the morning peak period. The respondents' preferences are estimated by means of a Multinomial Logit Model. The main findings indicate that drivers are sensitive to parking fees, and that, assuming low parking fees outside the central city, the major effect of a high parking fee downtown is the shifting to parking outside the downtown area and riding transit to internal destinations (P&R.) A high degree of travel demand elasticity is found and a specific parking fee policy for Beijing is recommended.

Keywords: Stated preference; Parking fees; Travel behavior; Price effects

1. Introduction

The downtown areas of major cities are typically congested. Experience in China and other countries indicates that parking demand control can be an efficient tool for restricting traffic flow in and out of a downtown district. Outside China, research on the effect of parking fees in reducing automobile usage began in the 1960s. Many economists hypothesized that zone-based parking pricing by geographic district was practical, but little information is currently available to systematically apply this theory to Chinese cities.

Two categories of factors affect people's choice of travel mode: Individual traveler attributes such as gender, age, income, car ownership and so on; and external factors such travel time, traffic policies, fare structure, quality of transit service, etc.

This paper is concerned with the results of a stated-preference (SP) survey that was designed to investigate the potential effects of changing the parking fee in downtown Beijing on travel mode choice. Based on the survey, a Multinomial Logit (MNL) model of mode choice was estimated to investigate commuter responses to changes in parking fees.

The stated-preference survey

Alternative set and presentation

The SP survey was aimed at cars arriving downtown during the morning peak period between 7:30 and 9:30 a.m. As elsewhere, parking fees constitute at least 20% of the total travel cost of these trips (Gila and Mahalel, 2006). For this

[1] School of Traffic and Transport, Beijing Jiaotong University, Beijing, China, 100044

[2] (Ed.), Dept. of Civil & Environmental Engineering, University of Hawaii, Honolulu, Hawaii, 96922

reason, we considered four levels of downtown parking fees: ￥8/h, ￥10/h, ￥15/h, ￥20/h. At the time of the survey $1=￥7.4.

The survey offered three travel mode alternatives:

Alternative 1: Continue commuting by car.

Alternative 2: Shift to public transport.

Alternative 3: park outside the district and ride public transport into the city.

We supposed that the low parking prices outside the downtown district would not change and, since our approach was based on the postulate that people choose alternatives indirectly on the basis of their attributes (Tversky and Kahneman, 1981), we described each alternative in the choice set by combinations of its main characteristics, specifically:

Parking fees in the downtown district: characterized by four levels.

Travel time, characterized by three levels: High, Medium, Low.

Travel cost, characterized by three levels: High, Medium, Low.

We classified the levels of those variables according to actual trips from various places of residence and then presented the corresponding levels of travel time and fare to the respondents. If all levels of the three attributes were included, the factorial design would entail four parking fee levels, three travel time levels and three travel cost levels, for 36 combinations. However, the three combinations shown in Table 1 were sufficient to describe the relevant modal attributes.

Table1. Combinations of attribute values for each alternative

Alternatives	Combinations of attribute levels		
	Parking fee	Travel cost	Travel time
1 (passenger car)	￥8（10/15/20）/h	High	Low
2 (public transport)	0	Low	High
3 (P&R)	￥2 /h	Medium	Medium

The alternatives were presented to the respondents as pair wise choices. Accordingly, for each value of the parking fee, three pairs were presented: Alternatives 1 and 2, Alternatives 1 and 3, and Alternatives 2 and 3. Thus, the four parking fee levels would result in 12 choice pairs. However, it is reasonable to assume that respondents who rejected a given parking fee level, would also reject all higher levels as well. Consequently, the questionnaires were shown to the respondents according to the increasing parking fee levels to avoid redundant data. An example of the questionnaire cards used is shown in Table 2 (Gila and Mahalel, 2006).

Table 2: Pair of alternative 1 and 2

1. To use a car and pay the parking fee	2. To use public transport
Parking price in downtown district: ￥8 /h	Parking price in downtown district: 0
Travel cost: 8×parking duration + fuel price×travel time	Travel cost: ￥1-3
Travel time: 30-40minutes	Travel time: delayed 30%-50%*
Which alternative would you prefer?	

*Compared with travel time by car.

Model calibration

The utility functions for alternatives 1, 2, and 3 were specified as follows (Guan, 2004):

$$V_{1n} = \beta_1 + \theta_2 carT + \theta_3 fuel\_fare + \theta_4 parking\_fee$$
$$V_{2n} = \beta_2 + \theta_1 fare + \theta_2 busT$$
$$V_{3n} = \theta_1 (P\,\&\,R) f + \theta_2 (P\,\&\,R) t$$

Where V_{1n} is the utility associated with using a passenger car; V_{2n} is the utility of public transport; V_{3n} is the utility of the park-and-ride option described above; fare = public transport fare; (P&R)f = public transport fares plus parking fees outside the central district; carT is travel time by car; busT is travel time using public transport, including waiting time; fuel fare = fuel price×travel time by car; parking fee = parking price in downtown district ×parking duration; β_1, β_2 are the mode-specific constants for alternatives 1 and 2, respectively; and θ_1 , θ_2, θ_3, θ_4 are model coefficients.

The respondents' preferences towards the alternatives were estimated by means of a Multinomial Logit Model, where the utility of alternative *i* to individual *n* is

$$U_{in} = V_{in} + e_{in}$$

Where U_{in} is the utility of alternative *i* to individual *n*; V_{in} is a linear function of the alternative's observed characteristics, and e_{in} is its random component. Accordingly, $P_n(i)$, the probability that an individual *n* will choose alternative *i* from a given set *C* of available alternatives is specified as

$$P_n(i) = \frac{e^{v_{in}}}{\sum_{i \in C} e^{v_{in}}}$$

Survey Results

Sample Characteristics

A sample size of 60 is considered appropriate for a stated preference surveys (Pearmain et al., 1991). In our survey, we obtained a sample of 572 valid responses out of 1039, representing sample characteristics as follows

(1) The percentage of respondents whose travel distance exceeded 20 kilometers was 18.28%; between 10 and 20 kilometers was 37.63%, and less than 10 kilometers was 44.09%.

(2) Based on the travel time distribution in the survey, we defined a travel time less than 30 minutes to be "low" (59.86%), between 30 and 60 minutes to be "medium" (32.38%), and longer than 60 minutes to be "high" (8.1%).

(3) The travel cost by automobile was considered equal to the fuel cost plus the parking fees. The survey showed that 52.61% of the respondents had a travel cost less than ￥15, 43.60% between ￥15 and ￥25, and only 3.79% higher than ￥25. Typical one-way public transport fares range from ￥1 to ￥3. Because of the comparatively high differences in costs, shifts between modes should be expected.

Model calibration results

We calibrated the model using the TransCad software package Tables 3 and 4 show the estimated coefficients and the resulting statistics.

Alternatives	AEC1	AEC2	fare	time	oilfare	parkingfee
1	ONE			carT	oilfare	parkingfee
2		ONE	fare	busT		
3			[P+Rf]	[P+Rt]		
Model	-0.744758	0.595019	0.026326	-0.997536	-0.039492	-0.007580

Table 3: Model specification and coefficients

Maximum likelihood reached at iteration 8

Parameter	Estimate	Std. Err.	t Test
AEC1	-0.744758	0.316524	-2.352929
AEC2	0.595019	0.251724	2.363774
fare	0.026326	0.017061	1.543041
time	-0.997536	0.317137	-3.145441
oilfare	-0.039492	0.062484	-0.632038
parkingfee	-0.007580	0.002376	3.190573

Table 4: Model coefficients, standard errors, and t-test results

Since $|t_k| > 1.96$ indicates significance at the 5% level, we eliminated fare and oil fare from the included attributes, and re-estimated the model coefficients. This resulted in a more accurate model with β_1=-1.11, θ_2=-0.89, θ_4=-0.01, $|t_k| > 1.96$, and ρ^2=0.544, $\overline{\rho}^2$=0.480.

To examine the nature of the model, we applied the SPSS software to forecast the market shares of each alternative if the parking price were set at ￥15/h (Table 5) As shown, 43.5% of the respondents would prefer to pay the higher parking fee and to continue to commute by car, 32.2% would park outside the downtown district and 24.3% would shift to public transport. The percentage of correctly "predicted" commuters by car reached 96.9%, and the corresponding percentages for P&R and public transport were above than 60%. Inclusion of other important determinants of mode choice, such as comfort and convenience, can result in an improved model.

Observed	Predicted			
	Car	Public transport	P&R	Percent Correct
Car	189	0	6	96.9%
Public transport	26	95	48	66.2%
P&R	34	44	130	62.5%
Overall Percentage	43.5%	24.3%	32.2%	75.1%

Table 5: Observed vs. Predicted mode shares

Table 6 shows the probabilities of choosing each of the three mode alternatives at different parking prices, as estimated via the SPSS software. At ￥8/h, shifts from the car to the other two modes would be very low, whereas at ￥20/h, only 5.6% of the commuters would persist in using their cars. It appears that ￥15/h may be a reasonable rate, with 43.5% of the commuters continuing to travel by car and paying the fee and 56.5% shifting to the other alternatives that avoid the fee.

Parking Fee	Predicted (%)		
	Car	Public transport	P&R
￥8/h	95.8	0.0	4.2
￥10/h	79.2	4.0	16.8
￥15/h	43.5	24.3	32.3
￥20/h	5.6	33.2	61.2

Table 6: Predicted mode shares

Demand elasticity

The demand elasticities at different parking price levels, which reflect the effects of changes in these fees, can be computed directly from the MNL models as follows (Gila and Mahalel, 2006):

$$E_{Xink} P_n(i) = \frac{dp_n(i)}{dx_{ink}} \frac{x_{ink}}{p_n(i)}$$

Where $P_n(i)$ is the probability that an individual n will choose alternative i, and x_{ink} is the relevant variable to alternative i.

The principal aim of our research is to determine the potential effect that parking fees have on commuting to the downtown district by car. The function relating auto travel demand and parking price can be described as follows:

$$E_{Xn}P_n(1) = \frac{dp_n(1)}{dx_n}\frac{x_n}{p_n(1)}$$

Where $P_n(1)$ is the probability that an individual n will choose alternative 1 into the downtown district and x_n is the highest parking price that the individual n is willing to pay. When the parking price is increased to￥8/h, the parking price elasticity of auto travel demand is -1.4. Therefore, a one-percent increase in the parking price will bring about a 1.4 percent decrease in the demand for travel by car.

Conclusion

This study shows that increasing the parking fees in downtown Beijing can affect the volume of automobile traffic entering and exiting the central district, and can change commuters' choice of travel mode. For example, at ￥8/h, the demand elasticity is -1.4, which means that a percent increase in parking price will bring a 1.4 percent decrease in the demand for traveling downtown by car. As a result, proper selection of parking prices in downtown areas can reduce travel congestion and can address to some degree the problem of parking facility shortages in those areas.

Our estimated model showed that when the parking fee in Beijing is set at ￥15/h, the probabilities of choosing the three alternatives were balanced. If it is set lower than ￥15/h, commuting in and out of the central district by car will be stimulated, resulting in relatively higher traffic congestion levels. If it is set at higher levels, it may restrain traffic flows excessively, which may lead to wasting resources and negative economic impacts.

References

Gila, A., and Mahalel, D. (2006). "Congestion tolls and parking fees: A comparison of the potential effect on travel behavior." *Transport Policy* 13, 496-502.

Guan, H. (2004). *Disaggregate Model-A tool used to analyze travel behavior.* Peoples Transportation Publishing Company (in chinese).

Pearmain, D., Swanson, J., Kroes, E., Bradiey, M. (1991). *Stated Preference Techniques: A Guide to Practice*. Steere Davies Gleave and Hague Consulting Group, London.

Tversky, A., and Kahneman, D. (1981). "The framing of decisions and the psychology of choice." *Science* 211, 453-458.

Performance Measures – A Tool to Enhance Transportation Decision-Making

By

Mr. Peter Lai, Vice President, PB International - Hong Kong

Ms. Holly Krambeck, Associate, PB Americas, Strategic Consulting Services

ABSTRACT

Making comprehensive urban transportation policies and decisions under the dynamic and complex environment of the Chinese institutional structure for planning is, to put it lightly, not a simple task. In addition to the traditional conflicting interests inherent in any transport project, there is an additional layer of challenge posed by Chinese planning approvals and incentive structures. Given this environment, the need for an intuitive, simple system to quickly evaluate project alternatives and to readily articulate the performance of an existing system is acute. Project sponsors need to be able to readily answer questions such as:

- Which is a better alternative to serve the transportation goals and needs of the project? (evaluate alternatives within project)
- What is a better choice between modal options to serve the city or communities? (new highway via public transport line)
- What are the best transportation options for the city in the near term and in the long term? (master planning)
- What are the transportation priorities under various levels of financial constraints? (capital programming)
- How do transportation policies impact the quality of life of the people who are living in the study area? (environmental and community impact assessment)

Establishment of standardized transportation "performance measures" could be an effective method to overcome these challenges. This paper will share the concept of and experiences in developing performance measures through examples.

Introduction

The availability of standard transportation system performance measures would help Chinese decision-makers: identify clear priorities for new transport projects and policies, understand how different options would hinder/promote their ability to meet goals, and help them quickly evaluate the performance of an existing system and identify its most critical needs, based on pre-determined priorities.

Other countries, such as the UK and the US, already have highly developed national and regional systems for performance measurement, with measures tied directly to funding and policy-making decisions. This paper focuses on particularly successful examples of these systems and provides some discussion as to how lessons learned from these systems may be applicable to the Chinese context.

Regional Performance Measures

To be effective, performance measures must be directly tied to transportation goals set forth by the leaders and critical stakeholders in the study area, since performance measures are used to determine whether the goals have been met. To illustrate, the North Jersey Transportation Planning Authority (NJTPA), a metropolitan planning organization that oversees a multi-billion dollar transportation system in the US, has established a set of regional transportation goals in its Regional Transportation Plan (RTP), where measures of accessibility, mobility and congestion have been extensively used to gauge the effectiveness of the transportation systems. They are:

- Accessibility – Can travel destinations be reached?
- Reliability – Can the transportation system always be counted on?
- Sustainability – Can the system maintain performance over time?
- Intermodality – Are different modes well-connected?
- Highway Mobility – Can roads be traveled without delay?
- Public transport Mobility -- Can public transport be used without delay or crowding?
- Walk/Bike Mobility – Are walking and bicycling effective modes of travel?
- Freight Mobility – Can freight be moved without delay?

To determine how well these goals and objectives are achieved by the transport system, the following set of performance measures were developed and implemented by the Authority:

1. Accessibility

Accessibility was considered for users of both private vehicles and public transport. Four performance measures were combined in equal proportions to create the Accessibility Index for measuring the degree of difficulty in reaching travel destinations.

- Number of jobs reachable within 40 minutes by highway.
- Labor force (population) reachable within 40 minutes by highway.
- Population plus jobs reachable within 40 minutes by public transport.
- Jobs reachable within 75 minutes by public transport.

2. Reliability

The most troubling form of congestion for travelers can be delays – traffic incidents, construction diversions, train or bus cancellations, and weather-related events – that can occur without warning. Reliable, predictable travel times are fundamentally important in both passenger and freight movements. Two measures were applied to create a Reliability Index:

- Non-recurring delay on roadways as compared to total delay (75% weight).
- “Mobility redundancy”, the percent of population that can reach high-level public transport services (rail, express bus, long-haul ferry) within 15 minutes (25% weight).

3. Sustainability

One of NJTPA's primary objectives is to create a plan where transportation and quality-of-life are maintained for future generations. Given the expense and difficulty in building roadway facilities, rapid growth in vehicle travel may contribute to environmental pollution and more crowded roads. One representative performance measure was used to create a Sustainability Index – vehicles miles traveled per capita.

The Sustainability Index is defined as measuring the efficient management of demand that must be supported by the roadway system, considering the amount of vehicle miles traveled per person. The maximum Sustainability Index value of 100 translates into very low vehicle travel on highways per person, as is usually the case in dense urban areas. Lower values of the Sustainability Index indicate more vehicle miles per capita due to more trips, longer trips, or less use of shared modes.

4. Intermodality

An efficient/well-coordinated transportation system depends on effective connections between modes, which provide options for travelers and make a variety of travel methods feasible. Nearly all public transport trips begin with another mode – on foot, by car, from another public transport line or by bicycle. Fixed route public transport, for example, depends highly on the quality of such connectivity. The Intermodality Index consists of three measures, the first weighted at 60%, the others at 20% each:

- Non-Single-Occupancy-Vehicle mode shares by district (60% weight)
- Ratio of bus/walk/bike access to public transport vs. auto access to public transport (20% weight)
- Percent of households within one mile of public transport services (20%).

5. Highway Mobility

Mobility on highways depends on a well-connected road network and on consistent traffic conditions. Congestion and unpredictable travel times hinder travel, and frustrate travelers. The principal costs of congestion are the delays experienced or time "wasted" in travel. The Highway Mobility Index contains two measures:

- Highway "delay ratio" – the number of jobs reachable within 40 minutes during peak travel periods compared to those reachable under freely flowing conditions. (25% weight)
- Percent of person-hours of travel occurring delayed below "acceptable" speeds. (75% weight)

6. Public Transport Mobility

Delays and uncertainty can present serious impediments to travel on public transport. As traffic congestion can affect bus movements and train congestion can occur on rail lines. The Authority adopted three factors to generate a Public Transport Mobility Index. They are:

- The number of jobs reachable within 60 minutes by public transport compared to the number reachable within 40 minutes by highway. (70% weight)
- Households having one-seat rail service to major regional destinations. (15%)

- Public transport crowding, as gauged by the number of passenger-miles at critical "over capacity" system locations. (15% weight)

7. Walk/Bike Mobility

Walking and bicycling were recognized as healthy, inexpensive, and community- and environmentally-friendly modes of travel for shorter trips. Walking and bicycling require facilities such as sidewalks, shoulders and bikeways. In addition, "calm" traffic conditions must be present, and destinations must be close enough to reach under human power. The index is formed by two equally weighted measures:

- Bicycle Compatibility – a gauge of roadway suitability for bicycle travel, considering average traffic volume and speed, the presence of shoulders, curb lane width, and residential land uses.
- Pedestrian Compatibility – a predictable analogue to the Pedestrian Potential *Index previously applied at the NJTPA reflecting synergies among land uses,* the availability of public transport service, the presence of pedestrian facilities, and street network densities.

8. Freight Mobility

Effective goods movement depends on many factors such as the extent of highway congestion experienced by trucks, right-of-way congestion experienced by rail providers, inter-modal transfer effectiveness, and efficient landside access to port and airport facilities. The Freight Mobility Index components are:

- Truck miles of travel near intermodal-freight facilities. (40% weight)
- Miles traveled by trucks under congested road conditions (on roads filled above 80% of capacity) compared to uncongested travel. (60% weight)

These performance measures or indices were quantified and forecasted and then compared to actual performance data to determine the regional transportation needs of the northern New Jersey region. The same sets of measures were also used to quantify the transportation benefits that can be generated different transportation strategies or programs. These measures illustrate how a set of performance measures need not be overly numerous or complex to evaluate a system – the important takeaway is that the measures should be tied to specific goals – measures not tied to critical priorities are not necessary in this context -- and are sufficiently designed to ensure reasonable measurement.

Congestion-Related Performance Measures

Given the rapid pace of motorization in Chinese cities, congestion is quickly becoming a critical issue. Establishing performance measures would help planners understand which measures and policies would be most effective, *given the other competing priorities of the transport system.* The following example shows how performance measures for congestion were developed in the US.

The US National Cooperate Highway Research Program (NCHRP) Report 398 reported that it is difficult to conceive of a single value that will describe all of the traveler's concerns about congestion, but there are four components that interact in a congested road network. These components are defined as follows:

- Duration – the amount of time congestion affects the travel system.
- Extent – the number of people or vehicles affected by congestion, and by the geographic distribution of congestion.
- Intensity – the severity of the congestion that affects travel. The level and total amount of congestion on the transportation systems.
- Reliability – a variation of the above three elements. Recurrent congestion is relatively stable and predictable, but non-recurrent (incidents, weather, etc.) delay causes greater variation in the amount of congestion and is most difficult to predict. Reliability is the impact of non-recurrent congestion on the transportation system.

The majority of performance measures developed for congestion can be summarized into five general categories that include: (1) level of service, (2) speed, (3) volume to capacity (v/c) ratio, (4) travel delay, and (5) travel time. There are many pros and cons associated with these categories – thus, in the development of these measures, a separate set of criteria were developed to determine which would be most effective.

- Representative of Congestion – the ability to truly reflect and differentiate various levels of congestion
- Comprehensible to the Public – the ability to be understood by non-technical general public, and the ease for the public to relate the measure(s) to the real world congestion
- Adaptability to the Region – the ability to represent a system-wide or area-wide average
- Predictability – the ability to predict and quantify future year performance
- Technical Adaptability -- how easily the measure can be produced through analytical tools/models, household surveys, and traffic volumes counts.

To determine which performance measures would be most suitable for evaluating congestion, a team of researchers funded by the US Department of Transportation tested each along these five criteria. Table 1 below summarizes the congestion performance measure assessment results.

Table 1. Performance Measure Assessment Matrix by Category

	Performance Criteria					**Technical Applicability**		
	Representative to Congestion	**Comprehensible to Public**	**Adaptability to Region-wide**	**Future Year Predictability**		**Analytical Tools / Models**	**Household Surveys**	**Traffic Volume Counts**
Performance Measure:								
Level of Service	Fair	Fair	Good	Good		Good	Poor	Fair
Speed	Fair	Good	Fair	Poor		Good	Poor	Poor
Volume-Capacity (V/C) Ratio	Fair	Poor	Good	Good		Good	Poor	Fair
Travel Delay	Good	Good	Fair	Fair		Fair	Poor	Poor
Travel Time	Good	Good	Fair	Poor		Fair	Poor	Poor

The "Level of service" and "Delay" are found to be most suitable performance measures under the five evaluation criteria; and analytical models are the better tools for these two performance measure categories than HH surveys and traffic counts.

To develop performance measures for China, a similar evaluation procedure would be undertaken to determine which measures would be most suitable, given local priorities, conditions, and technical capacity.

Assessment of Potential Performance Measures

To assemble a comprehensive list of potential performance measures, two areas of focus should be identified: measures that can be quantified from existing sources (especially by existing travel demand models) and measures that should be explored for future applications. The table below highlights some of the typical examples:

Measures available from existing sources

- Vehicle-Mile Traveled (VMT) by LOS by Facility Type
- Number of Link by LOS by Facility Type
- Lane-Mile by LOS by Facility Type
- Lane-Mile Hours by LOS by Facility Type
- VMT by LOS by Trip Purpose
- Recurring Vehicle-Delay by Delay Type by Facility Type
- Recurring Person-Delay by Delay Type by Facility Type
- Total Vehicle Time by Source of Delay by Facility Type
- Vehicle-Hour Traveled (VHT) by LOS by Facility Type
- Number of Hours at Estimated LOS
- Peak Period Vehicle Unacceptable Delay (in Vehicle Hours)

Measures to be explored for future use

- Percentage of employee within X minutes
- Percentage of employee within Y miles
- Excess delay (person hours of travel)
- Delay ratio (delay time/acceptable travel time)
- VMT per capita
- Person-mile-traveled (PMT) at high v/c
- Incident delay
- Incident rates
- Incident response times
- Pedestrian/Bicycle mode share
- Pedestrian potential index
- Bicycle compatibility index
- Delay experienced by trucks
- Freight rail capacity
- Tons of freight moved
- Transit ridership / capacity

The following criteria will be used to determine how well these measures will serve the overall technical objectives:

- Ability to reflect such factors as system-wide congestion, facility-based congestion, peaking duration and characteristics and trip purposes;
- Ability to consider demand management, transit and supply-side options
- Ability to monitor and verify results

Conclusion

The US experiences suggests that there is no single performance measure – e.g., cost or travel time savings – that can enable leaders to fully evaluate project alternatives and system needs. A well-defined set of transportation performance measures is a valuable tool that facilitates thoughtful short and long term planning, according to a region's most critical priorities. The development of this decision-supporting tool, however, can be a time consuming process – thus, lessons learned from countries and agencies that have already derived and tested measures are invaluable.

Economic Contribution Rate Analysis of Chinese Transportation Infrastructure Construction：Based on the Input-output Data Sheet in 2002

WU. Aihua[1], Mohamed Eisa[2], LI. Fengxia[3], and LIU. Xinhua[4]

[134] School of Management, Ludong University, Yantai, P.C. China., E-mail: hiwuaihua@yahoo.com.cn; fengxiali7866@yahoo.com.cn; xinhuaL@ytnc.edu.cn.
[2] Department of Civil Engineering, University of Sherbrooke, Sherbrooke, Quebec, Canada. mohamed.abdel.allah.nabil@USherbrooke.ca

ABSTRACT

Investments on the construction of transportation infrastructure have a great effect on the national economy of any country. Infrastructure includes the transport networks such as railways, highways, and waterways. This article examines the transportation infrastructure construction contribution on both of the national product growth and the personnel employment growth in China for the year of 2002. This examination was done based on input-output method and industry correlation theory which applies the output, investment, employment multiplication in order to quantify the investments effects on the transportation infrastructure construction for the Chinese economy. The obtained results showed that the scale of transportation infrastructure construction is small at present in comparison with many others countries which will lead to lower the growth rates and productivity of the Chinese economy in the near future. This paper concludes that the Chinese government should strength the administration of macro-economy in order to exert the function of the transportation infrastructure for the economy development.

CE Database Subject Headings: Transportation; Infrastructure; Economy; Construction;
GDP; China

1 Introduction

Analyzing the contribution of transportation industry to the national economy is one of the most important topics in the transport economic. The mainstream methods which are used generally to this research topic here in China are Keynes's multiplication theory and input-output method. The input-output model is used internationally to quantify the effects of economic investments on the construction of transportation infrastructure. In this paper the input-output data sheet in 2002 was integrated by the multiplication theory to input-output method. Moreover, the economic effects of the transportation on the gross domestic product (GDP) and the labor employment were considered.

The contribution to the GDP will not only reflect its increasing which consider as direct economic benefits, but also will manifest the increasing values of other economic sectors through the correlation between the different industries which consider as indirect economic benefits. Output multiplier, GDP multiplier and investment multiplier considered as the reflectors for the GDP, which means the more investments in the transportation construction the more of the total output and gross domestic product. The benefits arising from the output multiplier and investment multiplier are named output benefits and investment benefits respectively. Moreover, increasing the investments in the transportation infrastructure construction will contribute directly and indirectly the economy. For instance it will increase the employment sector by adding new jobs. In the same way, employment multiplier is an indicator reflecting the contribution of the employment; its benefits will be named with employment benefits. In sum, the total contribution to the GDP will include output benefits and investment benefits which will be expressed through direct and indirect investment benefits as well as employment benefits.

2 Analysis Method and Indicators for the Economic Contribution

2.1 Input-output table and the main coefficients

The input-output table can be named with department balance table or industry correlation table. It is based on the balance between input and output of various sectors of the national economy. Moreover, it can be used to reveal the quantitative relation of economic and technological interdependence and mutual restraint in various sectors, which means:

$$X = AX + Y \text{ or } X = (1\text{-}A)^{-1}\, Y, \qquad (1)$$

Where: $X= (X_i)_{nx1}$, X_i is GDP of sector i; $A = (a_{ij})_{nxn}$ is called a direct consumption coefficient matrix and $Y= (Y_i)_{nx1}$ is the final rank vector.

Before analyzing the data using the input-output table, it is important to calculate its coefficients. In this study the most basic coefficients are the direct consumption, the total consumption and Leontief inverse coefficient. Moreover, there are some coefficients will be included in the analysis such as the added value coefficient, the employment coefficient, and the influence coefficient. The added coefficient is the added value for the GDP generated by the unit output value for some sectors. The added coefficient value for sector (j) is noted by g_j, where $g_j = G_j / X_j$, and, G_j is the added output value for the sector j. In the same way, the employment coefficient is the employment amount needed by some sectors in case of adding unit value to the output values. In case of sector (j) the employment coefficient is noted by l_j, where $l_j = L_j / X_j$, and L_j is the employment amount needed by the sector j.

In this case the influence coefficient will be the indicator key for measuring the economic interdependence and mutual restrain between different sectors. Moreover, when some sectors increased with a unit value, the influence coefficient will reflect the needed influence degree created by various sectors of the national economy. The formula of influence coefficient is:

$$F_j = \sum_{i=1}^{n} \alpha_{ij} \Bigg/ \left(\left(\sum_{j=1}^{n} \sum_{i=1}^{n} \alpha_{ij} \right) / n \right) \qquad (2)$$

Where α_{ij} is Leontief inverse coefficient, $\sum_{i=1}^{n} \alpha_{ij}$ is the sum of the rank j in the Leontief inverse matrix $(I-A)^{-1} = (\alpha_{ij})$ and $\left(\sum_{j=1}^{n} \sum_{i=1}^{n} \alpha_{ij} \right) / n$ is the average sum of the ranks in the Leontief inverse matrix.

It can be noted from the previous equations that the influence degree created by the output sector (*j)* on the other sectors will be higher than the average level of social influence if the influence coefficient (F_j) is greater than unity. However, both of the influence degree and the average level of social influence will be the same if the influence coefficient (F_j) equal to 1. This in turn shows the tight relation between the influence coefficient (*Fj*) and the output sector (*j)*

2.2 Multipliers of the economic contribution
Direct and indirect economic benefits can be used as well as employment benefits to obtain the total outputs, investments and employment benefits from the transport construction investments. Assuming that the investment amount will be equal to unity, the output investments and employment multipliers can be obtained respectively. Moreover, when the transport industry increases one unit, the added total output, GDP, and employment amount will be increased for all other sectors.

The output multiplier (H_o) can be defined as the added total output for all other sectors when the transport industry increases one final demand unit.

$$H_o = DE_o + BE_o = \sum_{i=1}^{n} \alpha_{ij} \qquad (3)$$

However, the investment multiplier (H_{GDP}*)* or the *GDP* multiplier can be defined as the added GDP for all other sectors when the transport industry increases one final demand unit.

$$H_{GDP} = DE_G + BE_G = g_i \sum_{i=1}^{n} \alpha_{ij} \qquad (4)$$

In the same way, the employment multiplier (H_L) can be defined as the amount of the created new jobs for all other sectors when the transport industry increases one final demand unit.

$$H_L = l_i \sum_{i=1}^{n} \alpha_{ij} \qquad (5)$$

3 Quantitative Analysis of the Economic Contribution

Based on the input-output data sheet in 2002 formatted in table contains 122 sectors, combining with a refining for the transport industry, 15 × 15 sectors input-output table has been got as shown in table 1. The direct and the total consumption coefficient of railway, highway and waterway are shown in this table. Moreover, it can be noticed from this table that increasing the railway total output by only one unit will change the values of direct consumption of product by railway; highway and waterway to be 0.4086, 0.446 and 0.5983 units respectively. In the same way, when the railways itself increasing every one unit in the total output value. The values of GDP, highway and waterway will change to be 0.5537, 0.5914 and 0.4017 units respectively.

Tabel.1 Direct, indirect and total consumption coefficients of input-output railway,
highway and waterway in China.

	Railway			Highway			Waterway		
	Direct	Indirect	Total	Direct	Indirect	Total	Direct	Indirect	Total
1	0.0001	0.0494	0.0495	0.0002	0.0462	0.0464	0.0002	0.0639	0.0641
2	0.0148	0.0478	0.0625	0.0042	0.0409	0.0451	0.0012	0.0565	0.0577
3	0.1981	0.3887	0.5869	0.2149	0.3824	0.5973	0.3124	0.5496	0.8621
4	0.0412	0.0309	0.0721	0.0082	0.0285	0.0366	0.0030	0.0380	0.0410
5	0.0491	0.0051	0.0542	0.0036	0.0058	0.0094	0.0024	0.0062	0.0085
6	0.0228	0.0072	0.0300	0.0102	0.0073	0.0174	0.0052	0.0098	0.0151
7	0.0012	0.0131	0.0143	0.0649	1.0167	1.0816	0.1206	0.0322	0.1528
8	0.0036	0.0123	0.0160	0.0054	0.0115	0.0169	0.0590	1.0191	1.0782
9	0.0011	0.0033	0.0044	0.0021	0.0043	0.0064	0.0034	0.0053	0.0087
10	0.0169	0.0160	0.0329	0.0124	0.0170	0.0294	0.0137	0.0211	0.0348
11	0.0241	0.0478	0.0720	0.0328	0.0489	0.0817	0.0204	0.0661	0.0865
12	0.0212	0.0211	0.0424	0.0682	0.0286	0.0969	0.0486	0.0397	0.0883
13	0.0002	0.0055	0.0057	0.0013	0.0082	0.0095	0.0005	0.0085	0.0090
14	0.0000	0.0002	0.0002	0.0002	0.0002	0.0004	0.0001	0.0002	0.0003
15	0.0142	0.0267	0.0408	0.0177	0.0284	0.0461	0.0076	0.0363	0.0439
16	0.4086	0.6752	1.0838	0.4463	1.6749	2.1212	0.5983	1.9526	2.5509

Note: Each number indicates an industry sector. 1. Agriculture; 2. Mining industry; 3. Manufacturing; 4. Production and supply of electric& energy, gas and water; 5. Construction; 6. Railway transportation; 7. Highway transportation; 8. Waterway transportation; 9. Other transportation; 10. Warehousing, post and communications; 11. Wholesale and retail, trade and food lodging industry; 12. Finance and insurance industry; 13. Real Estate; 14. Tourism15. Other industries; 16.Total intermediate inputs.

According to the total consumption coefficient in Table 1, the output multiplier of the railway will be 2.0838 which mean if the railway construction adds one final product unit that will bring more than two units for the production needs from various sectors in the national economy. In the same way, the output multiplier of the highway is 2.1212, and the output multiplier of waterway is 2.5509. Moreover, the investment multiplier of the railway can be calculated to be 1.2324 from the added coefficient values and the total consumption coefficient which means that billion investments in the railway will create 1.2324 billion GDP values. The investment multiplier of the highway is 1.1745; and the investment multiplier of the waterway is 1.0247. According to the employment coefficient and the total consumption coefficient, the employment multiplier of the railway can be calculated to be 1506, which means per billion railway investments, 1506 jobs will be created. In the same way 487 and 164 jobs will be created in both of highway and waterway sectors respectively.

In addition, the influence coefficient can be calculated for the railway, the highway and the waterway sectors to be 0.9268, 0.9434 and 1.1346 respectively. Which in turn shows their deeply effect on other sectors specially the waterway effect that gives higher values than the average values of the social effect. The investments in the railroad sector have been increase with 3730 million Yuan from the year 2005 to 2006 as shown in Tables 2 and 3. This increasing will keep the employment post to be 561730 in parallel with creating GDP more than 4600 million Yuan and draws the GDP to grow 0.2%.

Tabel.2 Amount of the investments in the transportation and the GDP of China

Item	Investment (hundred millions of Yuan)				Increase rate of GDP (%)
	Railway	Highway	Waterway	GDP	
2004	516	4702	408	159878	10.1
2005	889	5495	689	183085	10.2
Increase	373	793	281	23207	——

Table.3 Results of the economic contribution of the transportation construction in 2005

Item	Railway	Highway	Waterway
Simple investment multiplier	1.2324	1.1745	1.0247
Total investment benefits induced by the increased investments (a hundred million Yuan)	460	931	288
Created GDP by the investment (a hundred million Yuan)	1096	6454	706
Rate of the increased economy caused by the investments (%)	0.2020	0.4094	0.1266
Contribution rate to the GDP by investments (%)	0.5984	3.5252	0.3856
Simple output multiplier	2.0838	2.1212	2.5509
Total output benefits induced by the increased investment (a hundred million Yuan)	777	1682	717
Direct output benefits (a hundred million Yuan)	373	793	281
Indirect output benefits (a hundred million Yuan)	404	889	436
Accumulative investment multiplier	3.0058	2.8647	2.4994
Total investment benefits induced by the increased investments (a hundred million Yuan)	1121	2272	702
Created GDP by the investments (a hundred million Yuan)	2672	15742	1722
Added value coefficient	0.5914	0.5537	0.4017
Direct investment benefits (a hundred million Yuan)	221	439	113
Indirect investment benefits (a hundred million Yuan)	239	492	175

Induced investment benefits (a hundred million Yuan)	661	1340	414
Rate of the increased economy caused by the investments (%)	0.4928	0.9985	0.3087
Contribution rate to the GDP by investments (%)	1.4595	8.5980	0.9406
Accumulative output multiplier	5.0824	5.1736	6.2217
Total output benefits induced by the increased investments (a hundred million Yuan)	1896	4103	1748
Direct output benefits (a hundred million Yuan)	373	793	281
Indirect output benefits (a hundred million Yuan)	404	889	436
Induced output benefits (a hundred million Yuan)	1118	2421	1032
Employment multiplier	1506	487	164
Total employment benefits induced by the increased investment (person)	561730	385950	46209
Employment coefficient	723	229	64
Direct employment benefits (person)	269574	181951	18115
Indirect employment benefits (person)	292156	203998	28094
Influence coefficient	0.9268	0.9434	1.1346

4 Conclusions

Input-output models are applied to quantify the effects of the investments in the railway, the highway and the waterway transport infrastructure construction on the national economy. The effects include direct, indirect, and induced economic benefits to an area. This area can be identified in terms of multipliers of output, gross domestic product and employment throughout as the affected area. The direct rate of the transport infrastructure contribution to the national economy for the developed countries is about 7-8%, however; in China this rate is much lower than this value. Most of the work done regarding the mutual mechanism between the transportation and the economy indicates that the function of the transportation infrastructure construction is based on the developed harmony between the economy, the society, the surrounding environment, and the transportation infrastructure. That must suit the increased demand for economy development as well as transportation service in different scales. Transportations have been developed quickly in recent years in China, however; in general it is still small in comparison with other countries and not enough for the Chinese growth economy. The government should strengthen the administration of macro-economy in order to exert the use of the transportation infrastructure in economic development.

5 References

[1] Banister D. and Berechman J., 2000 “Transport investment and economic development” UCL Press.

[2] L. ,Nan and Z., Qingming, 2006, “Quantitative Analysis of Effects of Transportation Infrastructure Investment on National Economy” (in Chinese), Journal of Highway and Transportation Research and Development, edition 5, pp 150-154.

[3] W., Yuan-qing, Z., Zhi-min, and Z., Wei., 2004 “Arithmetic research of investment multiplicator in highway construction investment (in Chinese)” China Journal of Highway and Transport, edition 17(2) pp105-108.

The Yangtze River Transportation-Economic Belt

CHEN Jun[1], *CHEN Chuan*[2]
Edited by G. Murray
[1] School of Traffic and Transportation Engineering, Tongji University, Shanghai, China, PH(+86-021)6958-4928;email:chenjuntom@126.com
[2] School of Traffic and Transportation Engineering, Tongji University, Shanghai, China, email: chenchuan_tj@hotmail.com

Abstract

The Yangtze River, at 6300 km long, is the longest river in China and a key transportation corridor in China's booming economy. The implementation of the *National Freeway Planning* and *National Long-term Railway Planning* programs will create a comprehensive transportation corridor along the river, consisting of railway, roads, waterway, aviation and pipeline infrastructure. The development of this transportation infrastructure will be the impetus for continued development within the Yangtze River economic region, further fuelling growth in the Chinese economy.

This paper applies the concept of the "transportation-economic belt"; i.e. - that construction of a comprehensive transportation corridor creates the impetus for formation of a "belt", or linear region, of high-growth economic activity around the corridor. Using the comparison of one of the world's first comprehensive transportation corridors – the Rhine River - the inherent relationship between industrial development and transportation is examined. Using this historical perspective and the transportation-economic belt concept, the paper suggests mitigation measures that could be adopted to manage growth, and to address the possible negative effects of rapid growth and industrialization.

Key words: Yangtze River; transportation corridor; transportation-economic belt; Rhine River

Introduction

At 6300 km (3900 miles), the Yangtze River is the longest river in China and the third longest in the world. The Yangtze River basin includes seven provinces and two major cities comprising a dense industrial zone which plays an important role in the Chinese economy. The region accounts for 18% of the country's land mass, 37% of the population and 41% of the nation's gross domestic product.

The planned transportation system within the Yangtze River region consists of a network of railways, roads and waterways. Some of these facilities are in existence, but many are planned or currently under construction. When completed, they will form a comprehensive transportation corridor along the river (see Figure 1). The focus of this paper is to apply the concept of the "Transportation-Economic Belt" to analyze how the transportation corridor promotes the formation of an economic "belt" or linear region. Using the example of the Rhine River corridor in Europe, the relationship between the development of the transportation corridor and industrial development within the corridor was examined at various stages. The results of this examination were used to assess the possible strategies for managing development and mitigating the negative effects of growth within the Yangtze economic belt.

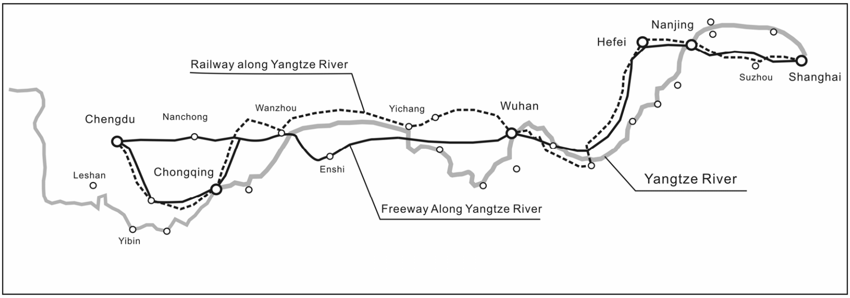


Fig.1 Layout of the Comprehensive Transportation Corridor of Yangtze River

Existing Conditions and Planned Transportation System Improvements

Waterways

Because of the Yangtze River's natural advantages as a transportation corridor, it is the busiest shipping waterway in China. It traverses much of China in an east-west direction and its tributaries and extensive canal network provide good "feeder routes" to the north and south. For centuries, the region to the south of the river has been the economic heart of China. The navigable mileage accounts for 53% of China's total navigable waterways, and it carries 80% of the country's shipping volume. In 2004, the Yangtze waterway carried 20% of the freight volume within the Yangtze River basin/region. The section between Nanjing and Shanghai is 437 km (272 miles) long and can accommodate up to 50,000 ton ships. Typical waterway traffic consists of fleets of kiloton barges. Currently, the (upstream) Wuhan to Nanjing section can accommodate only 5000 kiloton barges because of limited headroom on the bridges over the river(Ministry of Communication of China, 2005).

In 2003, the Chinese Ministry of Communication approved a report entitled "Development Planning of the Trunk Navigation Course on the Yangtze River". The report called for improvements that would increase the shipping capacity by 2010 to 150% of current (2003) levels, in order to keep pace with economic growth and shipping demand.

Railways

The main trunk railway in the Yangtze River corridor begins in Chongqing and ends in the seaport of Shanghai – a total length of 2024 km (1258 miles). Along the way, it passes through Hubei, Jiangxi, Anhui and Jiangsu. Only the section between Yinchuan and Wenzhou is not yet completed, and is scheduled to be opened in 2009.

Feeder lines and spur lines that feed into the trunk railway are continually under construction as part of a program of improving the rail network within the Region.

Roads and Freeways

The National Freeway also runs along the river corridor from Shanghai to Chongqing. With the planned completion of the Yichang-to-Enshi section in 2009, the entire freeway will be complete.

Concept of the Transportation-Economic Belt

Overview

The transportation-economic belt refers to the linear geographic region that includes integrated industries, population, resources, cities and towns; as well as the flow of goods, passengers and information within the corridor. The three primary elements of the belt are the trunk transportation facilities (waterway, rail and freeway); the industries and services; and the large population/industrial centers. The transportation trunks form the spine of the belt; large centers are the "growth poles", and the industrial clusters are the drivers of growth.

Growth and Evolution Pattern of the Transportation-Economic Belt

The creation of the transportation-economic belt starts with the old economic order. The initial construction of transportation facilities and introduction of new manufacturing technologies breaks the old order and provides the connections between industrial centers that allow the clustering of compatible production elements and diffusion of goods, people and information throughout the region, thereby increasing the overall economic efficiency of the region. Goods, resources and information are allowed to move freely and rapidly from where they are to where they are required. Economic "poles" develop in major centers that have favorable conditions for particular economic activities. As the economic poles develop stronger economies, they start to change the economic equilibrium of the surrounding areas by transferring the new technologies and, in turn, upsetting the old economies of the surrounding area as they become more-and-more linked to the major industrial poles. At this point, the entire belt experiences expansive growth which creates the need for additional transportation facilities. The completion of the transportation links and the diffusion of new technologies and supporting industries creates a new equilibrium, completing the transition from the old economy to the new economy (Zhang W.C. et al., 2002).

Case Study – The Rhine River Transportation-Economic Belt

Overview

At 2000 km (1243 miles) long, the Rhine River is the third longest river in Europe and carries the largest freight volumes of any inland waterway in the world. The Rhine Valley was one of the first areas in Europe to industrialize due to its resources of coal and iron, and is still continental Europe's primary industrial center. It is a classic example of the transportation-economic belt, complete with major industrial/population centers, a mature multi-modal transportation network and a diverse combination of manufacturing industries.

Stage One – Waterway Improvements (before 1850)

In order to improve shipping along the river and support the new coal-powered industries, dams, locks, canals and shoreline facilities were built and upgraded, improving the efficiency and capacity of the waterway as the spine of the transportation system.

Stage Two – Transportation Diversification and Rapid Industrial Growth (1850 to 1950)

In 1838, the first railway in the Ruhr District in Germany was officially opened and triggered the Rhine Valley rail construction boom, in order to move coal from the Ruhr Valley. By 1850, only 55% of the 1.6 million tons of coal was still transported by ship, the remainder being moved by rail. In spite of the growth in railroads, however, the Rhine River waterway continued to complement the railway network and provide an important transportation trunk due to its cost-efficiency.

Stage Three – Stable Development (1950 to present)

After 1950, West Germany entered a boom in highway construction within the Rhine corridor, while also continuing the reconstruction and in-filling of the railway network, improvements to inland shipping, and installation of oil and gas pipelines. This time was characterized by completion of a comprehensive multi-modal transportation network and maturing of the advanced industrial infrastructure. While rail is still the prime mover of bulk cargo today, the transportation-economic belt is supported by strong waterway shipping, highways, pipelines and aviation facilities (e.g. – Frankfurt is the largest cargo air-ferry center in Europe).

Industrial development within the Rhine River corridor was critically dependent on the construction and upgrading of the transportation network, which included main trunk facilities and the feeder canals, spur lines and roads. This interdependency of the transportation facilities with the region's economy formed the classic transportation-economic belt (Zhang W.C. et al., 2002).

Development of the Yangtze River Traffic Economic Belt

The ongoing development of the Yangtze River region's economy can be examined in light of this growth and evolution pattern, as epitomized by the Rhine River example. While the economy of the downstream section between Shanghai and Wanjiang is in a mature state, thanks largely to increases in shipping capacity in this section, and functions as an integrated economy having mostly completed the transition from old to new economic order (stage three); the middle and upstream sections of the region are still in transitional or formative stages (stage two), with only the major centers of Wuhan and Chongqing having made the transition to new economy. The completion of the transportation links will strengthen ties to the areas surrounding these major centers and complete the development of the Yangtze River transportation-economic belt.

Strategies for Continued Development of the Yangtze River Corridor

Due to its significant natural advantages, and its key location along the country's main east-west axis, the Yangtze River has the potential to be the largest economic belt surrounding an inland river in the world. While the downstream section has already reached a mature stage in the development of the transportation-economic belt, the middle and upstream sections are now in the formative stages.

The administrative districts along the Yangtze River corridor should strengthen intra-regional cooperation and initiate joint efforts to promote economic development within the belt by coordinating regional planning, and integrating their own resources and infrastructure into an overall comprehensive economic plan. This can be accomplished by improving and increasing the exchange of information, particularly on economic development issues that are relevant to the Yangtze River region(Tongji University, 2006).

The completion of transportation infrastructure is critical to integration of the economic belt and to the accommodation of the large volumes of freight and passengers in an efficient manner. Cities in the corridor must work together to agree on which projects have priority to support of the regional economy, and to bring these transportation projects to fruition. The three modes of transportation must be tightly integrated and developed in a coordinated manner. For instance, inter-modal connection/transfer points must be of high capacity and not create bottlenecks in the transportation network (Chen, Y.S. et al., 1999).

The Yangtze River shipping corridor is six times the length of the Rhine River corridor, however, the freight volume is currently only one-sixth of the Rhine's. While the completion of the Three Gorges project, the improvements to river ports, and construction of the Shanghai International Shipping Center provide much-needed capacity, the region's cities will need to focus on ways to improve the capacity, flows and efficiency of the transportation system along the corridor in order to protect the competitiveness of China's economy which is dependent on low price and large quantities (i.e. – efficient production and transportation). This is supported by policies and infrastructure that support low-cost shipping(Rong, C.H., 2002).

The development of the Rhine River corridor resulted in substantial destruction of the local environment. However, through the conservation efforts of the last few decades, much of this destruction has been reversed with the introduction of cleaner technologies and more responsible environmental practices. These technologies and practices must be incorporated into the construction, operation and maintenance of industries and infrastructure within the Yangtze River basin, and enforced by political jurisdictions within the Region. In this way, industrial development can continue on a sustainable basis.

Conclusion

With the rapid pace of transportation infrastructure construction along the Yangtze River corridor, the transportation system will be completed to the point where it effectively supports the economy. The example of the Rhine River and the concept of the transportation-economic belt provide a framework on which to build strategies and mitigation measures aimed at effectively completing the transportation network in a way that supports economic growth and protects the environment.

References

Chen, Y.S. et al. (1999). “The Necessity of Development of Comprehensive Transportation Corridor of Yangtze River Economic Belt”. China Highway Journal, 12 (1), 118-121.

Ministry of Communication of China (2005). “To Build Golden Waterway, to Promote Economic Development of Yangtze River”.

Rong, C.H. (2002). “Western Transport Economics” Economic Science Press, Beijing.

Tongji University (2006). “The Socio-economic Impacts of New Changes of Traffic Conditions to Yichang”.

Zhang, W.C. et al. (2002). “Traffic Economic Belt” Science Press, Beijing.

Cycling management: the success story of the city of Bolzano Bozen

Marcella Morandini
Ökoinstitut Südtirol/Alto Adige
e–mail: morandini@ecoistituto.it

ABSTRACT

Traffic jams, air pollution, lack of exercise – these are problems that a large majority of us face every day. Cycling is one of the best solutions to all of them: it is an active contribution to creating an urban environment with less traffic and better air quality, it helps people to improve their well-being and contributes to a better quality of life in urban areas.
The Municipality of Bolzano Bozen, located in South Tyrol (Italy), has been promoting bike mobility for several years. As an overall result, the modal split of bike mobility in this town rose from 17% to 25% in less than five years.
This paper gives an overview of the results achieved through several projects and activities implemented in collaboration with the Ökoinstitut Südtirol/Alto Adige.

BACKGROUND

Bicycles are particularly suitable as a means of transport for distances ranging between five and ten kilometres and therefore for small/medium towns. They are fast, silent, fun, sporty, economical to run and environmentally friendly. The environmental advantages are self-explanatory: bicycles require little space, are not noisy, do not use up energy resources and therefore do not pollute the environment.
There are also considerable economic advantages, not only for cyclists, but also for society in general. For instance, public administration offices, private companies and supermarkets offer costly parking facilities that use up a great deal of space, whereas bicycles require fewer and less costly infrastructure. Moreover, there are no ensuing health and environmental costs for the general public.
Starting from the 1980ies the administration of the city of Bolzano Bozen began to strongly favour bicycle mobility by creating a comprehensive network and a whole bike-system step by step.
The Municipality managed to increase the share of bicycle mobility from 17.5% (i.e. a yearly average of 59,740 trips per day in 2002) to 22,7% (2005, last survey) in the overall modal split. Recent estimates show that the current share of bike mobility in the city has already reached 25%.

BEGINNING OF A SUCCESS STORY

Cyclists should reach the destination of their transfers in the fastest and most direct way as possible, putting attention on safety. This is the only way cycling can become an attractive alternative. Figure 1. shows the network of bicycle lanes in Bolzano Bozen.

Figure 1. Bike lanes for everyday needs

TOWARDS A COHERENT BICYCLE MOBILITY SYSTEM

In order to set up to a long-term strategy for the development of bike mobility, in the year 2000 the Municipality commissioned the Ökoinstitut Südtirol/Alto Adige to elaborate a bike mobility plan.
The bike mobility plan is based on the survey and analysis of the demand for movement within the town. The analysis of the current supply is then compared with the outcomes of the demand analysis. This comparison allows to identify urgent short-term interventions and to plan strategies in the medium and long term. Along this, in order to enhance bike mobility in the city, information and marketing play a decisive role.

The four pillars of a coherent bike system hence are:

1. Demand analysis for bicycle mobility
2. Supply analysis of bicycle mobility
3. Information
4. Marketing

1. Demand analysis for bicycle mobility

In order to identify the demand for bicycle mobility, the general traffic flows through the city and the mobility choices made by the inhabitants of Bolzano Bozen are analysed. Two different studies are carried out following the KONTIV method (*Kontinuierliche Erhebung zum Verkehrsverhalten – Continuous Survey of Mobility Behaviour* [1,2]: the first one in 2002, the second one three years later, in 2005.
As a first step, such studies allow to sketch a picture of the mobility habits of the citizens. The results highlight <u>who is moving</u> through the city, <u>from where to where</u>, <u>when</u>, <u>how</u> and <u>why</u>. The 'mobility diary' of the inhabitants is designed in such a way that every trip is defined by the individual activity performed at each out-of-home destination. The basic idea behind is to obtain all information concerning all out-of-home activities performed on a specific date, not simply those which reflect the researcher's a priori views on what the 'formally correct' answers are.
The results of these studies represent an essential instrument for decision-making. They allows politicians, stakeholders and planners to elaborate strategies for the development of the bike network based on the real needs of the population, to

estimate the growth potential of bike mobility in town and to individuate the best target group on which to focus promotional activities.

First study: 2002

The first study was carried out in 2002 [3]. The following figures show the citizens' average movements among the different areas of Bolzano Bozen by all means of transport and by bicycle over the entire year of reference.

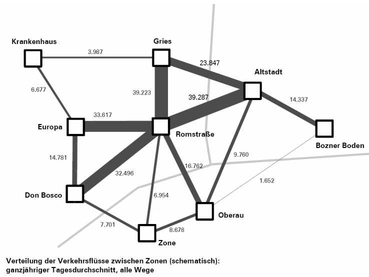


Figure 2. Distribution of traffic flows between city areas (daily average, <u>all means of transport</u>)

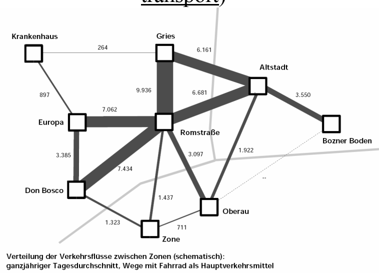


Figure 3. Distribution of traffic flows between city areas (daily average, <u>bicycle</u>)

Second study: 2005

The same study was repeated three years later, following the same method [4]. The general structure of traffic flows did not change dramatically: the main points of attraction in town had remained the same. Nevertheless, it is important to point out that the modal split had changed. The share of bike mobility increased more than that of all other means of transport, rising from a yearly average of 17.5% in 2002 to 22,7% in 2005. This result was achieved thanks to the measures implemented by the Municipality, which aimed at promoting bicycle mobility in town, both at the level of infrastructure and communication.

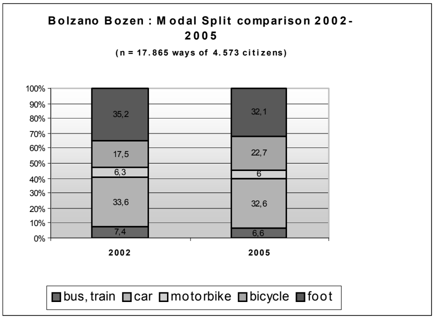


Figure 4. Comparison of the modal split on an annual average (2002-2005)

2. Supply analysis of bicycle mobility

The supply of bike mobility in the city must be based on the real demand, i.e. on the real needs of the citizens. In order to adapt the actual supply to the demand it is particularly important to set priorities, with particular attention to:

- points of attraction in town (points of interest) and origin-destination flows
- main axes on which to concentrate first investments
- non-expressed bike mobility: to be enhanced in the short term
- potential bike mobility: for the medium term

The network

Bike lanes should reflect the real traffic flows, forming a coherent network that allows cyclists to reach their destinations in the fastest and most direct way. The first step is guaranteeing a backbone of bike lanes and routes responding to the demand, following high-quality standard criteria.

Figure 5. Views of the bike network of the city of Bolzano Bozen

3. Communication and marketing

Along with the completion and extension of the bike lanes network a comprehensive and innovative communication and marketing strategy has been developed and it is in constant implementation.

Corporate identity

The communication system is based on a corporate identity, that connects all activities implemented for the promotion of bike mobility in the city. The logo of “Bici Bolzano – Fahrrad Bozen” is the founding element of it.

Figure 6. Logo “Bici Bolzano – Fahrrad Bozen”

In addition to the logo, a whole corporate design covering all aspects and all activities of bike mobility in Bolzano Bozen has been created.

Emotional marketing

The emotional appeal is strategic: promoting cycling in the city is often more a cultural than a technical matter. Therefore it is important to appeal to the emotions, to the sensations.

Bike mobility has to be perceived as a real mobility system, with the same dignity and importance as the other transportation systems.

In the city of Bolzano Bozen several marketing activities have been carried out, from poster and postcards actions to cinema spots.

Figure 7. Emotional marketing: big prints

Figure 8. Emotional marketing: postcards

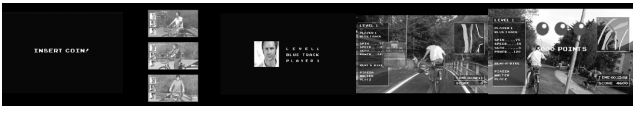
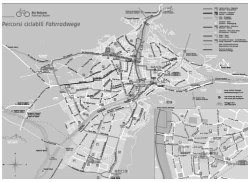


Figure 9. Emotional marketing: cinema spot as a “videogame”

Map in Z-Card format

A practical map in the Z-card system illustrates the whole bike network and its routes with information about related services and points of interest always with the typical corporate design. Routes are marked in different colours, almost as they were underground lines.



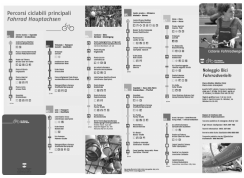


Figure 10. The map in Z-Card format

REFERENCES

1. Brög, W., Erl, E. (1999): Kenngrößen für Fußgänger- und Fahrradverkehr. Bericht der Bundesanstalt für Straßenwesen, Heft M 109.
2. Bundesamt für Raumentwicklung, Bundesamt für Statistik (2001): Mobilität in der Schweiz. Ergebnisse des Microzensus 2000 zum Verkehrsverhalten
3. Pörnbacher, H., Niederkofler H.P., (2002): Die Mobilität der Haushalte in Bozen Untersuchungen 2002. Illustrierter Ergebnisbericht zu einer empirischen Untersuchung, apollis, Bozen
4. Pörnbacher, H. (2005): Die Mobilität der Haushalte in Bozen Untersuchungen 2005. Illustrierter Ergebnisbericht zu einer empirischen Untersuchung, apollis, Bozen.

Can electric bikes provide sustainable transportation solutions to Chinese cities? An investigation of impacts on the transportation system

Christopher Cherry
Assistant Professor, Department of Civil and Environmental Engineering, University of Tennessee-Knoxville, 223 Perkins Hall, Knoxville, TN 37996-2010; PH 865-974-7710; FAX: 865-974-2669; e-mail: cherry@utk.edu

Abstract

Despite unprecedented motorization in China, most Chinese still rely on non-motorized modes. As a result of rising incomes and a long legacy of bicycle use, electric powered two-wheelers have risen in popularity over the past five years and the country has seen tremendous growth of this mode. Touted as environmentally friendly vehicles, they are capable of traveling about 40-50 kilometers on a single charge and emit zero tailpipe emissions. However, many cities are banning electric two-wheelers from city streets, citing safety and environmental problems. This research quantifies both the negative externalities (environmental and safety impacts) and positive mobility benefits of electric two-wheelers in China. Electric bikes perform well on most metrics, although their use results in much higher environmental lead emissions from battery production and disposal than any alternative. They do however provide unmatched, cost effective mobility and accessibility advantages compared to most alternative modes.

Introduction

Electric bikes have risen in popularity over the past several years, from several thousand sold in 2000 to 16-18 million sold in 2006 (Jamerson and Benjamin 2007). Estimates of total vehicle population range from 30-50 million currently in China. The adoption rate of this mode has far outpaced growth in automobile ownership and has created new challenges for the transportation system and environment in Chinese cities. Proponents of electric two wheelers tout them as environmentally friendly vehicles, reducing local air pollution and providing high mobility to users. Opponents suggest that they are unsafe, polluting and the cause of many of China's traffic problems (Ribet 2005). In fact, several cities have begun banning or heavily restricting electric two wheeler use in their cities (Guangzhou Daily 2006).

There are two classes of electric two wheelers, bicycle style electric bikes and scooter style electric bikes (Jamerson and Benjamin 2007), shown in Figure 1. Electric bikes have a range of 40-60 kilometers on a single charge and cost between US$150-300. Electric two wheelers are loosely regulated by size, weight and maximum speed and are thus classified as bicycles (China Central Government 1999; China Central Government 2004). As such, they are allowed the same privileges and adhere to the same regulations as bicycle riders, including shared right of way in bicycle lanes.

This research explores some of the impacts of electric bikes on the transportation system. The negative externalities (safety and environmental impacts) are explored and quantified, followed by the benefits to the transportation system, primarily mobility and access increases. Two case studies are conducted to identify the *net* impact on safety, environmental emissions, and mobility to the transportation system in the event of an electric bike ban in the cities of Shanghai or Kunming. The last section discusses policy implications and suggests areas where this research can be extended and improved.

Figure 1: Bicycle Style and Scooter Style Electric Bikes

Positive and Negative Impacts of Transportation

Electric two wheelers, like any mode of transportation, have positive and negative impacts to the transportation system. Electric bikes contribute to congestion, they emit pollution, and users are injured and killed in traffic crashes. To counter the negative impacts, they provide mobility in a city, increasing productivity and access to jobs or other destinations. The relevant question is the extent that electric two wheelers perform better or worse on these metrics than alternative modes of transportation. If electric two wheelers were taken out of the choice set, the distribution of the shift to alternative modes (cars, buses, bicycles or walking) would determine whether the policy to remove electric two wheelers increased or decreased the negative externalities of the transportation system. Any policy analysis related to use of electric two wheelers must first identify what modes electric two wheeler users would take in the absence of this mode and then identify the net positive and negative impacts based on environmental, safety and mobility metrics.

Methodology and Data

This research approach looks at electric two wheelers from several different angles. The methods and data are summarized here. For a full discussion, see Cherry (2007). From an environmental perspective, most major stages of an electric bike lifecycle are analyzed, including production processes, use, and disposal. Emissions of criteria pollutants are estimated for each process and the life cycle emission rate per kilometer can be estimated. This is also done for competing modes of transportation as well. Most of the emission data are from literature, interviews with electric two wheeler industry members, and statistical yearbooks (Sullivan, Williams et al. 1998; Zhang, Wang et al. 2001; Lawrence Berkeley National Laboratory 2004; Energy Foundation China 2005; National Bureau of Statistics 2005; Embarq 2006; Mao, Lu et al. 2006). Safety data originated from local public security bureaus. To identify mobility characteristics, published data were used from local sources commenting on the performance of different transportation modes (Fudan News 2004; Kunming

University of Science and Technology 2005). Mobility characteristics of bicycles and electric two wheelers were identified using floating vehicle studies using global positioning systems (GPS). An important input to the analysis is the modal direction that electric bike users would shift in the event of electric two wheeler prohibition. A survey was conducted in two cities, asking several questions, including alternative mode choice if electric bikes were prohibited. For details on this survey and mode choice analysis, see Cherry and Cervero (2007).

Unlike most modes, the environmental impacts vary from region to region, depending upon electricity generation mix. Some regions of China rely heavily on coal power, while others rely heavily on hydro power. The major benefits, accessibility and mobility, also vary from city to city depending on traffic conditions and distribution or housing and work locations. Two cities, Shanghai and Kunming, with different transportation systems and energy mixes were identified as case studies to conduct a hypothetical policy analysis.

Case Studies-What are Impacts of E-bike Ban in Kunming and Shanghai

In the two case study cities, more than 50% of electric bike users are previous bus riders and would shift back to the bus if electric bikes were banned. This has significant implications on bus capacity constraints and emissions. While this is an interesting and somewhat surprising finding, a similar study conducted in Shijiazhuang found that electric two-wheelers are mostly displacing traditional bicycle trips, as one might expect (Weinert, Ma et al. 2007). This could be because Shanghai and Kunming both have exceptional bus systems. Given the predominant alternative mode given by user surveys coupled with trip information from a travel survey and total vehicle fleet population in each city allows inference of the total yearly displaced kilometers traveled by mode. For instance, in 2006 Shanghai had about 1,000,000 electric two wheelers. According to the survey, 56% of all trips would shift to bus if electric bikes were banned. The average yearly travel of electric bike users that would otherwise choose bus is 2975 km, so the total displaced passenger kilometers travelled from the bus is 1,666,000,000 km (1,000,000 e-bikes x 56% x 2975 yearly vkt). Figure 2 shows the distribution of responses from survey respondents in Shanghai and Kunming when asked their alternative mode for a specific trip. From these data, net impacts can be calculated by mapping impact rates (such as SO_2 emissions/km or fatalities/mvkt) from electric bikes to alternative modes.

The negative environmental and safety impact rates in Shanghai and Kunming are shown in Table 1. The first seven columns show the lifecycle emission rates of criteria pollutants, including production, use and disposal phases. Electric two wheelers in Shanghai and Kunming have different emission rates because Shanghai has a 99% reliance on coal electricity generation, while Kunming relies on 48% coal and 52% hydro power. However, compared to Shanghai, Kunming has a larger percentage of scooter style electric bikes, which have higher electricity use and thus emission rates. Electric bikes perform better on some metric and worse on others compared to alternative modes. Most notably, electric bikes perform better than other motorized modes on NO_X, CO_2, and energy use rates. However, they emit much more lead (Pb) pollution into the environment from poor production and recycling processes in addition to relatively low recycling rates (Mao, Lu et al. 2006). The eighth column shows the average fatality rate of modes from several Chinese locations. The last column shows the mobility characteristics (average speed) of each mode.

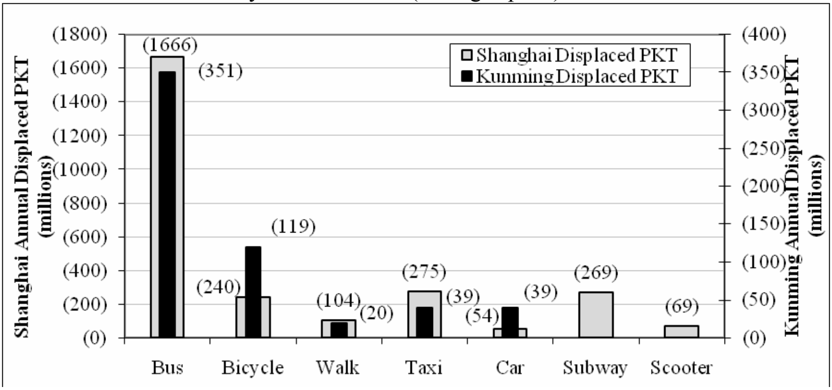


Figure 2: Displaced PKT by Mode-Shanghai and Kunming

Table 1: Impact Rates of Competing Modes in Shanghai and Kunming
(units: grams/passenger/km unless otherwise noted)

	CO*	CO_2	HC*	NO_X*	SO_2	PM	Lead (Pb)	Energy kWh/km	Safety** (fatalities per million vkt)	Avg Speed km/hr
Bus***	0.16	48.43	0.02	0.27	0.02	0.06	0.005	0.13	Unk	7.1
E-bike Shanghai	Unk	29.25	Unk	0.03	0.18	0.15	0.862	0.08	0.019	14.5
E-bike Kunming	Unk	23.17	Unk	0.02	0.11	0.16	0.922	0.07	0.019	14.7
Bicycle	0.00	4.70	Unk	0.00	0.01	0.06	0.000	0.05	0.013	11.0
Car****	9.43	306.0	1.11	1.01	0.69	0.28	0.299	1.40	0.199	15.0

Notes:
Unk=Unknown
*Only the use phase was included due to data availability (production phases emissions are unknown)
**Safety rates shown are averages from Zhejiang and Jiangsu Provinces and Kunming City
***All calculations assume an average bus is loaded to capacity-50 passengers
The average speed of buses is estimated using average access time, wait time, operating speed, and trip distance in both cities outlined in (Fudan News 2004; Kunming University of Science and Technology 2005)
****Emission rates from cars were adopted primarily from (Sullivan, Williams et al. 1998)

Combining data from Figure 2 and Table 1 results in *net* increases or decreases in impacts to the transportation system if electric two wheelers were banned in these case study cities. For instance, banning electric bikes would decrease lead emissions, but increase greenhouse gas emissions and reduce mobility. These net impacts are shown in Table 2. The positive numbers suggest a net increase in impacts, while the negative numbers suggest a net decrease as a result of a ban.

From Table 2, it is clear that electric two wheelers provide clear benefits and costs to the transportation system. For instance, banning electric two wheelers in

Kunming would cause former electric two wheeler riders to shift to a host of alternative modes. This shift would cause an increase of 11 fatalities per year, increased energy consumption, and increased CO_2, NO_X, and SO_2 emission. Moreover, electric bike users stand to spend an additional 145 hours per year commuting, reducing their productivity. The biggest drawback of electric two wheelers is the significant increase in lead (Pb) emitted into the environment. For every electric bike removed, two kilograms of lead would be removed from the environment. This is a very difficult and complex problem and should be remedied by significant improvement in battery production and recycling practices or shifts to more environmentally benign battery technology. The direction of impacts is similar in Shanghai.

Notably, banning electric bikes increases PM emissions in both cities. However, much of this PM emission is released at factories and power plants, away from the city. Therefore the exposure of people (and thus public health impacts) to such pollution is likely lower than tailpipe emissions (Marshall, Teoh et al. 2005; Zhou, Levy et al. 2006). Moreover, the large decreases in NO_X emissions provide health benefits that likely outweigh the health costs of increased PM emissions (Health Effects Institute 2004).

Table 2: Net Yearly Impact Per Electric Bike Banned from Roadway
(Units: Impact per electric two wheeler removed per year, unless noted)

	CO_2 (kg/yr)	NO_X (g/yr)	SO_2 (g/yr)	PM (g/yr)	Lead (Pb) (g/yr)	Energy (kWh/yr)	Safety* (total citywide fatalities per year)	Yearly Travel Time (hr/yr)
Shanghai	+119	+815	-129	-121	-1,908	+521	+39	+126
Kunming	+128	+830	+8	-140	-2,049	+557	+11	+145

Note:
*Safety is expressed as the increase in fatalities if all electric bikes were removed from the roadway in respective cities.

Conclusion and Policy Recommendations

Electric two wheelers have entered the market and grown as a significant mode share in Chinese cities. Much of this growth is in response to increasing trip length, restrictions on motorcycle use, and public transit service struggling with congestion and oversubscription. Many policy makers have praised electric two wheelers, while others have criticized them. This research quantifies a few of the more significant impacts of electric two wheelers on the transportation system. Electric two wheelers provide many benefits that are quantifiable, such as reduced emissions of many criteria pollutants, increased safety, and increased mobility. These benefits do come at a cost, primarily a large increase in lead pollution throughout the lead acid battery supply chain in China. Alternative battery technologies are available and commercially viable, however, the market does not encourage adoption because of their higher cost (Weinert, Burke et al. 2007). Because of their benefits to the transportation system, economic incentives or regulation should be introduced that prompts a shift from lead acid batteries to more expensive Li-ion or NiMH batteries that have fewer environmental externalities. None-the-less, when considering a broad range of costs and benefits, an electric two wheeler with a cleaner battery technology is one of the most cost effective modes of transportation China can offer.

Acknowledgments

This work was sponsored by the UC Berkeley Center for Future Urban Transport-A Volvo Center of Excellence, the NSF, and the UC Pacific Rim Research Program. The author wishes to thank Adib Kanafani, Robert Cervero, Arpad Horvath and Max Auffhammer for support through the research process. Thanks go to partners Jonathan Weinert, Ma Chaktan, Yang Xinmiao, Pan Haixiao, Xiong Jian and Ni Jie.

References

Cherry, C. (2007). Electric Two-Wheelers in China: Analysis of Environmental, Safety, and Mobility Impacts. Civil and Environemental Engineering, University of California-Berkeley. PhD Dissertation.

Cherry, C. and R. Cervero (2007). "Use Characteristics and Mode Choice Behavior of Electric Bike Users in China." Transport Policy **14**(3).

China Central Government (1999). General Technical Standards of E-Bike (GB17761-1999) from National E-bike compelling standards.

China Central Government (2004). National Road Transportation Law.

Embarq (2006). Componente III Pruebas de tecnología de Autobuses.

Energy Foundation China (2005). Electric Utilities.

Fudan News (2004). "上海交通发展第三道难题 机、非高度混杂干扰通行."

Guangzhou Daily (2006). Guangzhou bans Electric Bikes. Guangzhou Daily. Guangzhou, China.

Health Effects Institute (2004). Health Effects of Outdoor Air Pollution in Developing Countries of Asia: A Literature Review.

Jamerson, F. E. and E. Benjamin (2007). Electric Bikes Worldwide Reports -20,000,000 Light Electric Vehicles in 2007.

Kunming University of Science and Technology (2005). Kunming City Bus Network Optimization.

Lawrence Berkeley National Laboratory (2004). China Energy Databook 6.0.

Mao, J., Z. W. Lu, et al. (2006). "The Eco-efficiency of Lead in China's Lead-Acid Battery System." Journal of Industrial Ecology **10**(1-2): 185-197.

Marshall, J. D., S. K. Teoh, et al. (2005). "Intake fraction of nonreactive vehicle emissions in US urban areas." Atmospheric Environment **39**: 1363-1371.

National Bureau of Statistics (2005). China Statistical Yearbook.

Ribet, S. (2005). Two-wheel revolution. The Standard. Hong Kong.

Sullivan, J. L., R. L. Williams, et al. (1998). Life Cycle Inventory of a Generic U.S. Family Sedan Overview of Results USCAR AMP Project, Society of Automotive Engineers.

Weinert, J. X., A. F. Burke, et al. (2007). "Lead-acid and lithium-ion batteries for the Chinese electric bike market and implications on future technology advancement." Journal of Power Sources **In Press**.

Weinert, J. X., C. T. Ma, et al. (2007). "The Transition to Electric Bikes in China: Effect on Travel Behavior, Mode Shift, and User Safety Perceptions in a Medium-Sized City." Transportation Research Record **Forthcoming**.

Zhang, C., C. Wang, et al. (2001). Life Cycle Assessment of Electric Bike Application in Shanghai. Environmental Sustainability Conference and Exhibition, Graz, Austria, SAE International.

Zhou, Y., J. I. Levy, et al. (2006). "The influence of geographic location on population exposure to emissions from power plants throughout China." Environment International **32**(3): 365-373.

Travel Demand Based Optimization Model for Urban Mixed Traffic Systems

Bing-feng Si[1] and Zhili Gao[2]

[1]State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing, China, 100044, e-mail: sibingfeng@jtys.bjtu.edu.cn

[2]Department of Construction Management and Construction, North Dakota State University, Fargo, ND 58105, e-mail: zhili.gao@ndsu.edu

Abstract

The large population and urban mixed traffic flow have caused many urban traffic problems in China. This paper considers the basic factors that influence the choice behaviors of individual travelers, and presents a system optimization model for urban mixed traffic networks using bi-level programming. The model uses the variational inequality method and increases optimization for an urban mixed traffic system and provides a powerful tool to improve city traffic situations. A case study to illustrate the applications of the model and its algorithm is also included in the paper.

Introduction

The large population and the ever-increasing number of cars, together with existing bicycles, buses, and pedestrians on the same roads, have caused many urban traffic problems such as congestion, traffic pollution, and a shortage of fuel. The existing traffic organization and control theories, methods, and models, which were used successfully by developed countries, cannot solve the problems caused by such a mixed traffic flow. In general, many dimensions of travel choices should influence urban traffic assignment, including the travel mode and the route between origin and destination. The mode choice and route choice behaviors should be analyzed simultaneously according to the form mechanism of mixed flow. Past research (Smith, 1983; Sheffi, 1985; Yang & Huang, 2004) focused primarily on motor flow. Some research (Ben-Akiva & Lerman, 1985; Sheffi, 1985; McFadden, 1989; Dial, 1996) was based on urban mixed traffic flow; however, the interference between motor and non-motor vehicles was not significantly considered. Models from these studies, which usually focused on the model establishment and algorithm design and parameters, failed to describe and solve the urban mixed traffic flow problems.

This paper presents a bi-level optimization model for urban mixed traffic networks, including traffic flow-split and traffic assignment. The formulation of the model is established by analyzing the characteristics of different traffic modes. Such a formulation uses the variational inequality method. A type of link cost function is presented, in which the influences between different traffic modes are taken into account, and the traffic demand of different traffic modes is regarded as a variable. The objective of optimization is to reduce the total travel cost of a traffic network.

Model Formulation

This section of the paper presents the formulation of a bi-level optimization model. However, due to the length limit, most of the equations cannot be presented here. In total, 21 equations can be obtained by contacting the authors of this paper.

Link Resistance Function Based on Travel Demand

Commonly, urban traffic network flow is the summation of every mode flow changed into Passenger Car Units (PCU). A link resistance function mainly describes the relationship between travel time and link flow. In most of the previous research, link resistance functions, such as the Bureau of Public Roads (BPR) function and the Davidson (1966) function, were uncomplicated and easy to use but had their inherent limitations since they did not consider the characteristics of every travel mode and flow, and did not describe the obvious interference between traffic modes in China, especially between motor vehicles and non-motor vehicles. Therefore, this study establishes a new type of link resistance function based on travel demand, in which the independent variable is the number of travelers using different traffic modes.

Choice Behaviors in Urban Mixed Traffic Networks

Once a destination has been decided, a traveler always chooses a traffic mode and route simultaneously to either minimize the travel time or minimize the travel cost. In the formulas for this study, each travel option is represented by a unique symbol to show the benefit of choosing that option.

The general travel cost of the modes between the origin-destination (O-D) pairs in the urban mixed traffic network is then formulated. The mode choices made by a traveler becomes a probability issue. The probability of a mode between O-D pairs can be calculated as the probability of the minimum cost of that mode.

Assignment Model for Urban Mixed Traffic Networks

It is evident that the traffic flows of different traffic modes interfere with each other. Consequently, the Jacobian matrix becomes a nonsymmetrical matrix. Smith (1983) stated that it is hard to build an optimization model equivalent to the traffic assignment, but a variation inequality (VI) model can solve this problem. This study solves an urban mixed traffic assignment by using VI model to find $(\mathbf{x}^*, \mathbf{q}^*) \in \Omega$ to satisfy $\mathbf{t}(\mathbf{x}^*)^T (\mathbf{x} - \mathbf{x}^*) + \mathbf{g}(\mathbf{q}^*)^T (\mathbf{q} - \mathbf{q}^*) \geq 0$, where $\mathbf{x}$ and $\mathbf{q}$ are the vectors.

System Optimization Model for Urban Mixed Traffic Networks

A city government has to harmonize different traffic modes and to optimize the structure of its urban traffic system in order to manage the urban traffic successfully. The city government influences traveler choice behaviors by making traffic polices and laws, and can utilize these policies to arrange different travel modes and routes based on travel demands to improve urban traffic networks. Once a traffic flow assignment has been determined, a system optimization model can provide a decision support for the city government to improve the whole urban traffic network. Consequently, this study sets traffic policies (i.e., traffic pricing, financial allowance, and traffic priority) as parameters in the system optimization model. This process can be described as the upper level (U) of a bi-level programming model:

$$\text{(U)} \qquad \min Z(\boldsymbol{\lambda}) = \sum_k \sum_a x_a^k(\boldsymbol{\lambda}) \cdot t_a^k\{x_a^k(\boldsymbol{\lambda})\}$$

where $\boldsymbol{\lambda}$ represents the policy index such as traffic pricing.

On the other hand, travelers make decisions independently based on their experiences and actual situations. This process can be described by the lower level of a bi-level programming model (a VI model with parameters):

$$\text{(L)} \qquad \mathbf{t}(\mathbf{x}^*,\boldsymbol{\lambda})^T(\mathbf{x}-\mathbf{x}^*) + \mathbf{g}(\mathbf{q}^*,\boldsymbol{\lambda})^T(\mathbf{q}-\mathbf{q}^*) \geq \mathbf{0}, \text{ and}$$

$$(\mathbf{x}^*,\mathbf{q}^*) \in \Omega$$

In the upper level of the model, the objective function is used to minimize the total cost of an urban traffic network. In the lower level of the model, the objective function describes the choice behaviors of travel modes and routes under each traffic policy or road situation. It has been proven that the solution of this optimization satisfies a logit based modal split and user equilibrium (UE) condition (Si et al, 2006).

Algorithm Solving

An algorithm based on sensitivity analysis is adopted to solve the above bi-level programming model. Assuming the initial data and other conditions are fixed, the link flow matrix in a mixed traffic network can be obtained by solving the lower level of the model. From a sensitivity analysis of a VI model, the differential coefficient of a link flow matrix can be obtained. After inputting the link flow matrix into the upper level of model, model is simplified as a common optimization. The solution to such a common optimization is then inputted into the lower level of the model to start the next iteration. With repeated calculations, it is possible to obtain an optimum solution of the bi-level programming model. This process is summarized as following steps:

Step 1: Set the initial value and set the number of iterations to $n = 1$.

Step 2: Find the solution of the lower level of the model to obtain the link flow matrix.

Step 3: Find the linear equation of the matrix by sensitivity analysis and Taylor expansion.

Step 4: Input the linear equation of the matrix into the upper level of the model to update the parameters.

Step 5: Examine the convergence. If $\boldsymbol{\lambda}^{(n+1)} \approx \boldsymbol{\lambda}^{(n)}$ or $n = m$, the iteration stops, where m is the maximum number of iteration. Otherwise, set $n = n+1$ and start a new iteration.

Case Study

A simple case study is used to illustrate the application of the model and its algorithm above. The traffic network of this case study is shown in Figure 1, which contains one O-D pair (1-9), 9 nodes, 12 links, and 3 travel modes. The general travel cost of each mode is represented by equation $c_k^{rs} = \mu_k^{rs} + aF_k^{rs} - bE_k^{rs}, \forall k, r, s$ where F_k^{rs} and E_k^{rs} represent the price and accessibility of travel mode *k* between O-D pair *rs*; *a* and *b* are weight parameters, here a=0.2, b=0.1.

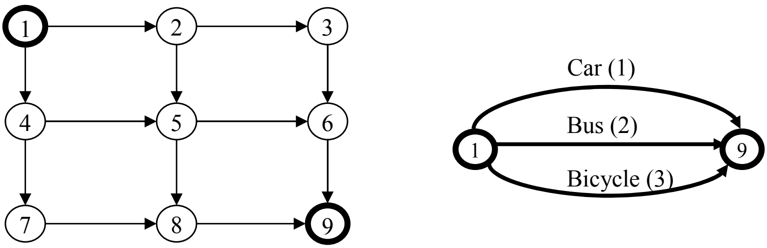


Figure 1. Traffic network structure (left) and traffic mode structure (right)

The input data of links is illustrated by Table 1, while the PCU conversion coefficient (F_k^{rs}), and the average number of passengers (E_k^{rs}) are illustrated by Table 2. It is recommended that parameter θ is 0.5, which describes the equilibrium relationship between travel time and other factors in the general cost equation. If travel time is given more weight, θ should be bigger; otherwise, θ should be smaller. By applying some management methods or policies, the city government can change λ_k^{rs} to guide the traveler choice behaviors so as to optimize the urban traffic network.

Table 1. Data of different link

Link	$t_a^{1(0)}$/(h)	$t_a^{2(0)}$/(h)	$t_a^{3(0)}$/(h)	C_a/(P.h^{-1})	Link	$t_a^{1(0)}$/(h)	$t_a^{2(0)}$/(h)	$t_a^{3(0)}$/(h)	C_a/(P.h^{-1})
(1,2)	0.111	0.178	0.261	1000	(5,6)	0.094	0.161	0.244	1000
(2,3)	0.128	0.194	0.278	700	(4,9)	0.133	0.200	0.283	900
(1,4)	0.100	0.167	0.250	1500	(5,8)	0.111	0.178	0.261	700
(2,5)	0.106	0.172	0.256	700	(6,9)	0.144	0.211	0.294	700
(3,6)	0.089	0.156	0.239	700	(7,8)	0.094	0.161	0.244	900
(4,5)	0.078	0.144	0.228	1000	(8,9)	0.100	0.167	0.250	900

Table 2. Data of different models

Mode	PCU coefficient	Average passengers	C_k^{rs}	E_k^{rs}
Private cars	1	4	10	10
Buses	1.5	20	4	5
Bicycles	0.25	1	0	1

The total travel demand denoted by q^{rs} is used to describe the congestion condition of the urban traffic network. Two extreme scenarios have been considered: (1) q^{rs}=10000/PCU, representing a network without any congestion; and (2) q^{rs}=30000/PCU, representing a network that has reached the congestion saturation point. Figure 2 shows the results for the second scenario.

It can be seen that, with the increase of the attractive rate for one mode, the occupation rate of this mode rises. Furthermore, once the whole network reaches saturation, the travel mode that has the least travel time is utilized more than others.

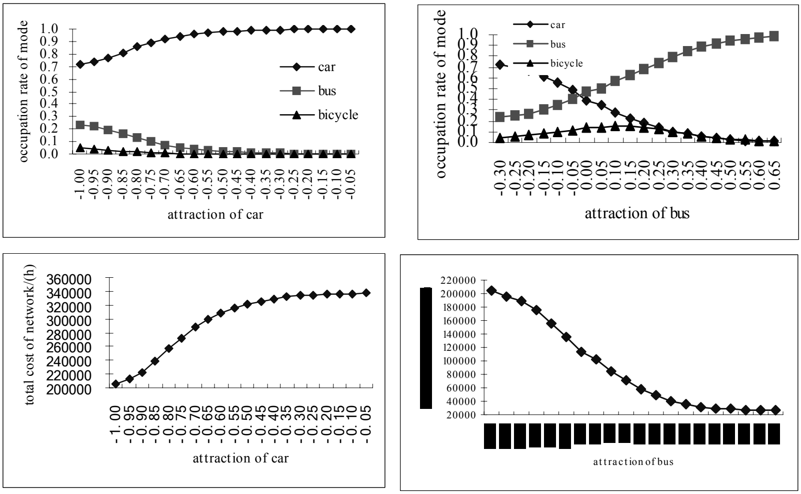


Figure 2. Occupation rate and total travel costs of cars and buses

In addition, if no congestion exists on the network, it is possible to reduce the total network cost by making cars and buses more attractive. Moreover, increasing the occupation rate of cars can reduce the total cost of the network more than increasing the rate of buses. In contrast, a higher occupation rate of bicycles can make the network worse due to the capabilities of bicycles. With the existence of high congestion in the network, increasing the number of buses would be effective to cut down the total network cost. When people choose either cars or bicycles but not buses the total network cost always increases. In some cases, this increase may be significant.

Based on the case study, with the increased travel demands in big cities, travelers tend to select travel modes that are less time consuming, such as cars. If an urban traffic system has not reached its saturation, the total travel cost of the network would go down if cars and buses share more transit responsibility than bicycles. When a network reaches its saturation, expanding the number of either cars or bicycles would increase both the cost and congestion of the network. To relieve congestion in this situation, developing and enhancing public transportation are the most successful and accessible methods. This explains why public transportation is given top priority in many countries such as China.

Conclusion

In this paper, the basic factors that influence travel choice behaviors of travelers are considered. A new link resistance function is established based on travel demand. The variation inequality assignment model is designed and an optimization model for urban mixed traffic flow networks is formulated. Based on the analysis of the case study, this new optimization model can solve the traffic flow assignment in an urban mixed traffic network.

However, this case study only optimized the total system cost in order to minimize the calculation. In actual urban transportation planning, additional factors must be considered such as the influence of the automotive industry, the impact on the environment, and the budget constraints. These problems require further research on urban flow assignment and optimization.

Acknowledgements

The work described in this paper was mainly supported by a grant from the National Natural Science Foundation of China (Project No. 70631001) and the National Basic Research Program of China (Project Nos. 2006CB705500).

Reference

Ben-Akiva, M. and Lerman, S. R. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. The MIT Press, Cambridge, MA.

Davidson, K. B. (1966). "A flow-travel time relationship for use in transportation planning." *Proceedings of Australian Road Research Board*, Belbourne, Australia, vol. 3, 183-194.

Dial, R. B. (1996). "Bicriterion traffic assignment: Basic theory and elementary algorithms." *Transportation Science*, 30(2), 93-111.

McFadden, D. (1989). "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration." *Econometria*, 57(5), 995-1026.

Sheffi, Y. (1985). *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, Englewood Cliffs, NJ.

Si, B. F. and Gao, Z. Y. (2000). "Sensitivity Analysis for the Relationship between Highway Prices and Passenger Demands for Different Services." *China Journal of Highway and Transport*, 13(2), 91-95.

Si, B. F. and Gao, Z. Y. (2005). "Optimization Model for Rail Passenger Transportation under the Condition of Integrated Transportation System." *Journal of the China Railway Society*, 27(5), 6-10.

Si, B. F., Sun, Z. Z. and Zhao, X. M. (2006). "Mixed Traffic Network Flow-Split Model Based on Stochastic User Equilibrium." *China Journal of Highway and Transport*, 19(1), 93-98.

Smith, M. J. (1983). "Algorithm for Solving Asymmetric Equilibrium Problem with a Continuous Cost-flow Function." *Transportation Research*, 17B(5), 365-371.

Yang, H. and Huang, H-J. (2004). "The multi-class, multi-criteria traffic network equilibrium and systems optimum problem." *Transportation Research*, 38B(1), 1-15.

EVALUATING THE QUALITY OF PEDESTRIAN INFRASTRUCTURE AND SERVICES IN CHINESE CITIES

Ms. Holly Krambeck, Associate, PB

Dr. Jitendra (Jitu) Shah, Country Sector Coordinator, EASES, SE Asia, World Bank

Abstract

Although a significant number of trips are made by foot in developing cities, pedestrian infrastructure, amenities, and services are often neglected in municipal planning and budgets. This paper presents the processes and results in the development of a Walkability Index, which would rank cities within a country and across the world based on the safety, security, and convenience of their pedestrian environments.

The Index is a tool that could be easily and economically implemented throughout Chinese cities to raise awareness of walkability as a critical issue, and more importantly, provide cities with the data and information they need to develop informed investment plans and policy decisions.

This paper provides a summary overview of how the Index is constructed and implemented.

1 INTRODUCTION

Although a significant number of trips are made by foot in developing cities, pedestrian infrastructure, amenities, and services are often neglected in municipal planning and budgets. This paper presents the processes and results in the development of a Walkability Index, which would rank cities within a country and across the world based on the safety, security, and convenience of their pedestrian environments.

The Index is a tool that could be easily and economically implemented throughout Chinese cities to raise awareness of walkability as a critical issue, and more importantly, provide cities with the data and information they need to develop informed investment plans and policy decisions.

Project objectives may be summarized as follows:

- Generating awareness of walkability as an important issue in developing cities;
- Providing city officials with an incentive to address walkability issues;
- Helping city planners understand scope and extent of local pedestrian conditions, relative to other cities; and

- Providing city planners with the information necessary to identify specific pedestrian-related shortcomings, as well with recommendations for next steps.

2 INDEX COMPONENTS

The Walkability Index, designed around the aforementioned project objectives, comprises three components: safety and security, convenience, and degree of policy support.

Component 1: Safety and Security is intended to determine the relative safety and security of the walking environment. For example, what are the odds a pedestrian will be hit by a motor vehicle? What safety measures are in place at major crossings and intersections? How safe from crime do pedestrians feel along walking paths?

Component 2: Convenience and Attractiveness reflects the relative convenience and attractiveness of the pedestrian network. For example, do pedestrians have to walk a kilometer out of their way just to cross a major road? Is there sufficient coverage from weather elements along major walking paths? Are paths blocked with temporary and permanent obstructions, such as parked cars or poorly placed telephone poles?

Component 3: Policy Support reflects the degree to which the municipal government supports improvements in pedestrian infrastructure and related services. Is there a non-motorized planning program? Is there a budget for pedestrian planning? Are pedestrian networks included in the city master plan?

In a previous iteration of the Index, these three components were further subdivided into 22 indicators and 45 variables. Since the overwhelming response from field testers was that the heavily detailed Index, while perhaps more technically robust and defensible, was not practical for implementation, subsequent surveys and trials were conducted to develop the more simplified Index presented in this paper. The new Index compromises thoroughness for practicality, yet still stands as a plausible indicator of walkability in cities around the world. The simplified Index variables are:

Safety and Security

1. Proportion of road accidents that resulted in pedestrian fatalities
2. Walking path modal conflict
3. Crossing safety
4. Perception of security from crime
5. Quality of motorist behavior

Convenience and Attractiveness

6. Maintenance and cleanliness of walking paths
7. Existence and quality of facilities fro blind and disabled persons
8. Amenities (e.g., coverage, benches, public toilets)
9. Permanent and temporary obstacles on walking paths
10. Availability of crossings along major roads

Policy Support

11 Funding and resources devoted to pedestrian planning
12 Presence of relevant urban design guidelines
13 Existence and enforcement of relevant pedestrian safety laws and regulations
14 Degree of public outreach for pedestrian and driving safety etiquette

These components, indicators, and variables are the final product of the evaluation of more than 20 different established methodologies for evaluating urban non-motorized transport; evaluation of three different econometric methods for compiling indices; consultations with experts in urban planning, pedestrian planning, transportation engineering, urban transport policy, pedestrian safety, accessibility for disabled persons, urban design, and economics; and comments from field testers in developed and developing cities, including: Alexandria, VA, USA; Beijing, PRC; Washington, DC, USA; Hanoi, Vietnam; Bangkok, Thailand; Manila, Philippines; Karachi, Pakistan; Delhi, India; Ahmedabad, India; Chicago, IL, USA; and Yogyakarta, Indonesia.

3 DATA COLLECTION METHODOLOGY

The quality of the data collection methodology largely determines the overall quality and usefulness of the Walkability Index. That said, while it is desirable that the data collection methods are thorough, they should also be very simple to ensure widespread, error-free implementation. To this end, two separate surveys were developed for collecting the data -- a public agency survey and a field survey – as well as a guidebook.

It is important that these surveys are conducted by local populations to prevent undue bias in results. We are more interested in attaining a walkability index that ranks cities on pedestrian facilities and services, *relative to their local political and economic condition*s, rather than an index that merely mirrors GDP rankings. This need was most acutely realized when comparing pilot survey results between Washington, DC, and Ahmedabad, India, where Washington testers gave consistently worse marks for motorist behavior, relative to the marks given in Ahmedabad, where most would agree traffic conditions and motorist behavior are relatively much more dangerous. A full set of survey materials may be found at: http://www.cleanairnet.org/caiasia/1412/article-60499.html

4 SURVEY AREA

It is important that selected survey areas within cities provide comparable results, and the areas surveyed are representative of a large cross-section of cities' varied neighborhoods and districts. Establishing a survey area selection methodology that suits these criteria is quite difficult, especially given that while some parts of a city may be very walkable, other parts may not – the walkability of cities is almost never uniform throughout.

Bearing this issue in mind, the study considered a variety of different survey area selection methodologies, finally deciding upon using a random spatial sampling technique – a method initially proposed by Judy Baker, an economist at the World

Bank. In this method, the urban area is segmented into a 500 meter and 500 meter grid, with each square populated by a randomly-generated number. First, four squares are "pre-selected" – we want to be certain that at least one high-income, low-income, commercial, and transport hub location are include in the survey. The remaining squares (the number is a pre-determined proportion of the total area) are then randomly selected.

This method is advantageous in that: 1) the random component mitigates some bias from the results, therefore making the survey data more readily comparable across cities; and 2) surveying a square area rather than a selection of single streets ensures issues such as connectivity can be captured in the data – that is, surveying whole areas give is a sense of general walkability for a whole neighborhood, as opposed an isolated road that may or may not be of import. One drawback is that a random spatial sample, inherently, will not cover all areas in the city and may miss important corridors. But, since this is the case for all cities, and since these surveys are conducted for the purposes of constructing an index, as opposed to an investment program, this loss may be considered acceptable. The more areas that are selected (and therefore the less-detailed the surveys are), the more this issue may be mitigated.

5 EXTENDED SURVEY MATERIALS

As mentioned previously, while the Global Walkability Index serves to raise awareness of walkability as an important issue, it is too general for use in devising an investment or policy strategy. Thus, the *Extended Survey Materials* have been developed, which would enable cities to pinpoint specific infrastructure and policy needs, in addition to deriving the simple Index for ranking. The extended materials comprise three sets of surveys: a physical infrastructure survey, a questionnaire to be administered to pedestrians, and a public agency interview form.

The <u>physical infrastructure survey</u> is a "supply side" tool used to collect raw data on the availability and quality of pedestrian infrastructure. The survey is relatively simple and could be conducted by volunteers with minimal training.

The <u>public agency survey</u> is used to collect important data that is not obtainable through physical infrastructure surveys, such as pedestrian fatality statistics and pedestrian-related laws and regulations.

The <u>pedestrian survey</u> is used to collect "demand side" data and enables residents most impacted by the walkability of a city to voice their opinions on current conditions and to suggest improvements. Topics covered in this survey include: perception of safety, quality of mode transfers, accessibility of low-income neighborhoods to places of work and public services, and general convenience afforded by the walking environment. As with the physical infrastructure surveys, these may be conducted by volunteers with minimal training.

To reiterate, in practice, the Global Walkability Index survey materials are simple and general, giving cities only a vague picture of their strengths and weakness and some sense of how their walkability compares to that of other cities. This process serves to generate awareness of walkability as an important issue and provides justification for more a more in-depth examination. The Extended Surveys are a simple tool Chinese cities can use to collect quantitative and qualitative data about existing pedestrian infrastructure conditions, feedback from residents on relevant pressing concerns, and a clear assessment of exiting institutional capacity and policies for ensuring safe, secure, and convenient pedestrian environments.

6 FULL SCALE PILOT

6.1 Overview

Pedestrian infrastructure, as far as transport infrastructure is concerned, is relatively simple and inexpensive to build, yet its impact on the quality of life and functioning of the whole urban transport system can be quite significant. Given this, we might assume that reasons for pedestrian planning neglect rest more closely with unawareness and lack of incentive than with any kind of inability or gross fiscal constraint. So, a logical first step towards improving walkability in a city would be to clearly define the problem and generate awareness of walkability as an important issue among city residents and officials. The data generated to define the problem (collected through physical infrastructure, public agency, and pedestrian surveys) could then be used towards developing a targeted investment program to improve walkability in select, high-profile areas, such as areas around schools or transport hubs. The support of local stakeholders should be solicited, to ensure light pressure is continuously applied to city officials to sustain walkability efforts. Also, ideally, a pedestrian infrastructure investment program would be tied to a larger existing project, such as a road upgrading project. Finally, an investment program should also include a policy component to ensure that the infrastructure is maintained, kept clear of obstructions, and so forth. The following discussion shows how these steps have been taken in Ahmedabad, India.

6.2 Pilot City Background

A full-scale pilot was conducted in Ahmedabad, India, in August 2005, using the extended survey materials. Ahmedabad, a city of 4.6 million people (150 people per hectare), is the commercial capital of one of India's wealthiest states, Gujarat. The local economy is primarily based on light industry and academic institutions, though there is an unusually large informal sector, comprising 77% of the workforce (SEWA interview 2005).In 2004, there were 1,490,000 registered motorized vehicles in Ahmedabad (including private automobiles, taxis, two-wheelers, etc.), and this number is expected to increase, on average, by 13% over the next few years (CEPT interview 2005). In terms of public transport provision, there were 540 public buses covering 150 routes and serving 385,682 passengers per day in 2003 (CEPT interview 2005). According to a study completed in 2000 by Louis Berger, about 38% of all commuting trips are walking trips.

6.3 Pilot Process

Walkability surveys have been conducted in eight-square kilometers of the city by 65 student volunteers from the Centre for Environmental Planning and Technology;

presentations have been made to local and state government officials; and the efforts of three local organizations, the Self-Employed Women's Association (SEWA), the Environmental Planning Collaborative (EPC), and the Center for Environmental Planning and Technology (CEPT) have been coordinated to oversee the development of the initial investment proposal, which will target walking conditions around schools, informal markets (where there tend to be large numbers of pedestrians), and transport hubs. Finally, the project will be tied to two larger, already established investment projects – a proposal for a BRT corridor, and a 125-kilometer road upgrading project.

6.4 Impact of the Ahmedabad Pilot

Because the collected dataset is site specific (physical infrastructure data is collected per road stretch and intersection, and pedestrian data is collected per survey area), it is very useful for developing targeted investment programs and generating rough cost estimates for infrastructure development. Further, when the pedestrian and infrastructure survey results are combined with results from the public agency survey, decision-makers have a valuable tool for developing effective, long-term policies directed at improving and maintaining the pedestrian environment.

The survey work was relatively simple and inexpensive. Most survey teams were able to complete the survey work and data entry in one day. Costs, including survey materials for volunteers, T-shirts, transportation, and a volunteer dinner totaled USD 222. Use of volunteers for survey work and data entry both kept costs down and generated visibility and excitement for the cause of improving walkability. To illustrate, after a workshop where volunteers presented results and worked with representatives from the Self-Employed Women's Association (SEWA) and the Environmental Planning Collaborative (EPC), the *Times of India* published a story about walkability in Ahemdabad, with the headline: "In Ahmedabad, Taking a Walk Can Be Tortuous" (September 4, 2005; page 3). The article drew the attention of the State government, and the author was invited to meet with the Gujarat State Joint-Secretary of urban development as well as officials from the Gujarat Infrastructure Development Board in Ghandinagar after the article's publication.

In short, the Index achieved exactly what it set out to do – it quickly and effectively raised awareness of walkability as an important issue to be addressed.

7 IMPLEMENTATION IN CHINA

Chinese cities, noted throughout the world for their competitive nature, would be very well suited to participate in a country-wide Walkability Index exercise. It is not unreasonable to assume that cities raking among the least walkable in China would take advantage of the free tools provided by the Walkability Index to improve their rankings in future years, or for high ranking cities to preemptively use Index guidelines for urban development to ensure sustained high rankings. Incentive programs could be developed in tandem with organizations such as the Global Environment Facility, Cities Alliance, or the World Bank to help cities implement investment plans derived from the extended Index surveys.

Simulation-Aided Airport Terminal Design

Dr. Michel A. Thomet[1], and Farzam Mostoufi[2]

[1]Bechtel Corp., 50 Beale St, San Francisco, CA 94105; PH (415) 768-8521, email: mthomet@bechtel.com
[2]Bechtel Corp., 50 Beale St, San Francisco, CA 94105; PH (415) 768-8525, email: fmostouf@bechtel.com

Abstract - This paper presents the application of simulation techniques to the design of the New Curaçao Airport Terminal. The purpose of the simulation is to confirm that the design will meet or exceed IATA Level of Service C (LOS C) planning standards during periods of peak activity of the design day.

The simulation model is a dynamic, object-oriented passenger movement analysis tool. The model is driven by a realistic flight schedule developed for a 24-hour design day, thereby providing passenger volumes and flows that reflect the arrival and departure of aircraft and passengers over the course of an entire day.

I THE NEW CURAÇAO AIRPORT TERMINAL

The new airport terminal started operations in July 2006. At opening it will be capable of handling 1.6 million annual passengers (Phase 1), although this traffic level is not expected to occur for at least five years after opening. Its ultimate capacity will be 2.5 MAP (phase 2). The airport's role is mainly as a terminal for Carribbean Basin traffic with some European (Dutch) tourists and US tourists (via Miami) together with a small business segment. There is also a very small number of transfer passengers to/from the other Netherland Antilles (Aruba, Bonaire, St.Maarten).

Curaçao Terminal and Concourses at Opening

Eighteen airlines will serve the airport, of which three will be US (American, Continental, and Delta). One airline, Dutch Carribbean Express, will carry almost half of all passengers, and connect Curaçao to the main Caribbean islands (Jamaica, Haiti, Santo Domingo, Trinidad & Tobago as well as the other Netherland Antilles, Aruba, Bonaire, St.Maarten) and to several cities in nearby Venezuela (Caracas, Valencia, Maracaibo). This airline will have long range flights to Amsterdam and Miami as well. Other airlines will serve several South American countries (Venezuela, Colombia, Surinam, and Peru) and Central American countries (Costa Rica). Flights to Cuba and Puerto Rico will also be available from Curaçao. This will make Curaçao a very flexible and convenient hub for the Caribbean Basin.

The equipment used to serve this market will go from long range, E-size aircraft with 400 seats, like the B-747, to small, twin turbos with 48 seats like the DeHavilland Dash 8. Each day, three B-747 flights from Europe will serve Curaçao from London, Madrid and Amsterdam. To serve these flights, the apron is designed with 12 positions, five of which will have access to the terminal by passenger loading bridges. There will be 3 positions for B-747, 3 for B-767, and 6 for the Dash 8.

The airport has a curfew from about eleven at night till six in the morning. From then on, the flights start building up and by mid-day a peak of activity is reached with 9 aircraft in the peak arrival hour (15%) and also 9 aircraft in the peak departure hour (not the same hour). During these hours, the passengers peak are also reached with 920 arrivals and 940 departures (over 18% of the daily passengers).

The worse conditions in the terminal will occur during the peak hour, when two B-747 arrive at the same time. The terminal is designed to have enough facilities and public space so that the Level-Of-Service (LOS), during this peak hour will not drop below LOS C (as per IATA definitions).

Because of this concentrated activity for such a compact airport, the planning methodologies advocated by IATA, and the FAA are more difficult to apply. For the design of the terminal, it was therefore decided to use a passenger simulation model, in addition to the IATA and FAA methodologies. The simulation model makes it possible to quantify the level of service (LOS) in each area of the terminal and for each type of airport patron, given a specific terminal size and layout and a specific scenario of arriving and departing flights during a whole day (the design day).

II PASSENGER SIMULATION MODEL - TERMSIM

The simulation model is driven by the design day flight schedule, where each flight has a time of arrival or departure, consists of a specific aircraft type, is assigned to a specific gate or remote stand, and belongs to a specific airline with a unique flight number.

In addition, the number of originating or terminating passengers and the number of transferring passengers in each flight are based on projected load factors. These passengers are further broken into first, business, and economy class passengers.

For each departing flight, originating passengers are generated in the model at a time determined by their scheduled departure time and an earliness-of-arrival distribution. Once created by this process, each originating passenger has the following attributes:

- Travel class in first, business, or economy
- Time and location of arrival at the airport
- Ground transportation mode used

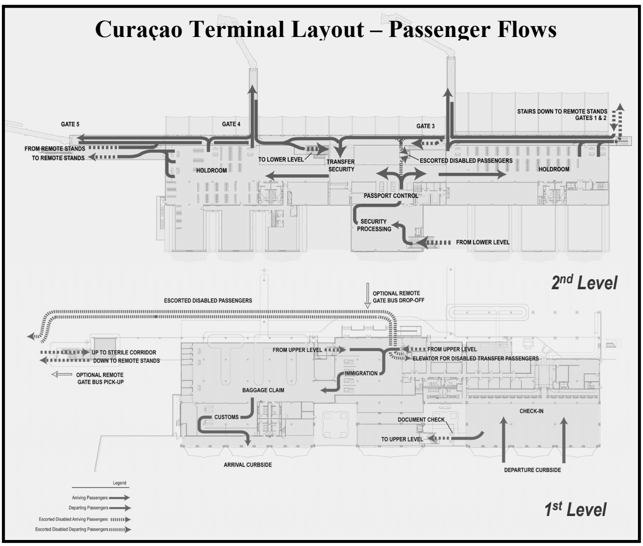
- Airline and flight number
- Departure boarding lounge

In addition, each originating passenger is assigned additional attributes:

- Number of people travelling together in a party (party size distribution)
- Number of checked bags per passenger based on a bag distribution range, as well as a probability distribution that a percentage of these bags are oversized or require special handling (i.e., animals)
- Number of carry-on bags per passenger sampled from a distribution range
- Number of well-wishers per party sampled from a well-wisher distribution range
- Special needs passenger assumptions to account for wheelchairs or electric carts

A similar process is used for generating terminating and transferring passengers in arriving flights at their specific gate or hardstand, with appropriate attributes assigned as the passengers exit the aircraft and move into the terminal or in a bus.

The passengers thus generated move within the terminal and concourses from one area to another according to their attributes. Each area or processing station has a location in the model specified by an x-y-z coordinate that is tied to a scaled CADD drawing of the terminal, such as the one showed in the Figure below. The distance between two areas is calculated as the most direct distance along a travel route or, when a straight line between two areas is not physically possible, intermediate points are defined through which the passengers must pass.



The walking time between areas is computed by giving each party a walking speed, sampled from a distribution, between a minimum and a maximum. Two such speed distributions are used: one for passengers with special needs (handicapped) and one for all other passengers. These speeds are reduced when the occupancy of the area that the passengers traverse rises above a given threshold (crowding effect). The randomization of walking speeds is utilized to reflect the reality of people moving in a terminal.

The model simulates real time behaviours. For example, while moving through the terminal, passengers have the option of walking, using a moving walk, or boarding the automated people-mover. When changing levels, they can use escalators, elevators, or stairs. On the escalators, some passengers will stand while some will walk, adding their speed to that of the escalator. Likewise, on the moving walks, some passengers will stand while some will walk, adding their speed to that of the moving walk.

When passengers arrive at a processing area, they first join a queue. Queues can be universal (a single queue serving several identical checkpoints or processes) or individual (one queue per checkpoint or process). When passengers reach the head of the queue, they are processed. The processing time is a value sampled from a distribution range specific to each facility and passenger attribute.

As passengers move through successive processing checkpoints or areas in the terminal and concourse, their movement is followed by the model and statistics are generated. The occupancy of each area (circulation or queuing) is tracked during the 24-hour simulation period. At each processing point the flow of passengers and queue length is tracked during the day and these data points are, in turn, used to assess the different Level of Service (LOS) metrics.

The model collects all these output variables on a minute-by-minute basis. This enables the planners and architects to design the facilities, public spaces, and corridor widths for the peak activity traffic of the design day.

III *INPUT ASSUMPTIONS*

These assumptions fall into four categories:

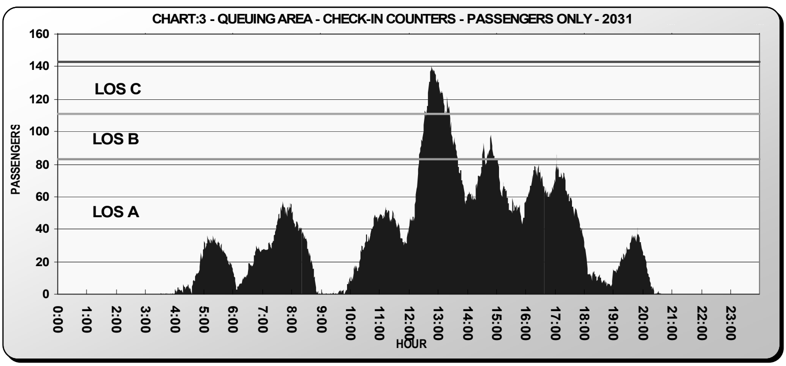
- Description of all the terminal spaces and processing facilities. This is summarized in the CAD drawings of each floor plan of the terminal building and of each concourse. On each drawing, the flow paths of originating, terminating, and transferring passengers are shown, as well as the paths of the electric carts.
- Flowchart of passenger circulation and processing. For each category of passengers (originating, terminating, and transferring), a flowchart describes all the facilities the passengers have to visit, and the order in which the facilities are traversed, from the time passengers arrive at the airport until they leave the airport.
- Functional description of each processing facility and sub-facility. For each category of passenger, a detailed table summarizes the facility parameters such as the percent of passengers using it, together with the processing time distribution (maximum, minimum, and average. Some facilities have no associated processing time, but passengers wait a specific length of time (e.g. well-wisher leaving point in departure hall) or wait for a specific event (e.g. boarding call in the departure lounges).
- Minimum Acceptable Level of Service in each Facility and Sub-facility. In addition to the IATA Levels-of-Service, which are based on areas available per passenger, a maximum dwell time not to be exceeded is specified (queuing time

plus processing time). Likewise a 90 percentile dwell time is also specified. This means that 90% of the passengers processed at that facility should have a dwell time shorter or at most equal to that last criterion.

IV SIMULATION OUTPUTS

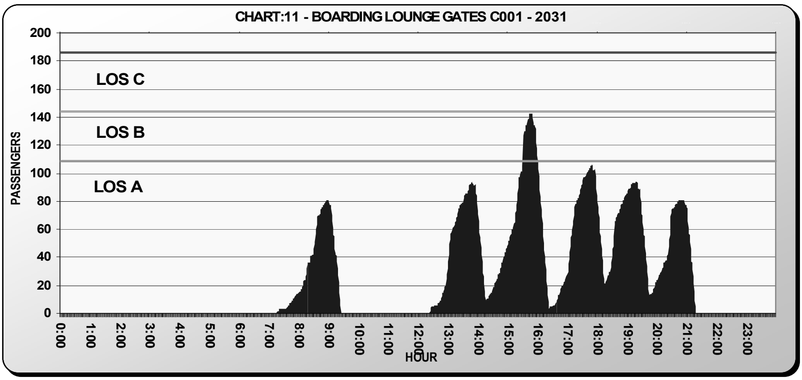
The results of the simulation are summarized in five categories:

- Determination of facilities requirements. When the simulation run is progressing, the model automatically adds facilities to ensure that the desired LOS is not exceeded when the demand for this facility keeps growing. For instance, at the economy check-in, when the waiting time for 90% of the users exceeds 10 minutes, a new counter is opened.
- Performance of each processing station. The number of passengers processed during the design day is summarized in a comprehensive table for each processing station, together with the percentage of passengers that did not have to wait in a queue, the mean wait time and max wait time for all passengers, as well as the max queue length.
- Queuing areas. Groups of processing stations are generally fed from a single queuing space (universal queue). For instance, the economy check-in counters are fed from such a queue. The LOS in the queuing area is determined by the number of passengers, and the size of the queuing area, using the IATA criteria. For each queuing area there is a graph showing the number of passengers in the queue every 10 minutes, together with lines showing the boundaries between LOS, as shown below.



- Clearance times. At the end of the simulation, each passenger from each category "remembers" the time spent in the airport, between the time of arrival or departure at a gate and the time of entrance or exit from a ground access mode. These clearance times are then ranked from the shortest to the longest and displayed in separate histograms for terminating passengers, originating passengers and transferring passengers. Such a histogram is shown below for transfer passengers.

- Space occupancy. For gate lounges and public lobbies, where people are waiting in a given space, an occupancy graph is given, as well as a graph of the corresponding LOS, based on IATA criteria. For corridors, where people are walking, similar graphs are given, based on people per minute walking past a cross-section. The LOS is calculated by cordoning off a section of the corridor with virtual doors and counting how many people are in this section, every 10 minutes. The number of people in that section is found by adding one every time a person passes the virtual entry door and subtracting one whenever a person passes the virtual exit door.



V CONCLUSIONS

Curaçao is a small, compact airport in a very dynamic environment. In the contract, one of the items of the performance specifications, was to demonstrate that even during the peak hours of the ultimate forecast, the Level-Of-Service should be C or better. The Bechtel proprietary simulation package TERMSIM, has made it possible to quickly investigate the performance of different terminal layouts and translate them into the design changes necessary to accommodate the traffic projections at the desired design standards. Because TERMSIM can be used at different levels of details, its use is practical and effective even for small airports like Curaçao.

Optimal toll-area and toll-level design of road networks on base of second-best multi-class based congestion pricing theory

Yao hongyun[1] Meng jingyu [2] Feng xiao [3] Christopher Garlick, P.E.[4]
[1]Transportation College, Chongqing Communication University, Chongqing 400074,China, Email:yaohongyun760610@163.com; [2]Chongqing yuda road and bridge company, Chongqing 400074, China, Email:mengjingyu@163.com; [3]Engineering and Structure College, Chongqing Communication University, Chongqing 400074,China [4]PBS&J, Orlando, FL, USA email: crgarlick@pbsj.com

Abstract: Congestion pricing strategy as a traffic demand management measure is widely accepted by the public. However, there is little focus on the research of optimal toll-areas. This paper puts forward the Second-best multi-class based bi-level planning model of road network congestion pricing in elastic demand, where the upper level model attempts to maximize the network social welfare by trying different tolls on some roads and the lower level model describes the route choice behavior of network users. A Multi-class user equilibrium Frank-Wolfe algorithm is used for the lower level model and a differential quotient algorithm of bi-level planning model is employed to find the optimal solution after function evaluations. Finally, on basis of the above model and algorithm, the optimal toll area is computed its toll level is simultaneously designed.

Keywords: congestion pricing strategy, second-best multi-class based bi-level-planning, congestion pricing model, optimal toll area , flow design.

1. Model and Algorithms

This section examines the multi-class based road network equilibrium and system optimum problem in a network with a discrete set of Value of Time (VOT) for several user classes. The travel demand is subdivided into $|M|$ classes corresponding to groups of users with different socio-economic characteristics (for example, income level), where M is the set of all user classes. Let β_m $(\beta_m > 0)$ be the average VOT for users of class m , $\beta_1 \langle \beta_2 \langle ... \beta_m ... \beta_M$, $m \in M$; W be the set of all OD pairs w; d_w^m be the demand for travel of class m between O-D pair $w \in W$, $D_w^m(\mu_w^m) = \sum_{m \in M} d_w^m$ (m＝1,2,1….M), and $D_w^m(.)$ be the rigorously decreasing and continuous function；correspondingly $D_w^{-1}(d_w)$ be the inverse of the demand function. Let μ_w denote the equilibrium minimum travel cost between Origin-Destination (O-D) pair $w \in W$, including both travel time and monetary cost; P be set of all path p between OD pairs w, p∈P_w. The link

travel cost function $t_a(v_a)$ is only dependent on the total flow of this link, and it strictly increases, v_a^m is the flow in link a in class m; f_P^m is flow in path p in class m; $\delta(0,1)$ be Kroenecker delta, $\delta_a^p = \begin{Bmatrix} 1 & \textit{if link} \text{ a in path p} \\ 0 & \text{otherwise} \end{Bmatrix}$; c_{pw}^m is the class m travel cost function in path p of OD pairs w; g_w^m is flow in O-D pair w in class m. The symbol x_a represents the toll charged on link $a \in \bar{A}$, and $x_a^{\max}$ and $x_a^{\min}$ are the upper and lower bounds of x_a, $\mathrm{x}_{\mathrm{a}}^{\min} \le x_a \le x_a^{\max}, a \in \bar{A}$. The generalized link travel cost (including both travel time and toll charge if any) is given by $\bar{t}_a(v_a, x_a) = t_a(v_a) + x_a / \beta, a \in \bar{A}$, or simply $\tilde{t}_a(v_a, x_a) = t_a(v_a)$.
Second-best multi-class based bi-level planning pricing model is an optimal congestion pricing problem of different users under elastic demand. When the toll is more than the demand, the total travel time of road network will go to zero. Therefore, we should not only consider demand but also consider supply of pricing design under an elastic demand. We formulate the following second-best multi-class based bi-level planning model of road network congestion pricing for optimal toll design problem with elastic demand. The upper level is to maximize total network system social welfare, and the lower level is the user equilibrium with given O-D demands. The model is given below:

$$\text{Upper Model: } \max_{\mathrm{x}} F_2 = \sum_{m=1}^{M} \{ \int_0^{d_w(x)} D_w^{m^{-1}}(\omega) d\omega - \sum_{a \in A} \bar{t}_a(v_a^m(x)) \bullet v_a^m(x) \} \tag{1}$$

Here $v_a(\mathbf{x}), a \in A$ is the solution of the following lower-level model:

$$\text{Lower Model} \min_{\mathrm{d,w}} (\sum_{a \in A} \int_0^{v_a} \tilde{t}_a(\omega, x_a) d\omega + \sum_{a \in A} \sum_{m=1}^{M} \frac{1}{\beta^m} v_a^m x_a - \sum_{w \in W} \sum_{m=1}^{M} \int_0^{d_w} D_w^{-1}(\omega) d\omega) \tag{2}$$

$$\text{subject to: } \begin{aligned} & \sum_{\mathrm{p} \in p_w} f_{pw}^m = d_w^m, w \in W, m = 1,2,\cdots M \\ & v_a = \sum_{w \in W} \sum_{m=1}^{M} \sum_{P \in P_W} f_{pw}^m \delta_{ap}^w, a \in A \\ & v_a^m = \sum_{w \in W} \sum_{P \in P_W} f_{pw}^m \delta_{ap}^w, a \in A, m = 1,2,\cdots M \\ & f_p^w \ge 0, p \in p_w, w \in W, m = 1,2,\cdots,M \end{aligned} \tag{3}$$

In the above two models, since the equilibrium link flow $v_a(\mathbf{x})$, $a \in A$ and the O-D demand $d_w(\mathbf{x})$, $w \in W$ are generally non-convex and non-differentiable functions with respect to the toll pattern $\mathbf{x}$, it is very difficult to obtain the global optimum by traditional programming methods. For this type of problem, the probability based optimization methods are advantageous because they are able to find the global optimum without the requirement of differentiability. The Multi-class user equilibrium Frank-Wolfe algorithm of the lower level model and differential quotient algorithm of bi-level planning model are employed to find the optimal solution after function evaluations. The detailed algorithms are described below.

1.2 Multi-class user equilibrium Frank-Wolfe algorithm

Because the above lower model involves with multi class, the user equilibrium Frank-Wolfe algorithm modifies traditional Frank-Wolfe algorithm. The detailed steps for these calculations are below:

Step 1: Initialization For every user class, perform all-or-nothing assignment based on $\mathrm{t}_a^0 = t_a(0)$. Obtain a set of link flows $\{v_a^{m(1)}\}$ and OD flows $\{\mathrm{d}_{\mathrm{w}}^{\mathrm{m}(1)}\}$, Set $v_a^{(1)} = \sum_{m=1}^{M} v_a^{m(1)}, d_w^{(1)} = \sum_{m=1}^{M} d_w^{m(1)}$ and iteration counter n=1;

Step 2: Update Set $t_a^{(n)} = t_a(v_a^{(n-1)})$, $a \in A$; compute $D_w^{m-1}\left(d_w^{m(n)}\right)$;for every class user, set $t_a^{m(n)} = t_a^{(n)} + \beta_m x_a, a \in \overline{A}$;

Step 3: Direction finding Compute the shortest path S that the user spends travel time $\mathrm{C}_{\mathrm{w}}^{\mathrm{m(n)}}$ between each O-D pair w based on $\{\mathrm{t}_a^n\}$; execute the assignment rule shown in Eqs.(4).

If $C_{sw}^{m(n)} \le D_w^{m-1}(q_w^{m(n)})$, set $g_{sw}^{m(n)} = \bar{d}_w, g_{pw}^{m(n)} = 0, \forall p \ne s$, (4)

if $C_{sw}^{m(n)} > D_w^{m-1}(q_w^{m(n)})$, set $g_{pw}^{m(n)} = 0, \forall p$.

Once this assignment is complete, the auxiliary link-flow and OD trip-rate variables can be calculated as follows Eqs.(5): $u_a^{m(n)} = \sum_{w \in W} \sum_{p \in P} g_{pw}^{m(n)} \delta_{ap}^{w}$,

$$q_w^{m(n)} = \sum_{p \in P_w} g_{pw}^{m(n)}, \quad u_a^{(n)} = \sum_{m=1}^{M} u_a^{m(n)}, \quad q_w^{(n)} = \sum_{m=1}^{M} q_w^{m(n)} \qquad (5)$$

This assignment yields an auxiliary flow pattern $\{\mathrm{y}_{\mathrm{a}}^{\mathrm{n}}\}$ and OD flow pattern $\{q_a^{(n)}\}$.

Step 4: Move size Find optimal move size $\alpha_n, 0 \le \alpha_n \le 1$,that solves program (6),

$$\min_{\alpha^{(n)}} \{\sum_{a\in A} \int_0^{v_a^{(n)}+\alpha^{(n)}\left(u_a^{(n)}-v_a^{(n)}\right)} t_a(\omega)d\omega + \quad + \sum_{a\in A}\sum_{m=1}^{M} \beta_m \left[v_a^{m(n)} + \alpha^{(n)}\left(u_a^{m(n)} - v_a^{m(n)}\right)\right]x_a \tag{6}$$
$$- \sum_{w\in W}\sum_{m=1}^{M} \int_0^{d_w^{m(n)}+\alpha^{(n)}\left(q_w^{m(n)}-d_w^{m(n)}\right)} D_w^{m^{-1}}(\omega)d\omega\}$$

Step 5: Flow update Find $\left\{x_a^{n+1}\right\}$ and $\left\{q_w^{n+1}\right\}$ by using Eqs.(7).

$$v_a^{m(n+1)} = v_a^{m(n)} + \alpha^{(n)}\left(u_a^{m(n)} - v_a^{m(n)}\right),\ d_w^{m(n+1)} = d_w^{m(n)} + \alpha^{(n)}\left(q_w^{m(n)} - d_w^{m(n)}\right) \tag{7}$$

$$v_a^{(n+1)} = \sum_{m=1}^{M} v_a^{m(n+1)},\ d_w^{(n+1)} = \sum_{m=1}^{M} d_w^{m(n+1)}$$

Step 6:Convergence criterion If $\left\{x_a^{n+1}\right\}$satisfies following inequality (8):

$$\sum_{m=1}^{M}\left\{\sum_{w\in W}\frac{\left|D_w^{m^{-1}}\left(d_w^{m(n)}\right) - c_w^{m(n)}\right|}{c_w^{m(n)}} + \sum_{w}\frac{\left|u_w^{m(n)} - c_w^{m(n-1)}\right|}{c_w^{m(n)}}\right\} \le \varepsilon \tag{8}$$

terminate, $\left\{v_a^{n+1}\right\}$is the final equilibrium solution; Otherwise, set n: n=n+1,and go to step 2.

1.3 Differential quotient algorithm for bi-level planning model

Because the differential coefficient is the change in the independent variable, it can be regarded as the quotient of function differential to independent variable differential. The solution for the differential quotient of travel time function increment can be approximately regarded as the final solution.

Before finding the solution, we should separate urban transportation equilibrium network congestion pricing with several interactive and smaller problems. Every smaller problem only takes into account the condition. If the increment of travel time increase is Δy_a, the increment of travel time will influence on the OD demand, that is $q_{ij} = q_{ij}(y)$. If the increment of travel time in link a is Δy_a, the variable of O-D demand is a discretional link of all road networks and will be $\Delta q_{ij}(y)$. The differential quotient $\partial q_{ij}(y)/\partial y_a$ can be considered equally as $\Delta q_{ij}(y)/\Delta y_a$. All links repeat the above steps, and then we can complete computing all differential quotients, $\partial q_{ij}(y)/\partial y_a$.

Suppose y' is the initial increment of travel time after implementing congestion pricing, and $q_{ij}(y)$ is the corresponding OD demand variable that can be obtained from the lower model of the above second-best multi-class based bi-level planning model of road network congestion pricing for optimal toll design. This is represented as:

$$q_{ij}(y) \approx q_{ij}(y') + \sum_{a \in A} [\frac{\Delta q_{ij}(y)}{\Delta y_a}]_{y=y'} (y_a - y_{a'}) \qquad (9)$$

The upper model becomes a traditional linearity optimization problem based on the increment of travel time as the variable, and there are many mature algorithms to find its solution. The optimization solution got from the upper model returns to compute the lower model, and results in new OD demands. After repeating above calculations, the final optimal solution with convergence of a bi-level planning solution can been found. The detailed method is as follows:

Step 1: Compute the differential quotient $\Delta q_{ij}(y) / \Delta y_a$;

Step2: Set initial increment of travel time y^0, and set k=0;

Step3: Compute the lower model according to y^k and obtain OD demand q^k;

Step4: Compute Eq.(9), with the solution used to compute the upper model and get a group of new increment of travel time y^{k+1};

Step5: if $\max \left| y_a^{k+1} - y_a^k \right| \le \sigma, (\forall a \in A)$, terminate; Otherwise, set $k = k+1$ go to Step3.

1.4 Reconcile road network O-D pairs according to known link-flows

The goal of implementing congestion pricing strategy using road networks is to ensure O-D pair volumes are accurately forecasted. We should investigate all link-flows of pricing road networks and then reconcile road network O-D pairs according to known link-flows. If the toll of pricing-link is zero, we distribute hypothetic O-D pairs volume according above second-best multi-class based bi-level planning model of road network congestion pricing. If the difference of distributive results between known link-flows is minimal, the O-D pairs matrix is O-D choice matrix of road networks. Otherwise we should continue to compute until the distributive results converge.

2. Compute-flow design of determining optimal toll area and its toll level

After the congestion pricing model and its algorithm are determined, we design the computing flow of optimal toll-area and toll-level. The design consists of 4 steps: 1.) building the network and determining function index, 2.) Reconciling road network OD pairs according to known link-flows 3.) Determining all toll-area and toll-level values, and 4.) Computing the optimal toll-area and toll-level values of the network. A detailed explanation of each step follows:

Step1: build road network and determine model index: The node of the network represents the O-D volume and the link represents travel time for the main road corridor. The main index of network includes link capacity and the free-flow travel times of all links. The index of the bi-level model includes the numbers of travel user class, their VOT, and model elastic index. Next we input all indexes to a data file of the model program.

Step2: Determine road network OD pairs according to known link-flows: According to the Section 1.4 algorithm of inducting road network OD pairs based on known link-flows in peak time, we obtain O-D pairs in peak time. Then we input O-D pairs choice matrix to above data file of the model program.

Step3: Determine all toll-area and toll-level: We determine all toll-area sections composed of closed transportation corridors from small sections to big sections. At the same time, we determine all toll-levels from lower toll-level to upper toll-level.

Step4: Compute the optimal toll-area and toll-level of network: We compute the social welfare of the road networks in all O-D pairs of different toll-area sections on all toll-levels according to the Section 1.2 model and algorithm. Then, we compare the road network social welfare of all toll-areas under different toll-level and compute optimal toll-area and toll-level of the network by the key social total welfare index.

Conclusion

This paper puts forward a Second-best multi-class based bi-level planning model of road network congestion pricing in elastic demand, where the upper level program intends to maximize the network social welfare by trying different tolls on some roads, and the lower level program describes the route choice behavior of network users. A multi-class user equilibrium Frank-Wolfe algorithm was employed for the lower level model and a differential quotient algorithm of bi-level planning model was employed to find the optimal solution after function evaluations. The basis of above model determines the optimal toll area and its toll level simultaneously.

References

[1] Yao hongyun, The research of city congestion pricing, Tongji university PHD thesis;

[2] Xiaoning Zhang *Optimal road pricing in transportation network.* Thesis for the Ph.D. Degree in Hong Kong University of Science and Technology 2003.

Optimization of Production Scheduling in Container Terminals using Computer Simulation

ZHANG Jie-shu ; SHA Mei; EASA Said
Management Department of Shanghai Maritime University
Shanghai, China 200135
Email: jszhang@cct.shmtu.edu.cn

Abstract. Scheduling of operational activities is a key step in port management. Port operation scheduling in China is normally based on empirical procedures that result in long ship berthing time, uneven distribution of duty cycle of machinery, and serious yard-block congestion. In view of this, this paper presents a simulation model for scheduling container terminals. The model provides information on equipment and port utilization as well as operational processes, thus providing managers with a valuable decision support tool for port operation scheduling.

Keywords: Operation scheduling, simulation model, container terminals

1. Application of Computer Simulation Technology in Port System

In recent years, an increasing number of simulation models have been used to provide decision support for port planning and operation management. The application of computer simulation technology in port systems is becoming more and more important and popular. For example, in Hong Kong Lai and Leung (2000) developed three expansion policies to improve the current gate operations and set up a facility expansion strategy. A simulation model is used to evaluate these policies under different demand growth rates. In the United States Das and Spasovic (2003) have presented a straddle scheduling procedure that can be used by a terminal scheduler to control the movement of straddle carriers. The objective of the terminal scheduler is to minimize the empty travel of straddle carriers, while at the same time minimize any delays in servicing the customers. The paper presents a simulation model for operation scheduling in container terminals using a real system. Using this model, the superiority of the proposed procedure over two alternative scheduling strategies is illustrated

2. Development of Simulation Model

2.1 Model Boundaries

The proposed model simulates the entire operation processes of a single ship, from loading to unloading after berthing. It provides the simulation operation for different plans under different equipment arrangement and operation organization based on specified ship distribution and the yard piling mode. By comparing a series of performance indices, the operation management personnel can evaluate various scheduling scenarios and select the best plan for the container terminal.

2.2 Main Input and Output Parameters

The port operation scheduling simulation system is based on a fundamental principle. With the development of the simulation model and the guide of its decision rules, the whole port operation process can be evaluated. In this system, the changes of state are recorded in order to collect, manage, and produce the best scheduling plan and operation result statistics, thus providing a valuable decision-support system for port scheduling managers.
The input statistics of the scheduling simulation system can be classified into three types: model parameters, controls parameters, and decision rules. The performance output results include two types: scheduling plan and operating indices (Figure 1). The following are brief descriptions of these input and output types:

(a) *Model Parameters*. Model parameter refers to the information applied to determining simulation model for scheduling, which can be input and modified by port scheduling managers. The main parameters are shown in Table 1.
(b)*Control Parameters*. Controls parameter provides necessary functions and mechanism for operation and experiment of simulation system and can influence the selection of simulation process, as shown in Table 2.
(c) Decision Rules. Decision rules are the basis of the operation of the simulation system. The model analyzes a single container or a single machine, and it must simulate the operating process exactly. Therefore, based on the scheduling plan input by terminal operators, some basic decision rules about the operation features of container terminals have to be summarized. For example, the decision rules for the operating process of unloading are shown in Table 3.
(d) *Performance Output Results*. The scheduling plan is based on the specified order of equipment and operation scheduling, which is drawn after being input by the users. The performance parameter is a series of indices that can reflect the effect of scheduling. The main performance parameters of the system for each operation process are shown in Table 4.

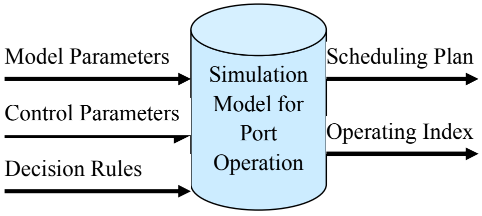


Figure 1 Basic principle of simulation model for port operation

Table 1 Main model parameters

Operation Process	Model Parameters	Variables
Ship Arrival	- Arrangement of ship berthing	- By left or by right
Loading/ Unloading	- Arrangement of operating lines, - Types of machinery at dock apron - Sequence of operation processes	- Determined by port scheduling managers - Ship-to-shore gantry crane or overhead crane - Determined by port scheduling managers
Horizontal Transport	- Plan of horizontal transport - Arrangement of operating lines - Arrangement of operating machinery	- 'fixed distribution 'or 'mixed distribution' - Determined by Port scheduling managers - Truck or trailer
Storage Yard	- Machinery distribution in stock yard	- Rubber tire gantry crane(RTG) or forklift

Table 2 Controls parameters

Controls Parameters	Explanations
- Selection of simulation step length - Time interval for counting loading and unloading - Selection of digital simulation or graphic simulation	- Counted by minute - Counted by hour - Digital simulation and/or graphic simulation can be chosen

Table 3 Decision rules of the operating process of unloading

Name	Explanations
Principles of deciding operation process of unloading	- Different operating lines unload at the same time - One operating line unloads at Operation Sequence 1 according to the input scheduling plan - The corresponding container of an operating line unloads from land side to ship side, from the top down - According to the user input scheduling plan, containers are unloaded at the same time from land side to ship side and from the top to down, while there are two containers at the same deck and row of neighboring bay, they should be unloaded simultaneously, otherwise, only one container can be unloaded.

Table 4 Performance parameters

Operation Processes	Parameter Indices
Loading and Unloading	- Total cargoes of each operating line - Comparison of cargoes of each operating line per hour - Operating time of each operating line - Operating efficiency of each operating line
Horizontal Transport	- Operating time of each machine of each operating line - Operating efficiency of each machine of each operating line - Operating cargoes of each machine of each operating line and their total
Storage Yard	- Stocking area where congestion occurred in loading/unloading - Frequency of congestion - Duration of congestion - Cargo areas that mold turnover occurs - Times of mold turnover

2.3 General Structure of Simulation Model

The general structure of the developed system is shown in Figure 2. The subsystem "Extraction of Basic Statistics" is applied to extracting the basic operating information from the operating statistics of container terminals. The storage situation of the container yard and the information about ship loading and unloading that requires simulation scheduling, including volume, types of containers, and stowed location. The subsystem "Definition of Parameters" provides the users with the option to select different parameters and the output interface. This subsystem includes the arrangement of operating lines, operating sequences, and selection of machines, whose design is made according to the operating features of the container terminals and trade practice.

In accordance with the user input parameters and control rules, the subsystem "Creation of Scheduling Plan" produces the operating sequence of all the containers. The subsystem "Simulation Operation" simulates the operating processes of each container, records relative statistics, and calculates the indices. The subsystem makes use of real-time animation which synchronizes with the animation system. All the statistics are entered in "Background Database" for processing, and all the changes in the operating process of the system can be shown simultaneously which can directly reflect the dynamic features of the operation of scheduling plan. The users can certainly use the "Digital Simulation Output" subsystem to view the corresponding parameter index of each scheduling plan to further compare the plans and select a favorable scheduling scenario.

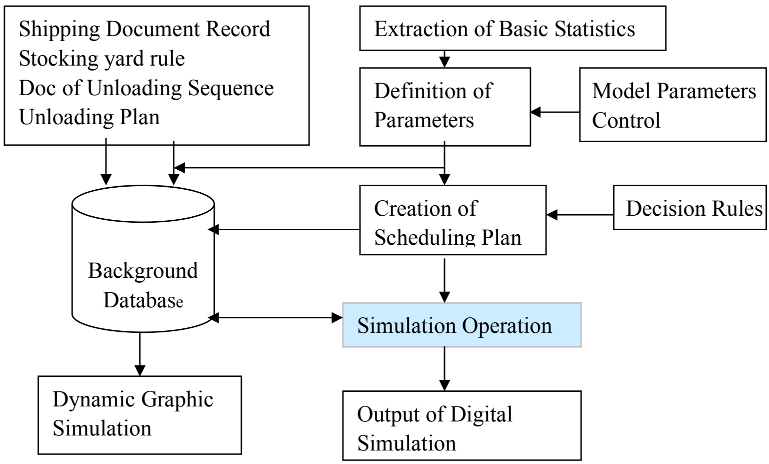


Figure 2 General design of simulation system for container terminal scheduling

3. Optimization and Selection of Scheduling Plans

The optimization and selection of the scheduling plan is a key function of the developed "Simulation Model for Container Terminal Scheduling." The model allows port operation managers to create different scheduling plans by changing such parameters as operating lines, operating sequences, type of operating machine, and distribution of machinery. By comparing the output indices of different simulated plans, a superior scheduling plan can be determined. Thus, the ship time in the port can be reduced and the service can be improved. To illustrate the process of selecting and optimizing scheduling plans, a specific ship (code E0010039) is used as an example.

The ship should be scheduled first to create Plan 1 (the scheduling at the dock apron in this plan is shown in Table 5). The results of the scheduling show that the latency time of the dock apron machine is long, with an average of above 10 hours and a maximum of 11 hours. The ship time in the port (the time from the beginning of the operation to the completion of the last line) is 45 hours which is also rather long. With the analysis of Plan 1, it can be noted that the volume of the trucks equipped for each operating line is relatively low, which may have resulted in the long latency time of the bridge crane. Therefore, while the other input parameters are the same, the volume of trucks equipped for each operating line increases to three in Plan 2 and

Table 5 Sample results of the simulation output [a]

Plan Code Number	Main Distributions of Plans		Main Output Parameters of Plans		
	Number of Operating Lines for a Ship	Number of Horizontal Transport Equipment for the Line	Working Efficiency of DAE [b] (TEU/h)	Longest Waiting Time of DAE (hours)	Ship Berthing Time in Port (hours)
Plan 1	3	2	9.9	11:10	45:28
Plan 2	3	3	14.6	5:34	31:56
Plan 3	3	4	18.27	3:36	26:32

[a] The type of dock apron equipment for all plans is ship-to-shore gantry crane
[b] Dock apron equipment (DAE)

to four in Plan 3, which results in a new input index. It can be noted in Plan 3 that the ship time in the port has been further reduced and the working efficiency of dock apron equipment is further improved. As a result, after the arrival of the ship, Plan 3 can be implemented for scheduling in order to reduce the ship time in port.

4. Conclusions

The proposed "Simulation System for Container Terminal Scheduling" makes use of the animated graphic simulation technology to provide a decision-support system for optimizing container terminal scheduling plans. The simulation output can prove to be valuable in the following important aspects: (a) selection of ship operation plans, horizontal transport organization plans, and storage yard operating plans, (b) identification of appropriate ratio of quay transport, horizontal transport, and yard transport machines, (c) effect analysis of ratio change in quay transport, horizontal transport, and yard transport machines, and (d) selection of appropriate solutions for emergency situations. The system has been implemented in the Waigaoqiao Dock in Shanghai Port. The methodology and techniques of the proposed system have been proven to be widely applicable to optimizing the operation scheduling of other container terminals.

References

1. Lai, K.K. and Leung, J. Analysis of gate house operations in a container terminal. *International Journal of Modelling and Simulation*, 2000, 20(1): 89-94.
2. Das, S.K., and Spasovic, L. Scheduling material handling vehicles in a container terminal. *Production Planning and Control,* 2003 14(7): 623-633.

Optimization Model for Locating Service Stations along the Freeway

Wei WANG[1], Shunxin YANG[2], and Xia Jin[3]

[1]Senior student of Transportation College, Southeast University, No.2, Sipailou, 210096, Nanjing, P. R. China; email:ww.movealong@gmail.com
[2]Lecturer of Transportation College, Southeast University, No.2, Sipailou, 210096, Nanjing, P. R. China; email:shunxin.yang@gmail.com
[3]Modeler, Cambridge Systematics, Inc; New York Metropolitan Transportation Council 199 Water St, 22nd Fl; New York, NY 10038; xjin@dot.state.ny.us

Abstract: The high economic cost and social impact associated with freeway incidents has prompted Traffic Management Agencies world-wide to develop incident management system (IMS) aiming to alleviate incident consequences. In order to reduce the response time and clear the incidents quickly and safely, an efficient and effective deployment of incident response resources, especially the service stations, is critical to IMS. The objective of this paper is to develop a tool to search for optimal locations of service stations along the freeway, to minimize the response time from the service stations to the incident sites. A dynamic programming approach is presented in this paper, followed by a case study to demonstrate the application of this optimization model in accordance with the status quo of China. With 20% reduction in total travel time, the model shows good potential in improving the incident response time.

1. Introduction

With the rapid development of freeway constructions and the fast growth of traffic volume in China, the incident rates have increased tremendously. Moreover, there is also a symbiotic relationship between traffic incidents and congestions. As congestion is one of the main reasons for traffic incidents, traffic incidents in turn cause more congestion. This vicious cycle is a major problem that threatens the mobility and the safety of transport in China. Increasing attention has been drawing to the development of a well-organized and coordinated incident management system.

The primary objective of any IMS is to minimize the incident duration time. In general, the duration time is defined as the time slice between the occurrence of an incident and the clearance of the incident as well as the restoration of the freeway capacity to its normal level (Kaan Ozbay 1999). More specifically, the freeway incident duration time can be divided into three components (K. G. Zografos 2002):

(1) Detection time: This is the time interval from the occurrence of an incident until the incident is detected and verified, and it is dependent on the detection systems such as emergency phones, patrolling vehicles, detectors and so on.

(2) Response time of service stations: This is the time interval between the incident verification and the arrival of service personnel and equipments at the scene of the incident.

(3) On-site service time: This is the time required by the dispatched service stations to clear the incident and restore the freeway capacity to its normal level. This time is contingent on the severity of the freeway incident.

Compared with the other two duration time components, the response time is controllable to some extent, and can be greatly improved as long as the response resources are well planned and located. Given the though above, the focus of this paper is on searching for the optimal locations of service stations so that the response time could be minimized accordingly.

Optimizing the locations of service stations along the freeway to reduce the response time has been proposed by numerous researchers using operational research approach since the middle 1970's. Chow, We-Min (1976) interpreted the stationary emergency service system as an ordinary queuing system and dynamic programming was used to determine the best locations for service facilities, the optimal number of service vehicles at each service location and the service area covered by an individual service location. Aly, Adel A (1978) assumed that the location of incidents were random variables, and the probability distribution for rectilinear travel time between a new facility location and the random location of the incident was developed for the case of being uniformly distributed over a rectangular region. Neebe, Alan W (1988) considered the problem of locating emergency-service facilities as assigning a set of demand points to a set of facilities. Berman, Oded (2003) indicated that there was a probability depending on the distance from the service facility, that the facility may not be able to provide satisfactory service to a demand point.

The objective of this paper is to search for the best locations of service stations along the freeway, so that their response time would be reduced. The remainder of this paper is organized as follows: Section 2 describes the current situations of freeway service stations in China, Section 3 presents an optimization model, through which, the best locations of service stations would be found, Section 4 presents a case study using field data from the northern part of Nanjing-Lianyungang freeway in China, and finally Section 5 summaries the conclusions of this paper and identifies the possible future research areas as well.

2. Description of the Current Situations of Freeway Service Stations in China

As mentioned above, incidents cause bottlenecks on the freeway, delay the traffic and increase the risk of a crash. In order to reduce the impacts on the freeway flow, time is essential. Equally important is the locations for the service stations, since well-located service stations would reduce the total response time to a great extent. To take a typical example, currently, there are many service stations along the northern part of Nanjing-Lianyungang freeway with their exact locations represented by the number of stake. And the service area covered by each individual service location has been determined as well, as demonstrated in figure 1.

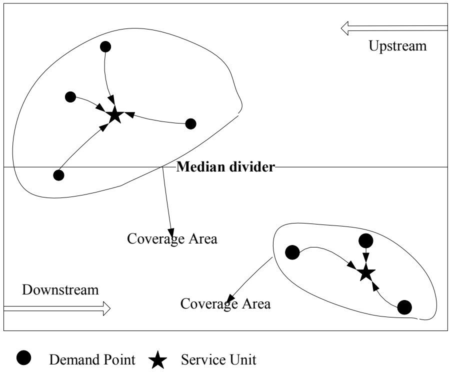


Figure 1. A schematic representation of current service stations and demand points.

By examining the locations and resource allocation of the service stations, we find that most of the stations are not rationally located and the resources are not properly allocated. Particularly, in some roadway sections with high incidents rate, the nearest service station is still far away and the quantity and quality of service facilities are far below the needs. Therefore, it is urgent to optimize the locations of service stations and reallocate the service facilities so that a safe and timely clearance of incidents can be ensured. Consequently the traffic delay and the recovery time to normal flow condition can be greatly reduced.

3. Optimization Model

Since the freeway demand changes from time to time, the same type of incidents occurred at the same location may cause different results if the incidents occurred at different time epochs (We-Min Chow and Adolf D.May 1972). To be specific, the effect on traffic flow by traffic incidents during the peak hours can be considerably severer than that during the off-peak hours. Logically, peak hour travel time skim should be utilized in the optimization model. In other words, if the service stations are located under peak hour condition, then they would certainly satisfy the constraints for other time epochs.

3.1 Notation and mathematical model

To formulate this model, we divide the freeway into many subsections so that each subsection can be treated as a uniform pipe, i.e., capacity is constant over the length of each subsection (We-Min Chow and Adolf D.May 1972). Besides, m potential locations for service stations are preselected for optimization.

Then the problem is to investigate the m potential locations for the service stations and to find the optimal set with minimum response times. The response time consists of two components, one is the dispatching time, and the other is the travel time from the service stations to the incident scene. Since the dispatching time is a function of the arrival pattern of the incidents, the type of agencies involved in the incident management operations, and the dispatching policy adopted by each of the agencies, it is beyond the scope of this study. So the focus in our problem is the travel time of service units. The mathematical formulation of this model is presented as:

$$Minimise \sum_{i}^{n} \sum_{j}^{m} t_{ij} x_{ij}$$

Where n denotes the number of subsections along a given freeway; m denotes the number of service stations along the freeway; t_{ij} denotes the shortest travel time from the service station j to the traffic incident scene in the subsection i; x_{ij}=1, when service station j is assigned to incident occurred in subsection i, 0 otherwise.

3.2 Dynamic programming approach

This type of optimization problem can be solved in several different ways. We use the dynamic programming approach, which is better suited for manual computations and requires less computation efforts (We-Min Chow and Adolf D.May 1972). Note that in order to convert the network of subsections and service station locations into a matrix form, we have to set certain upper limits for response time, and connect the service station locations with the subsections without violating the preselected limit.

For this approach, it is assumed that the incident spot is located at the mid point of the subsection and t_{ij} is the shortest travel time during peak hours from each of the preselected station locations to incident spots, with the travel route under the rule of typical “shortest path problems”.

4. A Case Study

In this section, the specific application of this optimization model is illustrated using field data from the northern part of Nanjing-Lianyungang freeway, which totals 138 kilometers and is divided into 4 subsections with a constant capacity over the length of each subsection, i.e., K204-K207, K207-K254, K254-K292, K292-K317, with the origin and destination of each subsection represented by the number of stake.

First, set the upper limit for the response time, called T. In this case, based on the statistical analysis of the historical data, July has the highest incident rate, and the maximal response time of this month during peak hours is 30 minutes, therefore, we set T= 30 minutes. Further, according to the records of the incident report, we find 4 locations along the northern part of Nanjing-Lianyungang freeway, where incidents frequently occurred. Here, we find these locations represented by the stake number as K207, K243, K254 and K274. So we set 4 potential locations for service stations around these 4 locations and each of them is located near the on-ramp. The traveling speed is assumed to be 68 kilometers per hour during peak hours, and therefore the travel time (length/speed) is ready to be evaluated. The estimated travel time from each potential service station to each incident spot along this freeway is presented in table 1.

Table 1. Travel time from potential service station x to the incident spot in subsection s during the most critical time period (in minutes)

x \ s	1	2	3	4
1	1.3	20	64	86
2	33.1	11	26.5	54
3	60.4	38.3	0.9	27
4	84.2	62.2	24.7	3

Next is to search for the best locations for service stations in terms of minimizing the total travel time.

For the dynamic programming approach, to prepare the incident matrix, a "1" is put into the cell (i, j) if the time required by service unit traveling from station i to the location of incidents occurred in the subsection j during peak hours is less than the preset time T (i.e., 30 minutes). Figure 2.demonstrates the initial connections between each service station and subsection.

		Subsections j			
		1	2	3	4
Locations i	1	1	1		
	2	1	1	1	
	3			1	1
	4			1	1

Figure 2. Incident matrix.

The problem is solved by searching for all minimal set solutions and one column is considered at a time. Note that a minimal set solution mentioned here is a set such that no proper subset of this set can be a solution. To this problem,

(1) Find the minimal set solutions for column 1 and treat these solutions as the current partial solutions. In this case, it is $\{1\}$.
(2) Find the minimal set solution for column 2: $\{1\},\{2\}$.
(3) Combine the partial solutions (1) and solutions just obtained (2) to find all the minimal set solutions: $\{1\}\cup\{1\},\{1\}\cup\{2\}$, then the current partial solution is $\{1\}$, since $\{1\}\subset\{1,2\}$. Treat this result as the current solution.
(4) In column 3, the current partial solutions: $\{1,2\},\{1,3\},\{1,4\}$.
(5) At the final stage, the complete solutions are obtained: $\{1,3\},\{1.4\}$, which means that the service stations should be established at locations 1 and 3, or 1 and 4.

Very importantly, another advantage of this approach is that when the preset limit changes, the post optimality problem is easy to handle for the simple reason that any minimal set solutions generated from the new column can satisfy the original column.

5. Conclusions and Future Research

A key component of an effective incident management system is to properly allocate service stations and resources so that the incidents can be cleared safely and timely. In this paper, using field data from the northern part of Nanjing-Lianyungang freeway, we employ the dynamic programming approach to search for the optimal locations along this freeway to minimize the total response travel time. The optimized locations are tested using historical incident report data along this freeway. The total travel time can be reduced by 20%, which indicates a good capability of the model in improving the response time.

Being related to the present work, two possible future research areas have been identified. First, the mathematical model proposed for minimizing the total travel time using dynamic programming approach is hard to solve for large scale networks. To make it more practicable for real world application, incorporating heuristic approaches like GA heuristic into this model may have a good potential. Second, although the best locations of service stations have been identified, the optimal number of service vehicles at each service location is not considered. In this paper, we assume that each incident demand point is served by only one service station, while in fact, two or more service stations may be required for a serious incident. Therefore, by taking into count different types of incidents, the minimum number of service stations and service equipment required by such incidents needs further study.

References

Aly, Adel A. and White, John A. (1978). “Probabilistic Formulation of the Emergency Service Location Problem”. *Journal of the Operation Research Society*. pp. 1167-1179.

Berman, Oded. , Drezner, Zvi and Wesolowsky, George O. (2003). “Locating Service Facilities Whose Reliability is Distance Dependent”. *Computers and Operations Research.* pp.1683-1695.

Kaan Ozbay, and Pushkin Kachroo. (1999). “Incident Management in Intelligent Transportation Systems”. *Artech House, Inc*. pp. 6-15.

K. G. Zografos, Konstantinos N. Androutsopoulos and George M.Vasilakis. (2002). “A Real-time Decision Support System for Roadway Network Incident Response Logistics”. *Transportation Research Part C* 10. pp.1-18.

Neebe, Alan W. (1988). “Procedure for Locating Emergency Service Facilities for All Possible Response Distances”. *Journal of the Operational Research Society*. pp.743-748.

We-Min Chow and Adolf D.May. (1972). “Searching for the Best Locations of Service Facilities along the Freeway”. *Research Report of University of California, Berkeley*. pp. 1-26.

We-Min Chow. (1976). “Optimization Model for Freeway Stationary Emergency Service Systems”. *Transportation Science*. pp.20-34.

A Method of Predicting NO_2 Hourly Concentrations near City Arteries Based on BP Neural Network

Min Xie, Ming Cai, Zhi Yu, and Weijia Xu

ITS Research Center, Sun Yat-sen University, Guangzhou Guangdong 510275, China; PH (8620) 84112638; FAX (8620) 84113689; email: caiming@mail.sysu.edu.cn

Abstract

In this paper, a method based on BP neural network is proposed to predict the hourly average concentrations of NO_2 on roadside near city arteries. The main factors that affect the hourly concentrations of NO_2 include O_3 concentrations, NO concentrations, traffic volumes and meteorological factors which contain atmospheric temperature, atmospheric pressure, wind speed, wind direction, relative humidity, rainfall and solar radiation. The hourly average data of these affecting factors were selected as the input neurons of the network. And the network was trained and validated with the data form an automatic monitoring experiment on roadside. As a comparison, a multiple linear regression model was established. When predicting NO_2 concentrations with the neural network approach, the mean absolute value of relative error of training and validation result was 7.19%, and the correlation coefficient between the predicted and monitored concentrations was 0.99, which was much better than the prediction with a multiple linear regression model.

Keywords: *NO_2; Back Propagation Neural Network; Prediction; Affecting Factors*

1 Introduction

Vehicular exhaust emission has been one of the main sources of urban air pollution. The high concentration of harmful air pollutant, such as NO_2, will threaten people's health, thus, it's significant to perform the research on the method of predicting air pollution in the public places. The underlying relationship for air pollution concentrations versus meteorology and vehicular emissions is complex and strongly non-linear, which makes the traditional deterministic techniques not particularly appropriate. Artificial Neural Network (ANN) is especially suitable for the research on the objects with the characteristics of multifactor, uncertainty, randomness and nonlinearity (Gardener and Dorling, 1999; Kukkonen, *et al.*, 2003). Hence, ANN is widely used on air pollution prediction for its powerful predicting capability, especially when Back Propagation (BP) algorithm is involved.

There are many applications of ANN technique on predicting the concentrations of air pollutants, such as O_3 (Elkamel, *et al.*, 2001), PM_{10} (Kukkonen, *et al.*, 2003) and CO (Nagendra and Khare, 2004). Meanwhile, there are a few applications on predicting the concentrations of traffic-related pollutants. Perez and Trier (2001) predicted the NO and NO_2 concentrations at noon near a street with heavy traffic in Santiago, Chile. Nagendra and Khare (2006) established an ANN based dispersion model of NO_2 to predict the 24-hour average NO_2 concentrations.

Different to the previous work, the input of network in this paper contained traffic volumes and the concentrations of NO and O_3, and the network was used to predict NO_2 concentrations of any hour. In the present work, a monitoring experiment on the roadside of Xingangxi Road in Guangzhou was conducted to obtain the samples of air pollution data, meteorological data and traffic flow data. And the prediction method of NO_2 hourly concentrations near an artery in Guangzhou based on BP neural network was presented and compared with a multiple linear regression model.

2 Data collection

The monitoring sites were located near Kanglecun bus station and Yilecun bus station on both sides of Xingangxi Road in Guangzhou. Xingangxi Road is a transmeridional artery in Haizhu District with 8 standardized lanes bidirectionally and a traffic volume of 4000 to 5000 vehicles per hour in daytime. The monitoring sites were about 5 meters off the road with a height of 8 meters. Eliminating the time for calibrating the instruments, the valid monitoring periods were from Jan 16th to Jan 22nd and from Jan 28th to Feb 3rd, in 2007. The automatic monitoring items included the real-time concentrations of NO, NO_2 and O_3, and the values of the 7 meteorological parameters which contained atmospheric temperature, atmospheric pressure, wind speed, wind direction, solar radiation, rainfall and relative humidity. The traffic flow of Xingangxi Road was record by video, and counted manually for the hourly traffic volumes. The concentrations of NO, NO_2 and O_3 were measured by an vehicular automatic air quality monitoring system consisting of a NO-NO_2-NO_x Analyzer, a UV Photometric O_3 Analyzer, a Dynamic Gas Calibrator and a Zero Air Supply. Details of the instruments for each parameter were shown in Table 1.

Table 1. Details of the monitoring instruments

Parameter	Reaction time	Accuracy	Manufacturer
NO, NO_2, NO_x	20s	1.0ppb	Thermo Electron, USA
O_3	20s	1.0ppb	Thermo Electron, USA
Wind speed	1s	±0.15m/s	MetOne, USA
Wind direction	1s	±5°	MetOne, USA
Relative humidity	1s	±5%	MetOne, USA
Temperature	1s	±0.5℃	MetOne, USA
Air pressure	1s	±0.04%	MetOne, USA
Rainfall	3s	±1%	MetOne, USA

3 Establishment of prediction model

3.1 Basis of BP neural network

As is known, Artificial Neural Networks are parallel calculation architectures whose structure is based on the human brain. ANN is extensively interconnected by a large amount of simple processing elements, known as neurons. Neurons are arranged in parallel layers connecting to the other layers with the weights. In the middle of the network, there is at least one hidden layer which acts as feature detector. The input neurons transmit their inputs to the neurons in the hidden layer. The outputs of the neurons in the hidden and output layers are calculated by transforming their inputs using a non-linear transfer function (Elkamel, *et al.*, 2001), such as the widely-used 'S'-shaped logistic sigmoid transfer function which was also used in this paper. The term back-propagation (BP) refers to the fact that the synaptic weights are all adjusted in accordance with an error signal, which is propagated backward through the network against the direction of synaptic connections.

3.2 Selection of the affecting factors

The dispersion and concentration distribution of air pollution are directly affected by meteorology. And atmospheric temperature, atmospheric pressure, relative humidity, wind speed, wind direction, rainfall and solar radiation are routine meteorological parameters that can be measured by almost every meteorological station. Thus, these variables were selected as the affecting factors of NO_2 concentration prediction. NO is a product of vehicle combustion. It can slowly oxidize to NO_2. The oxidation efficiency will rise when oxidation accelerator (*e.g.*, O_3) exist in the air. NO and O_3 play an important role in photochemical reactions which lead to the formation of secondary pollution of NO_2. Therefore, NO and O_3 concentrations were also selected as affecting factors to predict NO_2 concentrations. In addition, the vehicle emission is the source of NO and NO_2, and some researches indicated that traffic flow was also an important factor when predicting NO_2 concentrations (Nagendra and Khare, 2006). In this work, the vehicles were divided into light-duty and heavy-duty vehicles. According to the emission genes of light-duty and heavy-duty vehicles in Guangzhou (Zhu, 1997), the equivalent hourly traffic volumes were computed and used as the so-called traffic volumes. Therefore, 10 variables were selected as affecting factors of the prediction.

3.3 Determination and normalization of the samples

The affecting factors were used as input neurons of the neural network, and NO_2 concentration was the output neuron. The data of these parameters were used to train and test the network. The valid dataset contained 306 hourly average data from 0:00, Jan 17th to 21:00, Jan 22nd and from 0:00, Jan 28th to 20:00, Feb 3rd, in 2007. 278 data among them were selected randomly as training data, and the remaining 28 data were validation data. To avoid overflows of network due to very large or small the weights and eliminate the influence of different dimensions of data, the input of neural network had to be normalized (Gardener and Dorling, 1999; Elkamel, *et al.*, 2001). In this paper, the data were normalized into the range [0, 1] with the formula: $X_{norm}=(X_i-X_{min})/(X_{max}-X_{min})$, where X_{norm} was the normalized value, X_i was the original value, and X_{min} and X_{max} were the minimum and maximum values of X_i.

3.4 Establishment and training of the neural network

Having a sufficiently large number of neurons, a neural network with a single hidden layer can map any input to any output with an arbitrary degree of accuracy (Elkamel, *et al.*, 2001), thus the network with one hidden layer was used in this work. The number of hidden neurons was decided empirically, and differed in particular instances. Whereas, with few hidden neurons, the anticipant accuracy could not be reached, while too many hidden neurons would cause the 'overfitting' or 'overtraining' phenomenon of the input data, and reduced the generalization capability of network. In this work, it was found that when the number of hidden neurons was twice of the input neurons plus one, the network gave the best training effect, and the training can be stopped within the allowed error and the maximum training times of 10000.

The 10 affecting factors confirmed were split into 3 groups: the 7 meteorological parameters, traffic volume and the concentrations of O_3 and NO. Using the different combination of these groups as the input of the network, the training result was shown in Table 2 (MET was short for the 7 meteorological parameters, and Relative error was for average absolute value of relative error in this paper). By comparing the relative errors and the correlation coefficients of the predicted values and the actual values, the best set of input neurons were meteorological parameters, the concentrations of O_3 and NO and traffic volume, because of the minimal error and maximal correlation coefficient. Figure 1 shows a plot of predicted and actual values.

Table 2. Comparison of training effect with different input neurons

Input neurons	Best number of hidden neurons	Relative error	Correlation coefficient
MET	15	18.05%	0.939
MET, traffic volume	17	13.61%	0.958
MET, O_3 and NO	19	9.55%	0.957
MET, O_3, NO, traffic volume	21	6.67%	0.993

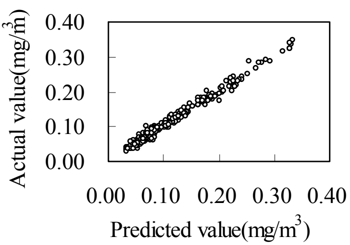


Figure 1. Plot of the actual and predicted NO_2 value with the training data.

4 Analysis of prediction effect

4.1 Predicting NO_2 hourly concentrations with the neural network

To check the generalization characteristics of the neural network with the new validation data is an important step to analyze the prediction effect of the network prepared. The 28 validation data, which were defined in section 3.3 with the meteorological parameters, traffic volumes and the concentrations of O_3 and NO, were put into the network, and the relative error for the corresponding NO_2 predicted values was 12.4%, while the correlation coefficient of the predicted and actual values was 0.968. This indicated that the generalization ability of the network was reasonable and it was feasible to predict the NO_2 hourly concentrations with neural network approach.

In Figure 2(a), a comparison of actual and predicted values of NO_2 with neural network approach was shown, and the data contained the training and validation data of 306 hours. The relative error for the predicted values was 7.19%, and the correlation coefficient of the predicted and actual values was 0.990.

4.2 Comparison with a multiple linear regression model

A multiple linear regression function was established with a dependent variable of NO_2 concentration and 10 independent variables of the meteorological parameters, traffic volume and the concentrations of O_3 and NO. The samples of this regression model were the training data used in the neural network, and the function was NO_2=0.002P-0.0000218WD-0.022WS+0.012T-0.001RAIN+0RH-0.000042SR-0.0000008V+0.176NO-1.157O_3-2.098, where P was air pressure, WD was wind direction, WS was wind speed, T was temperature, RH was relative humidity, RAIN was rainfall, SR was solar radiation, and V was traffic volume. The multiple correlation coefficient was 0.930, and the relative error was 20.7%.

According to the regression analysis, the regression function was significant, while the 6 regression coefficients of P, WD, RAIN, RH, V and the constant item were not significant, which indicated that these variables could be eliminated. Once this was done, the fitting effect was reduced with the multiple correlation coefficient of 0.915 and the relative error of 20.8%. This indicated that multiple linear regression model could not reflect the effect of all the factors comprehensively.

By inserting the 28 validation data into the function above, the predicted NO_2 values were obtained with the relative error of 31.35% and the correlation coefficient of 0.895. The error was more than twice of which from the neural network approach in section 4.1, and the correlation coefficient to actual values was also smaller.

Figure 2(b) gave a comparison of actual and predicted values of NO_2 with a multiple linear regression model, and the data contained the training and validation data. The average error for the predicted values was 21.71%, and the correlation coefficient of the predicted and actual values was 0.912.

5 Conclusions

The prediction method of roadside NO_2 concentrations based on BP neural network prepared in this paper was effective and performed obviously better than a multiple linear regression model not only in the fitting degree and generalization ability, but also in reflecting the effect of all the factors to NO_2 concentrations comprehensively. The meteorological parameters, traffic volume, and the concentrations of O_3 and NO were the 3 groups of factors that played important roles in predicting NO_2 hourly concentrations. And the affecting rules of them could be successfully achieved by neural network. In addition, due to the limited quantity of training samples, the affecting rules of those factors could not be represented completely, and that caused a limitation of prediction accuracy. Still, by enlarging the training data set and increasing the generalization ability of the network, a better prediction result could be achieved.

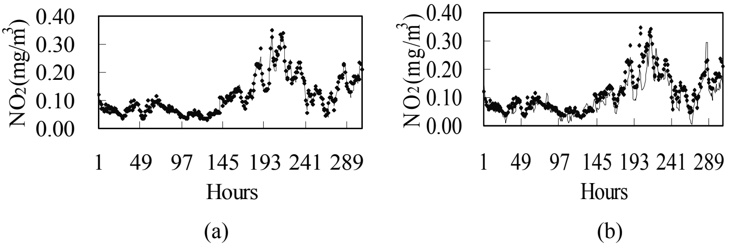


Figure 2. Comparison between actual (points) and predicted (line) NO_2 values with (a) the neural network and (b) multiple linear regression model.

References

Elkamel, A., *et al.* (2001). "Measurement and prediction of ozone levels around a heavily industrialized area: a neural network approach." *Advances in Environmental Research*, 5(1), 47-59.

Gardner, M.W. and Dorling, S.R. (1999). "Neural network modelling and prediction of hourly NO_x and NO_2 concentrations in urban air in London." *Atmospheric Environment*, 33 (5), 709-719.

Kukkonen, J.,*et al.* (2003). "Extensive evaluation of neural network models for the prediction of NO_2 and PM_{10} concentrations, compared with a deterministic modelling system and measurements in central Helsinki." *Atmospheric Environment*, 37 (32), 4539-4550.

Nagendra, S.M.S. and Khare, M. (2004). "Artificial neural network based line source models for vehicular exhaust emission predictions of an urban roadway." *Transportation Research, Part D: Transport and Environment*. 9 (3), 199-208.

Nagendra, S.M.S. and Khare, M. (2006). "Artificial neural network approach for modelling nitrogen dioxide dispersion from vehicular exhaust emissions." *Ecological Modelling*, 190(1-2), 99-115.

Perez, P. and Trier, A. (2001). "Prediction of NO and NO_2 concentrations near a street with heavy traffic in Santiago, Chile." *Atmospheric Environment*, 35 (10), 1783-1789.

Zhu C.J. (1997). “Coefficient of vehicular emission discharge and trend of pollution in Guangzhou City.” *China Environmental Science*, 17(23), 216-219.

Urban short-term traffic flow forecasting using the Markov switching model

Xianghai Sun[1], Tanqiu Liu[2], and Biswajit Basu[3] M ASCE

[1]School of Traffic and Transportation, Changsha University of Science & Technology, Changsha, Hunan, 410076, China; email: sunxianghai@163.com
[2]Post-Doctoral Researcher, Station of Mathematics, Central South University, Changsha, Hunan , 410075, China; email: liutanqiu@163.com
[3]School of Engineering, Trinity College Dublin, Ireland; email: basub@tcd.ie

Abstract: Urban short-term traffic flow data is highly nonlinear and with time-varying characteristics, especially changing abruptly before and after a period of congestion. In this paper, a Markov switching (MS) model is proposed to model the short-term traffic flow forecasting. The model is suitable to describe a nonlinear time series of urban traffic flow. The proposed model classifies traffic flow conditions into two states corresponding to two regimes of the MS model respectively. The traffic flow model changes under each state, switching at a certain probability controlled by a first-order Markov chain. Comparisons with real data for traffic flow validate the method with accurate short-term forecasts.
Keywords: short-term traffic flow forecasting, the Markov switching model, nonlinear time-series.

Introduction

Accurate short-term forecasting of traffic flow is of paramount importance in real-time traveler information and route guidance systems. However, short-term traffic flow prediction from available traffic data is still a challenge. Several researchers have used univariate time series models for traffic low forecasting (Lee and Fambro 1999; Williams 2001). These popular models, such as the autoregressive integrated moving average (ARIMA) model and its various extensions, are linear time series models which may fail to capture the dynamics and nonlinearity existing in the traffic flow behavior.

Urban traffic flow changes abruptly during the transition phases of entering or leaving the rush hours (Vlahogianni et al. 2006). Hence, a nonlinear time series model based on the Markov switching (MS) model is developed in this paper. As one of the most popular nonlinear time series model, the MS model has a framework of multiple regimes corresponding to different traffic flow conditions. Therefore, the model is used here to estimate the change in the pattern of traffic-flow.

The paper presents the general Markov switching model and the analysis of short-term traffic flow data with model estimation and prediction results.

Methodology

Preliminary analysis

Generally speaking, the word "congestion" is understood as a state in which several vehicles are in movement and each of them progress slowly and fitfully. Thomson and Bull (2002) redefined congestion as the situation that arises when the entry of one vehicle into a traffic flow increases the journey time of the rest. Van Arem et al. (1997) roughly classified traffic conditions into three different types: free-flow, recurrent congestion and non-recurrent congestion. The recurrent congestion is caused by the fact that at a certain location, traffic demand exceeds the supply of road capacity. Non-recurrent congestion is caused by conditions such as incidents, road works or special events. The state of non-recurrent congestion will not be considered in this paper as the probability of this type of traffic condition occurring is quite small compared with that of the recurrent congestion. In addition, data points representing the state of non-recurrent congestion are easily eliminated in the study of short-term traffic flow forecasting. Therefore, this paper categorizes traffic flow conditions of urban roads into two states, namely free-flow and congestion.

The MS model proposed here is based on the model by Hamilton (1989). Traffic flow time series is assumed to be approximated by an autoregressive process, whose parameters evolve through time. Moreover, their evolution is ruled by an unobservable variable, called the state variable, which in turn follows a first-order Markov chain. Therefore, the MS model involves multiples structures (AR equations) that can characterize the time series behaviors in different regimes. By permitting switching between these structures, this model is able to capture more complex dynamic patterns. In particular, the Markovian property of the model regulates that the current value of the state variable depends on its immediate past value. As such, a structure may prevail for a random period of time, and it will be replaced by another structure when a switch takes place. Obviously, traffic flow condition characteristics can easily be captured by a two-regime MS model, in which the two regimes correspond to free-flow state and congestion state respectively.

Markov switching model

Let y_t denote the traffic flow. Let us assume that the MS model has two regimes, represented by an unobservable process denoted by the state variable S_t, which can take the values 0 and 1, depending on the prevailing regime at time t. In this case the time series, y_t, will be different in each regime:

$$y_t = \phi_{0,0} + \phi_{1,0} y_{t-1} + \ldots + \phi_{m,0} y_{t-p} + \varepsilon_{t,0}, \quad \text{if } S_t = 0 \tag{1}$$

$$y_t = \phi_{0,1} + \phi_{1,1} y_{t-1} + \ldots + \phi_{n,1} y_{t-p} + \varepsilon_{t,1}, \quad \text{if } S_t = 1 \tag{2}$$

where $\varepsilon_{t,j} \sim N(0, \sigma_j^2), j = 0,1$. Following Hamilton (1989), further assume that S_t is a first-order Markov-process, which means that the current regime (S_t) depends only on the regime in the preceding period (S_{t-1}). This model is completed by defining the transition probabilities of moving from one regime to another:

$$p(S_t = j \mid S_{t-1} = i, S_{t-2} = k, \cdots) = p(S_t = j \mid S_{t-1} = i) = p_{ij} \qquad i, j, k = 0,1 \tag{3}$$

Since p_{01} = 1-p_{00} and p_{10} = 1-p_{11}, the transition probabilities are completely defined by $p_{00}$and p_{11}. These transition probabilities can be used to represent the transition matrix $\mathbf{P}$ = $[\,p_{ij}\,]$; $i, j = 1,2$ controlling the behavior of traffic flow. Maximum likelihood (ML) estimation of the MS model is based on implementation of the expectation maximization (EM) algorithm proposed by Hamilton (1994).

According to the assumption that the error terms of the MS model follow normal distribution, the short-term flow variable y_t has conditional probability densities based on different state as follows:

$$\boldsymbol{\eta}_t = \begin{bmatrix} f(y_t \mid S_t = 0, \mathbf{y}_{t-1}; \boldsymbol{\theta}) \\ f(y_t \mid S_t = 1, \mathbf{y}_{t-1}; \boldsymbol{\theta}) \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{(y_t - \phi_{0,0} - \phi_{1,0} y_{t-1} - \ldots - \phi_{m,0} y_{t-m})^2}{2\sigma_0^2}} \\ \dfrac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(y_t - \phi_{0,1} - \phi_{1,1} y_{t-1} - \ldots - \phi_{n,1} y_{t-n})^2}{2\sigma_1^2}} \end{bmatrix} \tag{4}$$

where, j=0,1; $\mathbf{y}_{t-1} = (y_{t-1}, y_{t-2}, ..., y_1)'$ represents all traffic flow observations up to time t-1; $\boldsymbol{\theta}$ denotes the population parameter vector, which includes $\phi_{0,0}, ..., \phi_{m,0}, \phi_{0,1}, ..., \phi_{n,1}$, and σ_0^2, σ_1^2.

Let the (2×1) vector $\hat{\boldsymbol{\xi}}_{t|t}$, called the filter probability, represent the analyst's inference about the value of S_t based on the data obtained through time t and on the knowledge of the population parameters $\boldsymbol{\theta}$. Let the (2×1) vector $\hat{\boldsymbol{\xi}}_{t+1|t}$, denote the prediction probability, representing the analyst's forecast of how likely the process is to be in regime j at the time point t+1 given the observations obtained through time t. According to Hamilton (1994), the optimal inference and forecast for each time t in the sample can be found by iterating on the following pair of equations:

$$\hat{\boldsymbol{\xi}}_{t|t} = \frac{(\hat{\boldsymbol{\xi}}_{t|t-1} \boldsymbol{-} \boldsymbol{\eta}_t)}{\mathbf{1}'(\hat{\boldsymbol{\xi}}_{t|t-1} \boldsymbol{-} \boldsymbol{\eta}_t)} \tag{5}$$

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{P} \cdot \hat{\boldsymbol{\xi}}_{t|t} \tag{6}$$

where $\mathbf{1}$ represents a (2×1) vector of 1s, and the symbol $\boldsymbol{-}$ denotes element-by-element multiplication. Given a starting value $\hat{\boldsymbol{\xi}}_{1|0}$ and an assumed value for the population parameter vector $\boldsymbol{\theta}$, one can iterate on equation (5) and

(6) for t=1, 2, …, T to calculate the value of $\hat{\boldsymbol{\xi}}_{t|t}$ and $\hat{\boldsymbol{\xi}}_{t+1|t}$ for each time t. Then, the log likelihood function $\mathcal{L}(\boldsymbol{\theta})$ can be calculated using the function $f(y_t \mid \mathbf{x}_t, \mathbf{y}_{t-1}; \boldsymbol{\theta}) = \mathbf{1}'(\hat{\boldsymbol{\xi}}_{t|t-1} \odot \boldsymbol{\eta}_t)$ and the estimates $\hat{\boldsymbol{\theta}}$ can be obtained by maximizing $\mathcal{L}(\boldsymbol{\theta})$. Then the 1-step-ahead forecasting value is:

$$\hat{y}_{t+1|t} = \mathrm{E}(y_{t+1} \mid \mathbf{y}_t; \boldsymbol{\theta}) = \mathbf{h}_t' \hat{\xi}_{t+1|t} \tag{7}$$

where

$$\mathbf{h}_t = \begin{bmatrix} \hat{\phi}_{0,0} + \hat{\phi}_{1,0} y_{t-1} + \ldots + \hat{\phi}_{m,0} y_{t-m} \\ \hat{\phi}_{0,1} + \hat{\phi}_{1,1} y_{t-1} + \ldots + \hat{\phi}_{n,1} y_{t-n} \end{bmatrix} \tag{8}$$

Similarly, the optimal k-step-ahead (k is an integer and larger than 1) forecast of the traffic flow is:

$$\hat{y}_{t+k|t} = \mathrm{E}(y_{t+k} \mid \mathbf{y}_t; \boldsymbol{\theta}) = \mathbf{h}_t' \hat{\xi}_{t+k|t} \tag{9}$$

where the k-step-ahead of the state variable is $\hat{\xi}_{t+k|t} = \mathbf{P}^k \cdot \hat{\xi}_{t|t}$.

Comparison of forecasting results

The traffic data for analysis is the discrete time series of vehicle flow rates recorded at every 5 minutes from 7:00 AM to 17:00 PM on each day for consecutive five weekdays from a road intersection in a capital city in China. The data is from Oct 15 to Oct 19, 2006 leading to a total of 600 observations, of which the first 480 observations are used to specify and estimate the MS model, and the remaining 120 to compare with the forecasting. During the sampling period there were no special events (e.g., traffic accidents, road maintenance etc.) occurring and so the data is consistent with the two different traffic flow conditions used in developing the model. Prior to estimating the MS model, test for stationarity is carried out in order to avoid "false regression" caused by non-stationary data. The augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests are used for this purpose by employing three test equations, including intercept, trend and intercept, and none respectively. Table 1 shows that the traffic flow sequence is non-stationary. To remove the non-stationarity, the average value is subtracted from the time series. The results in Table 1 indicate that the null hypothesis of a unit root for the resultant time series can be significantly refused, ensuring stationarity. Hence, the new time series $\{y_t'\}$ is used to develop the MS model. Table 2 shows that under the free-flow state, the current traffic flow is significantly correlated to the second-lagged traffic flow; under the congestion state (S_t=1) the current traffic flow is significantly correlated to the first and second lagged traffic flows. Obviously, traffic flow has different data generating processes (DGPs) under the different traffic condition states.

The parameter estimates in Table 2 are used to carry out one-step-ahead predictions of 5-minute traffic flow and compared with observations on the fifth day. Once a predicted value is obtained, observations at the past two time points and the parameter estimates are substituted in Eq. (7) and (8) in order to obtain traffic flow in the next 5-minute interval. The comparison of the measured traffic flow and that of the predicted traffic flow in Fig. 1 suggests that the MS is able to precisely make real-time predictions for 5-minute-interval traffic flow at urban roads.

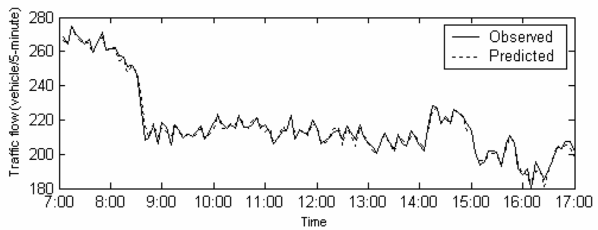


Figure 1. Predicted flow versus observed flow

Table 1. Results of Unit Root Tests

Test	y			y'		
	Intercept	Trend and intercept	None	Intercept	Trend and intercept	None
ADF	-1.3974 (0.5837)	0.14156 (0.9975)	-0.5951 (0.4589)	-9.2026 (0.0000)	-9.3154 (0.0000)	-8.6646 (0.0000)
PP	-1.2070 (0.6722)	-0.5224 (0.9821)	-0.4978 (0.5000)	-9.3349 (0.0000)	-9.3969 (0.0000)	-8.8831 (0.0000)

Note: These values in the table are t statistics of the ADF test and PP test respectively, and the p-value corresponding to these statistics is reported in parentheses.

Table 2. Estimation Results for the MS Model

The free-flow state: $S_t = 0$		The congestion state: $S_t = 1$	
$\phi_{0,0}$	0.3878 (0.3698)	$\phi_{0,1}$	1.9338 (3.0172)
$\phi_{1,0}$	1.0259* (11.6242)	$\phi_{1,1}$	—
$\phi_{2,0}$	-0.4837* (-4.1158)	$\phi_{2,1}$	0.5062* (6.4287)
Transition probabilities	$\begin{bmatrix} 0.6111 & 0.6228 \\ 0.3889 & 0.3772 \end{bmatrix}$		

Note: The parameters are described in the text. Values given in parentheses below the parameter estimates are t statistics. Significance at 1% level is indicated by *.

Conclusions

This paper proposes a nonlinear time series model as a powerful approach for capturing the dynamics of urban short-term traffic flow. Since, similar studies seem to be absent in the existing traffic flow forecasting literature, this paper is possibly one of the first attempts in the application of the proposed Markov switching time series model to traffic and transportation engineering.

The model is conceptually appealing as traffic flow conditions in urban cities can be classified into two states, namely the free-flow state and the congestion state. The model can switch between these two states at a certain probability controlled by a first-order Markov chain. In this study the two traffic flow conditions are corresponding to two regimes of the MS model respectively, so there are two different DGPs of short-term traffic flow. This study confirms that the MS model not only captures nonlinear characteristics displayed in the short-term traffic flow very well, but also makes precise real-time predictions for 5-minute traffic flow at urban roads.

References

Hamilton, J.D. (1989). "A new approach to the economic analysis of nonstationary time series and the business cycle." *Econometrica*, 57, 357-384.

Hamilton, J.D. (1994). *Time Series Analysis, Princeton*, NJ: Princeton University Press.

Lee, S., and Fambro, D. B. (1999). "Application of subset autoregressive integrated moving average model for short-term freeway traffic volume forecasting." *Transportation Research Record*. *1678,* Transportation Research Board, Washington, D.C., 179-188.

Thomson, I., and Bull, A. (2002). "Urban traffic congestion: its economic and social causes and consequences." *CEPAL REVIEW 76*, 105-116.

Van Arem B., Kirby H.R., Van Der Vlist M.J.M., and Whittaker J.C. (1997). "Recent advances and applications in the field of short-term traffic forecasting." *International Journal of Forecasting*, 13, 1-12.

Vlahogianni E.I., Karlaftis M.G., and Golias J C. (2006). "Statistical methods for detecting nonlinearity and non-stationarity in univariate short-term time-series of traffic volume." *Transportation Research Part C*, 14, 351-367.

Williams B. M. (2001). "Multivariate Vehicular Traffic Flow Prediction: An Evaluation of ARIMAX Modeling." *Transportation Research Record. 1776,* Transportation Research Board, Washington D.C., 194-200.

Evaluation of Cellular Probe Traffic Data: Issues and Case Study

Saini Yang[1] and Ali Haghani[2]

[1]Assistant Professor, Cleveland State University, Department of Civil and Environmental Engineering, Stilwell Hall 122, Cleveland, OH 44115; PH (216) 687-2230; FAX (216) 687-6593; email: s.yang27@csuohio.edu
[2]Professor and Chair, University of Maryland, Department of Civil and Environmental Engineering, College Park, MD 20742; PH (301) 405-1963; FAX (301) 405-3100; email: haghani@eng.umd.edu

Abstract

While cellular probe traffic data shows its great potential in applications of Intelligent Transportation System (ITS), especially, the Traveler Information System (TIS), the data quality becomes a more and more important issue for real world application. The purpose of the paper is to discuss the issues in probe data evaluation and their potential solution approaches. The selections of data source, sampling strategies and data analysis techniques can be complicated problems. A case study is presented. It is observed in the study that cellular probe traffic data provide ideal spatial and temporal coverage and the precision of the cellular probe traffic travel times on freeways qualifies immediate ITS application. It indicates that it is necessary to further improve the algorithm on arterials and to coordinate signal setting within the algorithm. With future improvement of cellular probe technology, ITS applications will be further benefited. (142 words)

1. Introduction

Development of wireless technology and the increasing penetration rate of cellular phones provide an enormously valuable tool to improve the performance of transportation system. By mapping cellular carriers' trajectories on road networks, we can get not only the real time traffic speeds but also the real time travel time information. Compared to the traditional traffic data collection approaches, cellular probe data has its unique advantages, such as the ready-to-use infrastructure, high network (includes freeways, highways, arterials and streets) coverage, and low installation and maintenance cost. However, there are some concerns regarding to the data quality. The potential bias in sampling cellular carriers and interference of pedestrians or bicyclists who carry cellular phone may cause problem in data accuracy. The signal quality and cellular network coverage may influence the probe data quality, especially, some parts of the road network having no cell service will not be able to provide real time traffic information. Therefore, before applying cellular probe data in real world application, it is important to evaluate precision and quality of data.

In 1999, Intelligent Transportation Society (ITS) of America and the U.S. DOT developed guidelines for quality advanced traveler information system data (2000) to support the expansion of traveler information products and services, explicitly, to control the quality of traffic data being collected. Federal Highway Administration (2003) selected accuracy, confidence, delay, availability, breadth of coverage and depth of coverage (density) as data quality measures. A 20% error tolerance level for traffic speed accuracy in traveler information applications and 5-10% for speed accuracy in traffic management were recommended by Tarnoff (2002).

In the past decade, a number of simulation tests and field tests have been conducted to evaluate cellular probe-based traffic information. Simulation frameworks were used to test cellular applications (Ygnace, et al. 2000; Fontaine and Smith, 2005). Simulation has its limitation in representing the complexity of the real world condition. University of Maryland implemented the first field test on wireless location technology (WLT) (University of Maryland Transportation Center, 1997) in U.S. on several interstate and state routes. Ygnace (2001) reported an average traffic speed difference of 24%-32% on urban freeway and 10% difference on inter-city freeway from their field test in Lyon, France. The loop detector data was used for these evaluations. As pointed out in 1999, the mean error of loop detector data itself ranges from 10% to 20%. Smith, et al. (2003) used point sensor data in an evaluation of link travel speeds by WLT. A van-mounted video detection system was used to derive spot speeds and counts by processing video from a camera. However, this point sensor could not provide travel time information which is essential in travel information systems. Bar-Gera (2007) used vehicular probe and loop detector data in their evaluation. Graphic identification in cellular probe evaluation is recommended. Though the graphic presentation shows the big picture of data similarity, when the travel speeds or travel times have significant variation, a short delay in cellular probe data will cause large error for users. It is important to have multiple data sources in an evaluation and figures of detailed performance measures need to be provided.

2. Issues in Evaluation

An evaluation procedure includes several major tasks, such as initial study, sampling scheme design, data quality control and data analysis.

The preliminary information such as the cellular probe data type, data format, and the important factors that may influence the cellular probe data need to be collected before selecting the data sources for comparison and sampling strategies. How to select the data source for evaluation is the first major issue. In most studies, the vehicular probe data or sensor data is treated as the “Ground Truth” in the calculation of error terms. In reality, there is no universally accepted measure of “Ground Truth” and each measure has advantages and disadvantages as mentioned before. The accuracy of the “Ground Truth” directly affects the validity of evaluation results. Currently, there are two major types of data sources used in cellular probe data evaluation: probe vehicles and point sensors. Probe vehicles equipped with GPS systems can be used to collect traffic information in real time while traveling in the traffic stream. Ideally, when drivers strictly follow floating car method, the average travel speed on a road segment should be obtained. In practice, when the traffic is congested on road, it is difficult for a driver to pass one car for every car that passes him or her, especially, on arterials or streets with traffic lights installed. Loop detector/sensor data has been used in data quality evaluation for a long time. Since the loop detector is only available on certain location, it is impossible to use them in network wide evaluation. Furthermore, due to detector health and hardware reasons, not all detectors/sensors can provide accurate and reliable traffic information. It is important to have multiple data sources for comparison purpose in an evolution.

In data collection and data preparation, there are different error sources. For cellular probe data, sampling error may be caused by bias of demographic characteristics of phone users, or may be caused by insufficient sample size. For vehicular probe data, errors may be introduced during the data collection and data processing. For example, operations such as editing, reviewing, or keying data may introduce error into the estimates. It is suggested to have data quality control during the entire data collection procedure. The sampling scheme should be adjusted in time in order to reduce potential waste of efforts and improve accuracy of evaluations.

For data analysis, it is necessary to provide figures of performance measure and to conduct statistical analyses on error terms. Besides average error and average absolute error, percentile error term and percentage of absolute error are necessary. For traveler information users, a 20 kilometer per hour (kph) difference may be acceptable on freeways with an average traffic speed of 120 kph, while on a 35 kph arterial, this difference will be too large. The percentile error term is important due to the outlier effect. This impact is caused by some extremely large errors. For example, when the probe vehicular speed is 10 kph and the cellular speed is 40 kph, the absolute error is 300 percent. These outlier data points are rare, but they contribute to large average percentage of absolute error. The histogram analysis of the data can help analyze the data quality as well. Randomness property is essential in selecting statistical tests. If an observed value is influenced by its position in the sequence, or by the observations which precede it, the process is not truly random. Traffic information is influenced by many factors, such as the time of day, day of week, road type, weather, and so on. Their randomness is questionable when we are exploring the traffic information during time interval of interest. Though parametric tests are stronger than nonparametric tests in general, it is crucial to test the normality of data before applying any parametric tests. The selection of statistical test is based upon the nature of data.

These issues discussed above can bring large variability to the data and greatly degrade data quality measures. Clarifications of these issues can help not only interpret the data but also design the evaluation process.

3. Case Study

This case study is a part of a Multimodal Traveler Information System project. The project is an independent study sponsored by state DOT, private sector, Federal Highway Administration (FHWA) and other government agencies. A real metropolitan road network is used in this case study, as shown in Figure 1.

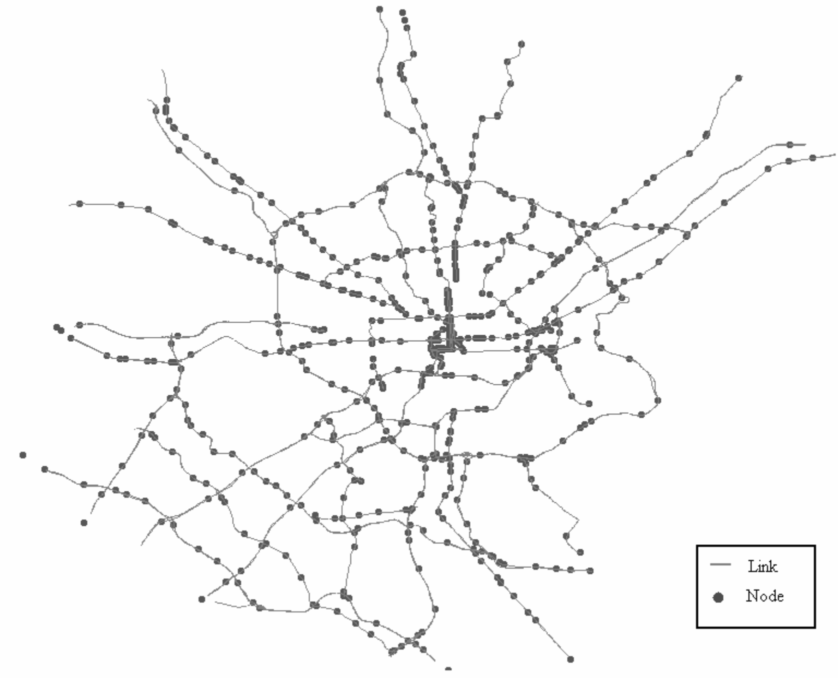
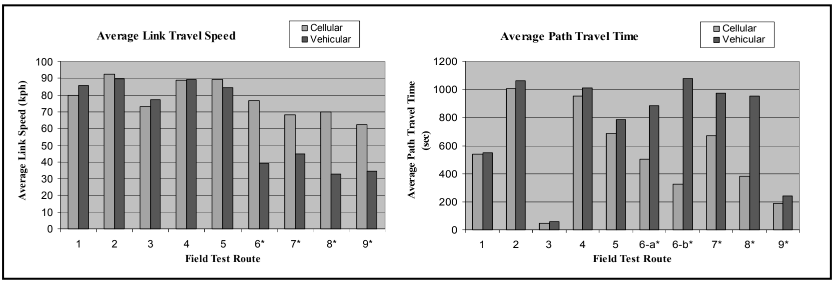


Figure 1. Road network of case study area.

Two major data sources were used in evaluation: traffic information from probe vehicle data and the Remote Traffic Microwave Sensor (RTMS) data. A total of 500 probe vehicle survey runs covering 14,500 kilometer were conducted and travel speed information from 10 sensors location is collected every 5 minutes on a beltway. The cellular probe data and RTMS data span the entire probe vehicle survey duration. The cellular travel speed and travel time on each link are provided in every 5-minute interval. In this case study, the data quality evaluation is focused on the travel speed on short links and travel time on longer paths. Five error measures are considered: average error, average absolute error, 50 percentile absolute error, 85 percentile absolute error, and average absolute percentage of error. Microsoft Access is used for data processing and statistics software SPSS 14.0 is used in statistical analysis.

It is observed in the study that cellular probe traffic data provides ideal spatial and temporal coverage for the entire road network during the whole test period. The average error terms of travel speed and travel time are shown in Figure 2. Both evaluation results on freeways are promising. For speed evaluation, the average absolute errors on freeways are about 17 kph. The 50 percentile absolute errors on freeways are in a range of 9 kph to 16 kph. For travel time evaluation, the deviations between path travel times obtained from cellular probe data and those obtained from vehicular probe data are about 10%. On arterials, on the other hand, the travel speed obtained from cellular probe data is significantly larger than those from probe vehicle data. The cellular probe speeds consistently overestimate vehicular link speeds and the 50 percentile absolute errors ranges from 24 kph to 40 kph, which are about 80% to 100% of the vehicular average link speeds. The cellular path travel times are significantly smaller than those obtained from probe vehicle data on arterials, the difference ranges from 40% to 60% of the vehicular travel times. In short, the precision of the cellular probe traffic travel times on freeways is good enough (around 10% error terms) for immediate ITS application and that of the traffic speeds on freeway is acceptable (around 20% error terms). On arterials, the cellular probe data cannot provide accurate travel speed and travel time information in this study.



Notes: Routes 1-5 are freeways and routes 6-9 are arterials. Route 6 is segmented to two paths in travel time tests.

Figure 2. Comparison of travel speeds and travel times.

The comparison among cellular probe data, sensor data and vehicular probe data are based upon traffic speeds only. The cellular data outperforms the sensor data, especially in capturing the overall picture of the variations in vehicular speeds.

4. Conclusion

The application of cellular probe vehicle data in TIS heavily depends on the quality of the data itself, especially in large scale real world operations. Cellular probe data quality evaluation has great importance in providing a solid base for future operations. This study discusses the issues in probe data evaluation and their potential solution approaches. Cellular phone tracking, at least the one implementation in this case study, shows promise and would benefit from additional technology improvement. On most freeways, the evaluation results are consistently good. On arterials in which traffic signals exist, the cellular probe data tends to over-predict the travel speeds and under-predict the travel times significantly. It is necessary to conduct further analysis for arterials. Additional information such as signal phase setting could be useful and should be provided for analyzing cellular probe data quality. With future improvements of cellular probe technology, ITS applications will be further benefited.

References

1. ITS America and U.S. Department of Transportation. "Closing the Data Gap: Guidelines for Quality Advanced Traveler Information System (ATIS) Data." Version 1.0, September 2000.
2. "Defining and Measuring Traffic Data Quality." Office of Policy, Federal Freeway Administration (2003), Washington, DC.
3. Tarnoff, P.J (2002),"Getting to the INFOstructure." White Paper prepared for the TRB Roadway INFOstructure Conference.
4. Ygnace, J., Drane, C., Yim, Y.B. and de Lacvivier, R., Travel Time Estimation on the San Francisco Bay Area Network Using Cellular Phones as Probes. University of California, Berkeley, PATH Working Paper UCB-ITS-PWP-2000-18, 2000.
5. Fontaine, M.D., and Smith, B.L. Probe-Based Traffic Monitoring Systems with Wireless Location Technology: An Investigation of the Relationship Between System Design and Effectiveness, Transportation Research Record: Journal of the Transportation Research Board No. 1925, 2005, 3-11.
6. University of Maryland Transportation Studies Center (1997), "Final Evaluation Report for the CAPITAL-ITS Operational Test and Demonstration Program." University of Maryland, College Park, MD, 1997.
7. Ygnace, J. (2001), "Travel Time/Speed Estimates On The French Rhone Corridor Network Using Cellular Phones As Probes." SERTI V Program, STRIP (System for Traffic Information and Positioning) Project, INRESTS – LESCOT 0201, Lyon, France.
8. Smith, Brian L., Han Zhang, Mike Fontaine and Matt Green (2001), "Cell Phone Probes as an ATMS Tool." Project report for the Center for ITS Implementation Research (ITS).
9. Bar-Gera, Hillel (2007), "Evaluation of a Cellular Phone-based System for Measurements of Traffic Speeds and Travel Times: A Case Study from Israel." Proceeding CD-ROM of 86th TRB Annual Meeting, Washington D.C.

Microscopic Traffic Simulation for Transportation Planning
– An Application on Transit System Planning

Xiaobo Liu[1], Mei Chen[2] and Hu Su[3]
1 Jacobs Engineering Group, Department of Planning, Morristown, NJ, (973) 267-8830, ext 1143, xiaobo.liu@jacobs.com
2 University of Kentucky, Department of Civil Engineering, Lexington, KY, (859) 257-9262, mchen@engr.uky.edu
3 Southwest Jiaotong University, School of Electric Engineering, Chengdu, China, (28)66162503, suhu@home.swjtu.edu.cn

Abstract

This study presents a practical approach to transit system planning using microscopic simulation with emphasis on operational issues. The microscopic simulation model can generate detailed estimates of traffic conditions considering time-varying demand and individual driver's behavior. Therefore, it can evaluate drivers' interactions with complex traffic environment and analyze the impact of major roadwork improvement and/or advanced transportation management strategies on the transit and traffic operations.

A microsimulation-based transit planning practice is presented as case study. Exhaustive analyses combining various alignments and three transit technologies including BRT, LGR, and AGT are conducted to evaluate the performance of each alternative. The case study demonstrates the feasibility, applicability and benefits of the proposed approach.
Keywords: Microsimulation, Transportation Planning, Transit System

Introduction

Traffic simulation modeling has been widely implemented in traffic analysis and evaluation. By replicating the traffic operational dynamics, the simulation model can generate measures of effectiveness (MOE) to test and evaluate tentative alternatives without actual implementation. In addition, the simulation model can effectively analyze the complex traffic flow process that is infeasible to be analytically formulated.

Based on the representation of traffic flow or vehicle movement (Chu, 2002), traffic simulation models can be classified into microscopic (CORSIM, TRANSIM, MITSIM, PARAMICS, VISSIM, AIMSUN2, etc.), mesoscopic (FREFLO, METANET, AUTOS, VISUM, etc.), and macroscopic (DYNASMART, DYNAMIT, INTEGRATION, METROPOLIS, etc.). Microsimulation is the dynamic and stochastic modeling of individual vehicle movements within a system of transportation facilities. It can identify a vehicle by type (e.g., auto, truck and bus), specify driver characteristics (e.g., aggressive or cautious), and emulate stochastic driving properties behavior (e.g., lane-change maneuver, car following logic and driver decision process) individually. It is capable of describing dynamic interactions among vehicles and generating fully disaggregated vehicle behavior in a complicated surface traffic environment. It has been widely applied in dynamic network analysis, ITS applications, transportation infrastructure design, traffic operations, safety study, etc.

There are several challenges to building a high-quality traffic simulation model; they include computational performance, accuracy real traffic system representation, and integration with advanced traffic management and traveler information systems (Erol, etc). In this paper, a microsimulation model uniquely designed for transit system planning application is developed to address these issues.

Case Study

The case study is to provide simulation-based alternative analysis for a transit development in a corridor of a major city in the Midwest US. There are a number of alternative alignment designs to be evaluated for the area. Critical segments along the transit line that pose the most challenge to operation and/or are representative of other segments of the alignments are chosen for detailed analysis.

Three prevailing transit technologies are studied including Automated Guideway Transit (AGT), Light Rail Transit (LRT), and Bus Rapid Transit (BRT). AGT uses elevated right-of-way along the median of the public streets. Steel wheels are placed on the rail for guidance only. LRT also uses steel wheels on rail, but for traction as well as guidance. The proposed LRT system operates at-grade in this study. Where it operates longitudinally along a street, it is located either (desirably) in the median between opposing directions of traffic, or (as a second-best option) along one side of a public street. BRT is a rubber-wheeled technology that uses right-of-way similar to LRT. It is capable of operating in either isolated right-of-way or shared right-of-way together with general-purpose traffic. BRT vehicles have higher capacity; they are longer and more comfortable than standard urban buses.

Modeling Approach

The Paramics microsimulation technique is chosen for this analysis, in which stochastic (i.e. automobile) and fixed (i.e. transit vehicle movements with given schedule) demands are converted into the operations of individually simulated vehicles. The Paramics model has a very low level of granularity. Operations are analyzed at the level of the individual transit vehicles. Intersection analysis covers technical details such as the turning lane attributes and signal operations. Microsimulation at the segment or intersection level can help to identify the implications and trade-offs among key elements such as physical right-of-way (ROW) availability, transit quality of service, traffic level of service, pedestrian accessibility, and community character /visual impacts.

In order to ensure an exhaustive analysis of combinations of alignments and technologies, all three transit technologies with 5-mintue service headway are analyzed for each alignment. The conceptual station locations have been previously identified for each alignment.

Network Development

The automobile and commercial traffic in the Paramics is generated after the origin-destination (OD) matrices representing travel demand are developed.

Therefore, the transportation analysis zones (TAZ's) structure is designed to ensure that individual zones represent each roadway segment and transit alignment that is cut off by the location boundaries. Within the locations under study, minor streets and major private driveways are set as discrete zones that attract or produce traffic. The model development process includes a matrix generation process, wherein traffic patterns are replicated based on turning movement from field count data at each intersection.

The Paramics model uses three different themes (shortest path, stochastic algorithm, and dynamic assignment) to determine the travel path between each OD pair. The shortest path assumes all traffic between a particular OD pair always choose their routes with the lowest cost (in terms of travel time, distance, or associated cost). The stochastic algorithm applies the perturbation cost on different travelers to generate alternative routes between each OD pair. The dynamic assignment will recalculate the travel cost based on the feedback of the travel cost of each link during each time interval, therefore the travel path are dynamically updated to reflect the impact of the traffic congestion on the route choice of automobiles. Figures 1 and 2 show the shortest path and stochastic path for one same pair of zones, respectively. Since traffic in this area is mostly generated by regular commuting trips; therefore, most drivers are assumed to be able to choose alternative paths when traffic congestion occurs on their usual routes. Thus, the dynamic assignment with stochastic perturbation is adopted to determine the travel path between each pair of OD in this study. The assigned traffic in the simulation model is compared with real traffic count to ensure that the matrices that represent traffic demand are accurate to the level of the individual turning movement. The desire line shown in Figure 3 illustrates the traffic patterns between each pair of OD.

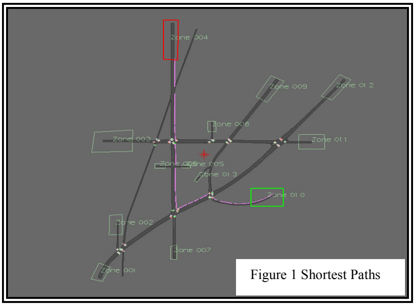


Figure 1 Shortest Paths

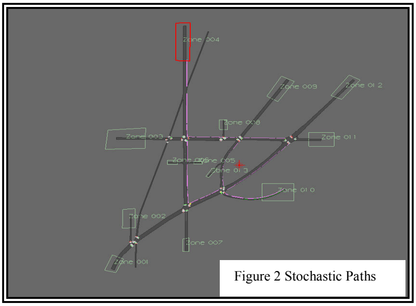


Figure 2 Stochastic Paths

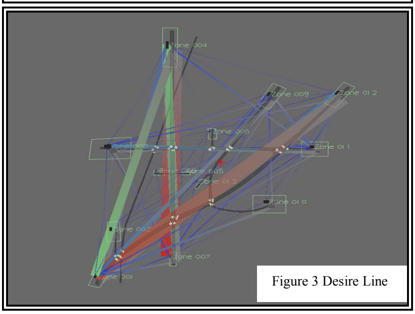


Figure 3 Desire Line

Potential Intersection and Segment Treatments

Intersection and segment treatments are developed as each of the locations and roadway alignments are analyzed in detail. Since AGT operates on the elevated track, there is no conflict between the AGT train and the general traffic. LRT and BRT operate at the grade, it is necessary to give an efficient and safe operation plan at the junctions where transit line need to cross. Conceptually, three potential solutions are provided for at-grade LRT and BRT operations on existing roadway.

1. When transit operates in the median between the general-purpose traffic lanes of two opposite directions, the prohibition of left turns across the transit facility may be needed on a site-specific basis. Waiting areas at transit stations located within a median would require additional separation from the general-purpose traffic lanes, which can be mitigated by offsetting the platform on each direction.
2. When transit operates along the curb lanes, this requires site-specific consideration of how right-turn movements in and out of both public streets and private properties.
3. When bi-directional transit operates along only one side of the street's right-of-way, complex signalization is needed to manage conflicts between transit vehicles and general-purpose turning traffic. Depending on the roadway layout, it may require physical relocation of existing lanes, signal infrastructure, utility lines, etc.

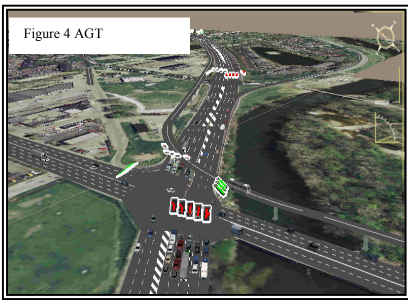


Figure 4 AGT

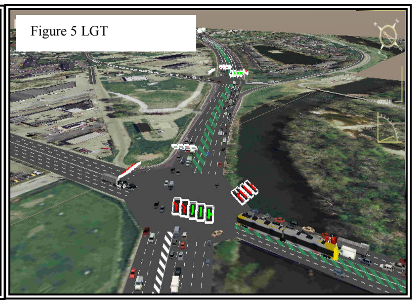


Figure 5 LGT

Examples of intersection treatments are shown in Figure 4, 5 and 6 for AGT, LRT and BRT crossing at one same location, respectively. Figure 4 shows that the grade-separated AGT track avoids the conflict of AGT and general traffic. Only minor geometry change is needed at the intersection to provide enough room to install the poles of AGT guideway track. Figures 5 and 6 show the designed intersection for LRT and BRT operation, respectively. The geometry is modified to give transit priority at the intersection. Exclusive transit lane marked in green color is designated for the AGT system in the median, while BRT operates at the curbside of the roadway. BRT is designated with exclusive lane only on certain sections, while on other sections BRT bus shares a lane with the general-purpose traffic. Conflicting vehicular traffic of perpendicular roadways is controlled by

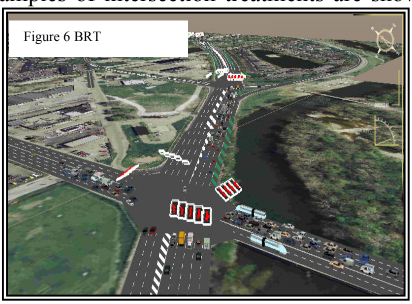


Figure 6 BRT

customized signal timing plan to ensure an independent phase assigned for bus turning movement. Motorists crossing the transit way will encounter tricolor traffic signals as at any signalized intersection.

Performance Measures

The transit plan for each alignment with each transit technology is developed based on corresponding simulation models. The traffic operation along the major corridor and transit system performance can then be evaluated based on the measures of effectiveness (MOE) generated by the simulation model. Various MOEs, such as the average travel times of general-purpose traffic and transit can be generated to indicate the difference of traffic operation of different alternatives.

Besides of the traffic operational analysis, simulation model can provide additional performance measures to address other issues for transit system development, including

- Requirements for property acquisition
- Level of service for transit vehicles
  - Average travel speeds
  - Reliability (standard deviations in transit travel time reliability
- Degree of complexity of physical and signalization treatments
- Ability to prevent or mitigate impacts to general purpose traffic operations
  - Requirements for turning movement restrictions
  - Intersection delay
  - Queuing
  - Requirement for displacement of on-street and/or off-street parking spaces
  - Pedestrian access to stations and impacts to general pedestrian mobility

For example, Table 1 summaries performance measures of various alternatives at one location. It includes transit reliability, impact of transit on general traffic operation and circulation pattern, efficiency of proposed traffic signals, complexity of geometric and physical treatments, impacts of geometric change on ROW and parking, pedestrian accessibility, etc. With such information, policy-makers can make effective assessment of alternative plans based on different preference or criteria for transit system planning.

Table 1 Performance Measures of Alternatives

Parameter	Alignment A			Alignment B		
	AGT	BRT	LRT	AGT	BRT	LRT
Transit Operating Speed	High	Medium	Medium	High	Low / Medium	Low
Transit Reliability	Highest	Low	High	Highest	Low	High
Traffic Impacts	N/A	Low	Medium	N/A	Low	High
Efficiency of New / Impacted Traffic Signals	N/A	(Auto); Medium	(Auto); Medium	N/A	(Auto); Low	(Auto); Low
Impacts to Circulation Patterns	Minimal	Low	Medium	Minimal	Low	High
Complexity of Geometric & Street-Level Physical Treatments	Minimal	Medium	Medium	Minimal	Medium	High
Right-of-Way / Parking Impacts	Minimal	Low	Low	Minimal	High	Medium
Degree of Pedestrian Accessibility (Generally and to Transit Stations)	High	Medium	Medium	Medium	Medium	Medium

Conclusion

The case study presented in the paper uses microsimulation model to conduct alternative analysis for a transit system development. The Paramics model employed in this study has demonstrated sufficient capability to combine the travel demand planning and operational analysis together. The stochastic traffic pattern of general traffic and fixed transit line can be accurately represented in the simulation model. Meanwhile, the microsimulation model can effectively mimic the individual vehicles' interaction with transportation environment including geometry, signalization plan, and advanced traffic management strategies. It allows transportation professionals to conduct an in-depth analysis for traffic operational characteristics even during planning phase.

Reference

Balakrishna, R., Antoniou, C., Ben-Akiva, M., Koutsopoulos, H., Wen, Y., (2007) "Calibration of Microscopic Traffic Simulation Models: Methods and Application," the 86th Transportation Research Board Annual Meeting, Washington D.C.

Chu, L., Liu, H., Recker, W., and Zhang, M., (2001) "Development of a Simulation Laboratory for Evaluating Ramp Metering Algorithms," Center for Traffic Simulation Studies. http://repositories.cdlib.org/itsirvine/ctss/UCI-ITS-TS-WP-01-13

Erol, K., Levy, R., Wentworth, J., "Application of Agent Technology to Traffic Simulation," http://www.tfhrc.gov/advanc/agent.htm.

Kitamura, R., Kuwahara, M., (2005) "Simulation Approaches in Transportation Analysis: Recent Advances and Challenges," Springer press.

Lieberman, E., Rathi, A., "Traffic Simulation," www.tfhrc.gov/its/tft/chap10.pdf.

Traffic Analysis Toolbox, Volume III: Guidelines for Applying Traffic Microsimulation Modeling Software, June 2004. Publication No. FHWA-HRT-04-040. http://ops.fhwa.dot.gov/trafficanalysistools/tat_vol3/index.htm

Stochastic User Equilibrium Traffic Assignment Model Based on Travel Time Budget

Aiwu Kuang*, Zhongxiang Huang*, and W. K. Victor Chan**

* School of Traffic & Transportation, Changsha University of Science and Technology, 410076, Changsha, P. R. China; email: jxgakaw@126.com.
** Department of Decision Sciences and Engineering Systems, Rensselaer Polytechnic Institute, CII 5015, 110 Eight Street, Troy, NY 12180, U. S. A.

Abstract: This paper first analyzes link and route travel time distributions in a degradable road network by assuming a uniform distribution on the capacity of each link. It is postulated that travelers, by acquiring route travel time variability from past experiences, make route decisions based on a travel time budget, which is the summation of the mean route travel time and the safety margin of the travel time. This paper then formulates a novel travel time budget-based stochastic user equilibrium traffic assignment model with multiple user classes and elastic demand as a variational inequality problem. Finally, a numerical example of a small road network is presented to illustrate the properties of the proposed model.
Key words: *travel time reliability, travel time budget, stochastic user equilibrium, multiple user classes, elastic demand*

1 Introduction

In reality, uncertain factors, such as natural disasters, traffic accidents, daily traffic congestion, and road maintenance etc., are randomly influencing the state of a road network, resulting in the degradations of link capacities. In a road network with uncertain link capacities, the link travel times and route travel times are also stochastic. In order to reflect the network performance with degradable link capacities, one can use the concept of travel time reliability—defined as the probability that a trip between a given origin-destination (OD) pair can be made successfully within a specified interval of time (Bell and Iida 1997).

With the presence of travel time variations, travelers do not know exactly the outcome of taking a particular route decision. For example, if travelers plan their trips based on the average route travel time, they could reach the destination sooner or later depending on the network conditions. As a result, travelers, especially commuters, should take into account the travel time variations in order to increase the chance of arriving on time. One common approach is to add an extra route travel time to the mean route travel time. Such an extra time is called the travel time safety margin in this paper. Accordingly, the route choice problem has to consider both the mean route travel time and the route specified safety margin, which together are referred to as a travel time budget in the literature (Lo et al. 2006). As in Lo et al. (2006), this paper postulates that the travel time budget varies with individuals and trip purposes. The travel time is related to the requirement on punctual arrivals as well as the travelers' objective of minimizing their travel time budgets. This paper defines a novel network stochastic user equilibrium (SUE) model based on the analyses of the random variations of link and route travel times as: ***at equilibrium point, for each OD pair, no travelers can improve their perceived travel time budget by unilaterally changing routes.***

2 Formulation of travel time budget

2.1 Link and route travel time distributions

Consider a road network $G(N, A)$ where N is the set of nodes and A is the set of links. Denoted the capacity of link a by C_a, which is a stochastic degradation random variable. We assume that the link capacity follows a uniform distribution with a lower bound being a fraction, θ_a ($0 \le \theta_a \le 1$), of the upper bound. We also adopt the Bureau of Public Road link performance function, which defines the travel time, T_a, under flow x_a using two deterministic parameters α and β as:

$$T_a(x_a,C_a) = t_a^0[1+\alpha\,(x_a/C_a)^{\beta}] \quad \forall a\in A \tag{1}$$

where t_a^0 is the deterministic free-flow travel time on link a.

According to (1) and assuming that the degradation random variable C_a is independent of the amount of traffic flow on it (i.e., x_a), the mean and variance of link a's travel time can be expressed as:

$$E(T_a) = t_a^0+\alpha t_a^0 x_a^{\beta} E[1/C_a^{\beta}] \quad \forall a\in A \tag{2}$$

$$Var(T_a) = \alpha^2 (t_a^0)^2 (x_a)^{2\beta} Var[1/C_a^{\beta}] \quad \forall a\in A \tag{3}$$

For uniformly distributed link capacities, the mean and variance of T_a can be derived as the following (for the general case of $\beta \neq 1$):

$$E(T_a) = t_a^0 + \alpha t_a^0 x_a^{\beta} \frac{(1-\theta_a^{1-\beta})}{\bar{c}_a^{\beta}(1-\theta_a)(1-\beta)} \quad \forall a \in A \tag{4}$$

$$Var(T_a) = \alpha^2 (t_a^0)^2 x_a^{2\beta} \left\{ \frac{1-\theta_a^{1-2\beta}}{\bar{c}_a^{2\beta}(1-\theta_a)(1-2\beta)} - \left[\frac{1-\theta_a^{1-\beta}}{\bar{c}_a^{\beta}(1-\theta_a)(1-\beta)} \right]^2 \right\} \quad \forall a \in A \tag{5}$$

where $\bar{c}_a$ is the design capacity of link a (i.e., the upper bound).

To simplify the calculation, this paper assumes that the link capacity (and hence link travel time) distributions are independent. Thus, the travel time T_k^w on route k between OD pair w can be simply expressed as the summation of the independent link travel time variables along route k :

$$T_k^w = \Sigma_a T_a \delta_{a,k}^w \quad \forall w\in W,\ k\in R_w \tag{6}$$

where W is the set of OD pairs, R_w is the set of routes between OD pair w, and $\delta_{a,k}^w$ is the link-route incidence parameter whose value is one if link a is on route k and zero otherwise. According to the Central Limit Theorem, in a network with routes consisting of many links, the route travel time follows a normal distribution regardless of the link travel time distributions. Therefore, the route travel time distribution, the mean route travel time t_k^w, and the route travel time standard deviation $\sigma_{t,k}^w$ can be expressed as:

$$T_k^w \sim N(t_k^w, (\sigma_{t,k}^w)^2) \quad \forall w \in W, k \in R_w \tag{7}$$

$$t_k^w = \Sigma_a E(T_a)\delta_{a,k}^w \qquad (\sigma_{t,k}^w)^2 = \Sigma_a \delta_{a,k}^w Var(T_a) \qquad \forall w \in W, k \in R_w \tag{8}$$

Combining (4) and (5) give the expressions for t_k^w and $\sigma_{t,k}^w$ as the following:

$$t_k^w = \sum_a \delta_{a,k}^w \left[t_a^0 + \alpha t_a^0 x_a^\beta \frac{(1-\theta_a^{1-\beta})}{\bar{c}_a^\beta (1-\theta_a)(1-\beta)} \right] \quad \forall w \in W, \ k \in R_w \tag{9}$$

$$\sigma_{t,k}^w = \sqrt{\sum_a \delta_{a,k}^w \alpha^2 (t_a^0)^2 x_a^{2\beta} \left\{ \frac{1-\theta_a^{1-2\beta}}{\bar{c}_a^{2\beta}(1-\theta_a)(1-2\beta)} - \left[\frac{1-\theta_a^{1-\beta}}{\bar{c}_a^\beta (1-\theta_a)(1-\beta)} \right]^2 \right\}} \quad \forall w, k \tag{10}$$

2.2 Model of travel time budget

As mentioned above, the travel time budget is the summation of the mean route travel time and the safety margin. It can also be expressed as:

$$c_k^w = t_k^w + m_k^w \quad \forall w \in W, k \in R_w \tag{11}$$

where c_k^w is the travel time budget associated with route k between OD pair w and m_k^w is the travel time safety margin determined by the following chance-constrained model:

$$\min c_k^w \tag{12}$$

$$\text{s.t.} \quad \Pr[T_k^w \le c_k^w] \ge \rho \quad \forall w \in W, k \in R_w \tag{13}$$

where ρ is the travel time reliability requirement.

Since T_k^w follows a normal distribution, m_k^w can be obtained by directly solving constraint (13). As a result, m_k^w and c_k^w can be expressed as:

$$m_k^w = \sigma_{t,k}^w \Phi^{-1}(\rho) \qquad c_k^w = t_k^w + \sigma_{t,k}^w \Phi^{-1}(\rho) \qquad \forall w \in W, k \in R_w \tag{14}$$

where $\Phi(x)$ is the standard normal cumulative function of random variable x.

3 SUE condition and variational inequality model formulation

In general, since different travelers have different attitudes toward risk, it is unreasonable to assume that the travel time reliability requirement ρ is identical to all travelers in a real road network. Therefore, in this paper, the OD traffic demand q_w between OD pair w is divided into I classes. The travel time reliability requirement ρ is identical in each user class i ($\forall i \in I$) and the travelers belongs to different user classes will have different travel time reliability requirements.

In a real road network, due to the complexity of network structure and a high degree of uncertainty in traffic conditions, travelers usually have partial information about the congestion status and the actual route travel time of the network. Therefore, it is necessary for a traveler to consider the perception error when choosing a route. In this paper, a perceived travel time budget is introduced as follows:

$$C_{k,i}^{\ w} = c_{k,i}^{\ w} + \xi_{k,i}^{\ w} \qquad \forall w \in W,\ k \in R_w,\ i \in I \tag{15}$$

$$c_{k,i}^{\ w} = t_k^{\ w} + \sigma_{t,k}^{\ w}\Phi^{-1}(\rho_i) \qquad \forall w \in W,\ k \in R_w,\ i \in I \tag{16}$$

where $C_{k,i}^{\ w}$, $c_{k,i}^{\ w}$, $\xi_{k,i}^{\ w}$, and ρ_i are, respectively, the perceived travel time budget, the actual travel time budget, the perception error term of user class i on route k between OD pair w, and the travel time reliability requirement of user class i. It is assumed here that $E[\xi_{k,i}^{\ w}] = 0$ and therefore, $E[C_{k,i}^{\ w}] = c_{k,i}^{\ w}$.

In reality, the OD traffic demand may be influenced by level of service on the network (Sheffi 1985). To take this phenomenon into account, for each user class i, the OD traffic demand $q_{w,i}$ is assumed to be a strictly monotonic decreasing function with respect to the expected minimal perceived travel time budget. In other words,

$$q_{w,i} = D_{w,i}(C_{w,i}) \qquad \forall w \in W,\ i \in I \tag{17}$$

where $D_{w,i}()$ and $C_{w,i}$ are, respectively, the OD traffic demand function and the expected minimal perceived travel time budget of user class i between OD pair w. $C_{w,i}$ can be computed by using the following equation.

$$C_{w,i}(\mathbf{c_w}) = -\ln\Sigma_{k\in R_w}\exp(-\theta_i c_{k,i}^{\ w})/\ \theta_i \qquad \forall w \in W,\ i \in I \tag{18}$$

where parameter θ_i is a constant related to the perception error of travelers.

We assume that the perception error terms on a traveler's routes between OD pair w are independently and identically Gumble distributed. Therefore, the probability $p_{k,i}^{\ w}$ for travelers in user class i to choose route k from the set of routes between OD pair w is given by:

$$p_{k,i}^{\ w} = \exp(-\theta_i c_{k,i}^{\ w})/\Sigma_r \exp(-\theta_i c_{r,i}^{\ w}) \qquad \forall w \in W,\ k \in R_w,\ i \in I \tag{19}$$

For travelers in each user class i ($\forall i \in I$), the stochastic user equilibrium condition can be written as:

$$f_{k,i}^{\ w} = q_{w,i}\, p_{k,i}^{\ w} \qquad \forall\ k \in R_w,\ w \in W,\ i \in I \tag{20}$$

where $f_{k,i}^{\ w}$ denotes the mean route flow of user class i on route k between OD pair w.

We formulate the travel time budget-based stochastic user equilibrium traffic assignment model with multiple user classes and elastic demand as an equivalent variational inequality problem as follows.

Find a mean route flow vector $\mathbf{f}^*$ and an OD demand vector $\mathbf{q}^* \in \Psi$, such that:

$$\sum_w\sum_k\sum_i\left(c_{k,i}^{w^*} + \ln\left(f_{k,i}^{w^*}/q_{w,i}^*\right)/\theta_i\right)\left(f_{k,i}^{w} - f_{k,i}^{w^*}\right) - \sum_w\sum_i D_{w,i}^{-1}\left(q_{w,i}^*\right)\left(q_{w,i} - q_{w,i}^*\right) \geq 0$$
$$\forall f_{k,i}^{w}, q_{w,i} \in \Psi \tag{21}$$

where the superscript "*" is used to designate the solution of the variational inequality problem; $D_{w,i}^{-1}()$ denotes the inverse function of the OD traffic demand function; and Ψ is the feasible set for the mean route flows and OD demands determined by the following constraints.

$$\Sigma_k f_{k,i}^{\ w} = q_{w,i} \qquad \forall w \in W,\ i \in I \tag{22}$$

$$f_{k,i}^{\ w} \geq 0 \qquad \forall\ k \in R_w,\ w \in W,\ i \in I \tag{23}$$

$$q_{w,i} \geq 0 \qquad \forall w \in W,\ i \in I \tag{24}$$

$$\Sigma_w \Sigma_k \Sigma_i\, f_{k,i}^{\ w} \delta_{a,k}^{\ w} = x_a \qquad \forall a \in \mathrm{A} \tag{25}$$

The proposed variational inequality problem can be solved by a number of route-based algorithms such as sequential quadratic programming or gradient projection methods etc.

4 A numerical example

The example network shown in Figure 1 includes six nodes, seven links, and one OD pair (from node 1 to node 6). Associated with each link are four numbers: the index, the free-flow travel time (h), the design capacity (pcu/h), and the worst degraded coefficient θ_a. The link performance function is given by Eq. (1) with α = 0.15 and β = 4. The following linear demand function is adopted.

$$q_i(C_i) = q_{i,max} - 500C_i \tag{26}$$

where $q_{i,max}$ is the maximum (or potential) OD demand. For ease of exposition, it is assumed that there are only two user classes and their potential OD demand are set to be $q_{1,max}$ = 2500 pcu/h and $q_{2,max}$ = 2000 pcu/h.

In this example, we first assume that the dispersion parameters for the two user classes are θ_1 = 10; θ_2 = 2. Then, we solve the proposed SUE model by using a route-based gradient projection algorithm under different combinations of the two user classes' travel time reliability requirements ρ_1 and ρ_2. Table 1 depicts the corresponding results, which include the traffic flow on each route, the OD demand of each user class, and the travel time budget of each route for each user class at the equilibrium point.

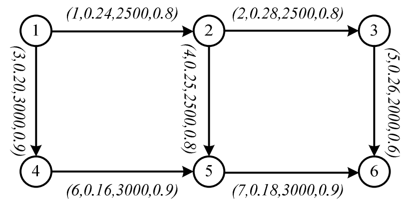


Figure 1. A six-node network for testing the proposed model

Table 1 Equilibrium flow patterns of the travel time budget-based SUE model

ρ_i	Route	Link sequence	Route flow	c_1(h)	c_2(h)	q_1	q_2
ρ_1=0.50 ρ_2=0.95	1	1-2-5	777	0.80	0.81	2211	1914
	2	1-4-7	1040	0.74	0.75		
	3	3-6-7	2308	0.61	0.62		
ρ_1=0.90 ρ_2=0.90	1	1-2-5	780	0.81	0.81	2207	1915
	2	1-4-7	1035	0.75	0.75		
	3	3-6-7	2308	0.62	0.62		

From Table 1, we can find that if the travel time reliability requirement increases, the OD demand will decrease and the corresponding route travel time budget will increase accordingly.

Acknowledgement

This research is sponsored by the National Natural Science Foundation of China under grant number 50578019 and the Ministry of Communications of China Application Foundation under grant number 2005319825050.

References

Bell, M. G. H., and Iida, Y. (1997). "Transportation network analysis." *John Wiley and Sons*, New York.

Lo, H. K., Luo, X. W., Siu, B. W. Y. (2006). "Degradable transport network: Travel time budget of travelers with heterogeneous risk aversion". *Transportation Research 40B*, 792-806.

Sheffi, Y. (1985). "Urban transportation networks: equilibrium analysis with mathematical programming methods". *Prentice-Hall*, Englewood Cliff, New Jersey.

Calibration of Lowry Model Using Immune Genetic Algorithm

Zhong-Bin Xiao[1] and Wei Wang[2]

[1] Transportation College, Southeast University, Postcode: 210096, Nanjing, Jiangsu, China; Cellphone +86-13584051701; email: xzb_seu@126.com
[2] Transportation College, Southeast University, Postcode: 210096, Nanjing, Jiangsu, China; email: wangwei@seu.edu.cn

Abstract: In order to investigate the suitability and potential benefits of applying land-use forecasting models in China, this paper presents an immune genetic algorithm for the calibration of Lowry model based on a maximum likelihood approach. The calibration procedure comprises three stages. In the first stage, an immune genetic algorithm is employed to calibrate the population and employment potentials, together with a coefficient associated with the travel impedance function in the study area. The second stage investigates the relationship between the calibrated potentials and various land-use variables, using a multivariate stepwise regression analysis. The third stage is model validation. A case study of Hu Zhou city, in Zhe Jiang province of China, was employed to demonstrate the performance of the proposed methodology. The results indicate that the calibrated Lowry model is acceptable for forecasting the future population and employment distribution in China.
Key words: *Land-use and transportation; Lowry model; Calibration; Immune genetic algorithm*

1 Introduction

Land use and transportation interaction is a dynamic process that involves changes over spatial and temporal dimensions between the two systems. Since the 1960s, many theories and models have been used to study land use and transportation interaction (Waddell et al., 2006). The Lowry model (Lowry, 1964) has been widely used for this purpose. It was developed to simulate location patterns of residential and service activities. A maximum likelihood approach was employed to calibrate the Lowry model(Putman and Ducca, 1978).

The maximum likelihood approach is a large-scale nonlinear optimization problem with a large number of variable and uncertain parameters. Recently, in order to get a better solution, the artificial intelligence methods, such as genetic algorithm (GA) (Wong, 1998), parallelized GA(Wong, 2001), neural networks (Rodrigue, 1997), etc, were proposed in many papers, and their promising performance was approved. But analyzing these methods, problem solutions are not very optimistic due to objections coming from these algorithms, e.g. local optimization solution, slow constringency speed etc. According to the disadvantages, this paper cites the theory of biological immune system (Huang, 1999), and constructs an immune genetic algorithm (IGA) to calibrate the Lowry model.

The remainder of the article is organized as follows. Section 2 describes the structure of the maximum likelihood function to calibrate the Lowry model. IGA for the calibration process is given in section 3. A case study is presented in section 4, where modeling results are given and interpreted. Finally, section 5 concludes the article with a brief summary and implications for further research.

2 The Lowry Model

The Lowry model uses three categories of activities: basic employment, service employment, and the household sector or residential population. Two spatial interaction functions form the basis of the Lowry model; the first distributes zonal employment to residences, and the second distributes service employment required by the residential population dependent on these employees. Beginning from an exogenously supplied distribution of basic employment, these interaction functions are used iteratively until no further increments to population are generated and a static equilibrium is achieved.

The parameters used to predict the population and employment distributions can be specified with the vector format as follows: $\mathbf{B}=(\beta_i, i=1,2,\ldots,N_P)$, $\mathbf{W}_P=(w_{Pik}, i=1,2,\ldots,N_P, k=1,2,\ldots,N_Z)$ and $\mathbf{W}_E=(w_{Ejl}, j=1,2,\ldots,N_E, l=1,2,\ldots,N_Z)$; and the results from the Lowry model as $\mathbf{P}=(P_{ik}, i=1,2,\ldots,N_P, k=1,2,\ldots,N_Z)$ and $\mathbf{E}=(E_{jl}, j=1,2,\ldots,N_E, l=1,2,\ldots,N_Z)$.

where

$$P_{ik}=\alpha\sum_{j=1}^{N_E}\sum_{l=1}^{N_Z}E_{jl}\frac{w_{Pik}\exp(-\beta_i c_{lk})}{\sum_{i=1}^{N_P}\sum_{k=1}^{N_Z}w_{Pik}\exp(-\beta_i c_{lk})}$$

$$E_{jl}=\frac{1}{\alpha}\sum_{i=1}^{N_P}\sum_{k=1}^{N_Z}P_{ik}\frac{w_{Ejl}\exp(-\beta_i c_{kl})}{\sum_{j=1}^{N_E}\sum_{l=1}^{N_Z}w_{Ejl}\exp(-\beta_i c_{kl})}$$

P_{ik} and w_{Pik} are, respectively, the number of and housing potential for household of category i in zone k; E_{jl} and w_{Ejl} are, respectively, the number of and employment potential for employment of category j in zone l; N_P, N_E and N_Z are, respectively, the number of household categories, employment categories and spatial zones in the study area; c_{lk} is the travel cost from zone l to zone k; β_i is the travel impedance coefficient of household category i; α is the regional household to employment ratio. The household and employment potential are defined as measures of the relative zonal attractiveness to residents and employees, respectively, in the study area. These potentials state only the aggregate effects and it is believed that these attractions are contributed by the level of development or the land-use intensities of a zone. The above-mentioned procedures can be summarized by

$$(\mathrm{P},\mathrm{E})=\Gamma(\mathrm{B},\mathrm{W}_P,\mathrm{W}_E) \tag{1}$$

Let $\overline{P}_{ik}$ and $\overline{E}_{jk}$ be respectively the observed population for population category *i* and observed employment for employment category *j*, at zone *k*. Assuming that the population or employment in a zone for a particular category is normally distributed around the results of the Lowry model. The probability of obtaining the observed values can be calculated as

$$P_r\left\{\overline{P_{ik}} = P_{ik}\right\} = \frac{1}{\sigma_{Pik}\sqrt{2\pi}}\exp(-(\overline{P_{ik}} - P_{ik})^2 / 2\sigma_{Pik}^2)$$

$$P_r\left\{\overline{E_{jk}} = E_{jk}\right\} = \frac{1}{\sigma_{Ejk}\sqrt{2\pi}}\exp(-(\overline{E_{jk}} - E_{jk})^2 / 2\sigma_{Ejk}^2)$$

Where σ_{Pik}^2 and σ_{Ejk}^2 are respectively the variances of the observed population for population category *i* and observed employment for employment category *j*, at zone *k*. Further assuming that the occurrence of the observed values are independently distributed, the likelihood function *L* is

$$L = \prod_{i=1}^{N_P}\prod_{k=1}^{N_Z}\frac{1}{\sigma_{Pik}\sqrt{2\pi}}\exp(-(\overline{P_{ik}} - P_{ik})^2 / 2\sigma_{Pik}^2)\prod_{j=1}^{N_E}\prod_{k=1}^{N_Z}\frac{1}{\sigma_{Ejk}\sqrt{2\pi}}\exp(-(\overline{E_{jk}} - E_{jk})^2 / 2\sigma_{Ejk}^2)$$

The problem is solved by finding a set of parameters (**B,** $\mathbf{W}_P$, $\mathbf{W}_E$) such that the household and employment distribution obtained from the Lowry model as specified in equation (1) maximizes the objective function ln*L*, i.e.

$$\underset{\mathrm{B,W_P,W_E}}{Maximize} \ln L = -\left[\sum_{i=1}^{N_P}\sum_{k=1}^{N_Z}\frac{(\overline{P_{ik}} - P_{ik})^2}{\sigma_{Pik}^2} + \sum_{j=1}^{N_E}\sum_{k=1}^{N_Z}\frac{(\overline{E_{jk}} - E_{jk})^2}{\sigma_{Ejk}^2}\right] \qquad (2)$$

Subject to

$$\mathrm{(P,E)}=\Gamma(\mathrm{B,W}_P,\mathrm{W}_E)$$

3 Application IGA to Calibrate the Lowry Model

Immune genetic algorithm (IGA) is put forward by adding the theory immune system to genetic algorithm in order to calibrate the Lowry model. IGA regards evolutional individuals as antibodies and objective function as antigens. With its ability of self-regulate， the antibody population achieves a good regulation of dynamic balance between individual diversity and population convergence after encountering foreign invading (Ma, 2006). Main processes using IGA to calibrate the Lowry model are followed as:

Step 1 *Generation the initial antibodies*

Gene coding adopts the decimal coding rather than binary coding. This method avoids the process of frequent coding and recoding, increases the speed and accuracy of calculation, and has advantage to solve the large-scale optimization problems. The control variables including the travel impedance coefficient **B**, the population potential $\mathbf{W}_\mathrm{P}$ and the employment potential $\mathbf{W}_\mathrm{E}$ are coded into antibodies. **B** adopts the real number code, while $\mathbf{W}_\mathrm{P}$ and $\mathbf{W}_\mathrm{E}$ adopt the integer code. The initial antibodies population of control variables is generated randomly from the set of uniformly distributed control variables ranging over their upper and lower limits.

Step 2 *Calculation the affinities*

The affinity A_{bgw} between antibodies and antigens is calculated as follows:

$$A_{bgw} = \mu\left[f(v)\right].$$

Where $f(v)$ is objective function, $\mu(x)$ is the monotony function of x. Here, the negative of the reciprocal of objective function is used to express affinity:

$$A_{bgw} = -\frac{1}{\ln L}.$$

Affinity between antibodies can reflect their analogical extent. In other words, a greater affinity value indicates a greater similarity between antibodies. The affinity between two antibodies *w* and *v* can be expressed by followed expression:

$$B_{w,v} = 1/(1+H_{w,v}).$$

Where $H_{w,v}$ is Euclidean space between antibody *w* and *v*, and it can be calculated as follows:

$$H_{w,v} = \left\{\sum\left[(B_{iw}-B_{iv})(B_{iw}-B_{iv})^T + (W_{Piw}-W_{Piv})(W_{Piw}-W_{Piv})^T + (W_{Eiw}-W_{Eiv})(W_{Eiw}-W_{Eiv})^T\right]\right\}^{\frac{1}{2}}$$

Where $B_{iw}, B_{iv}, W_{Piw}, W_{Piv}, W_{Eiw}, W_{Eiv}$ are values of the *ith* item of antibody *w* and *v*.

Step 3 *Calculation the density of antibodies*

The density of antibody w, c_w can be defined as $c_w = \left(\sum_{v=1}^{N} B_{w,v}\right)\Big/N$. Where N is the number of antibodies. According to their corresponding densities, antibodies are ranked in ascending order, and the antibodies with great density are eliminated by the abandoning rate. The new ones by random generation are substituted for the eliminating antibodies. Suppressing the high density antibodies can greatly keep the diversity of population, and avoid trapping into the local optimal solution.

Step 4 *Selection calculation*

The selection operation is executed with the rank method, while adopting the rank method, the fitness of antibody is only decided by its order in the population rather than by its actual value of the objective function. The fitness of antibody *w* is composed of two sections: the affinity A_{bgw} and the density c_w, that is,

$$p = \alpha A_{bgw} + (1-\alpha)c_w.$$

Where α is a proportional factor, $0<\alpha<1$. The selecting operation can retain the antibodies with the small affinity to antigen by activating and suppressing based on the densities of antibodies, which can ensure the diversity of population and improve the convergence near the optimal solution.

Step 5 *Crossover and mutation*

The antibodies with smaller fitness have higher probability to be selected by the rank method, and to carry out the crossover and mutation operation. The crossover operator adopts the middle recombination suitable for the real variables, after crossover, the antibodies possess a good distribution. The mutation operation adopts the even real number mutation.

Step 6 *Judging the conditions of end*

If the number of end generation of evolution is arrived and the average density of antibodies is stable, then the program is ended, else goes back to step 2.

4 Numerical Analysis of Example

Hu Zhou city, in Zhe Jiang province of China, is used as a case to demonstrate the performance of the IGA. The travel characteristics survey (TCS) was carried out by Southeast University to provide information for transportation planning in 2004. It consists mainly of a large scale home interview survey to collect information on household, personal and trip characteristics. Some land use data are measured from the outline zoning plans for the present study.

Using MATLAB6.5 toolbox, two kinds of algorithm are performed. The first is simple genetic algorithm (SGA) with only the GA operators and the second is IGA adding the immune operators. The city is divided into 32 zones. The number of unknowns is 65. In the IGA, the size of population is 100 and the mutation rate is set as 1/500,000. The crossover rate is 0.8 and the maximum number of generation is taken as 1000. The fitness function proportional factor α is set as 0.98. Fig.1 shows the evolving curves of the SGA vs. IGA. It is apparent that the proposed method is superior to the SGA.

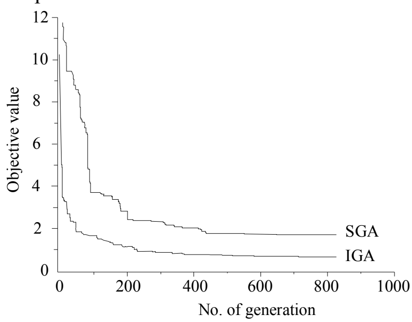


Fig. 1. Evolving curves of SGA vs. IGA

By minimizing the discrepancy between the observed and modeled distributions of population and employment, a set of zonal population potential and employment potential are obtained. They are set as dependent variables and a set of ten types of land use data are set as independent variables in the stepwise regression analysis. The coefficient of determination, R^2, is employed as a statistical measure for the assessment. Table 1 summarizes some performance index. It was observed that very high R^2 were obtained in IGA. The resulting potentials are reliable and able to replicate the base year conditions. It is shows that the Lowry model can be applied to forecast future distributions of population and employment based on different land use zoning policies.

Table 1 calibration results of the Lowry model

	Population R^2	Employment R^2	Travel impedance coef. β
SGA	0.778	0.753	5.75×10^{-4}
IGA	0.929	0.899	9.03×10^{-4}

5 Conclusions

A novel algorithm, immune genetic algorithm (IGA), is proposed to calibrate the Lowry model, which can be applied to forecast future population and employment distribution. IGA, compared with SGA, possesses a better global convergence and a quicker calculation speed. The results indicate that the calibrated Lowry model is acceptable for forecasting the future population and employment distribution due to different land-use planning and policy schemes, as well as changes in transportation systems.

References

Huang, S. (1999). "Enhancement of thermal unit commitment using immune algorithms based optimization approaches." *Electrical Power and Energy System,* 21, 245-252.

Lowry, I.S. (1964). *A Model of Metropolis,* The Rand Corporation, Santa Monica.

Ma, J., Zou, H., and Gao, L.Q. (2006). "Immune genetic algorithm for vehicle routing problem with time windows." *Proceedings of the Fifth Intl. Conf. on Machine Learning and Cybernetics,* IEEE, Dalian, China.

Putman S.H. and Ducca F.W. (1978). "Calibrating urban residential models 1: procedures and strategies." *Envir. and planning A,* 10, 633-650.

Rodrigue, J.-P. (1997). "Parallel modeling and neural networks: an overview for transportation/land use systems." *Transport. Res. Part C (Emerging Technologies),* 5C, 259-271.

Waddell, P. et al. (2006). "Incorporating land use in metropolitan transportation planning." *Transport. Res. Part A (Policy Prac.),* doi:10.1016/j.tra.2006.09.008.

Wong, C.K., Wong, S.C., and Tong, C.O. (1998). "A new methodology for calibrating the Lowry model." *J. Urban Plng. and Devel.*, ASCE, 124, 72-91.

Wong, S.C., Wong, C.K., and Tong, C.O. (2001). "A parallelized genetic algorithm for the calibration of Lowry model." *Parallel computing,* 27, 1523-1536.

Signal Control for Oversaturated Intersections Using Fuzzy Logic

Lin Zhang, Ph.D., P.E.[1]; Honglong Li, Ph.D., P.E.[2]; Panos D. Prevedouros, Ph.D.[3]
[1] Cambridge Systematics, Inc., 555 12th St., Suite 1600, Oakland, CA 94607, USA; PH (510)873-8700; FAX (510)873-8701; e-mail: lzhang@camsys.com
[2] PB Americas, Inc., 1001 Bishop Street, Suite 2400, Honolulu, HI 96813, USA; PH (808)566-2211; FAX (808)528-2368; e-mail: lih@pbworld.com
[3] Department of Civil and Environmental Engineering, University of Hawaii at Manoa, 2540 Dole Street, 383, Honolulu, HI 96822, USA; PH (808)956-9698; FAX (808)956-5014; e-mail: pdp@hawaii.edu

ABSTRACT

The fuzzy logic controller (FLC) presented in this paper simulates the control logic of experienced human traffic controllers such as police officers who supersede signal controls at over-saturated intersections during special events. Given real-time traffic information, the FLC controller decides on whether to extend or terminate the current green phase based on a set of fuzzy rules. A microscopic simulator, the Intersection Control Simulator (ICS), was developed to facilitate the evaluation of the proposed FLC strategy. The FLC strategy was compared with pretimed and actuated control strategies using a typical intersection with varying traffic volume levels. Based on delay, speed, % stops, time in queue and throughput-to-demand ratio statistics, the FLC strategy produced significant improvements over pretimed and actuated control strategies under heavy traffic volumes. This indicates that FLC has the potential to improve operations at over-saturated intersections.

INTRODUCTION

Intersections are common bottlenecks in roadway systems. It has been recognized that signal improvement is one of the most cost-effective methods to reduce congestion. Most signal controls are implemented with either pretimed controls or actuated controls. A pretimed controller repeats preset signal timings derived from historical traffic patterns. An actuated controller computes phase durations based on real-time traffic demand obtained from detection. Actuated control has an efficiency handicap rooted to its simple operating principle of phase extension until a preset maximum is reached. Its performance deteriorates under heavy traffic conditions as it extends green phases to the maximum green times on all phases when high traffic volumes occur on all approaches (Tarnoff et al. 1981).

Adaptive control is designed to take account of the traffic conditions for the whole intersection. It has the ability to adjust signal phasing and timing settings in response to real-time traffic demands at all or some approaches. However, current control methods typically cannot accommodate heavy traffic well. Human control logic may be superior to existing signal control logic because of its ability to realize the prevailing traffic demands. This paper explores the potential application of fuzzy logic on traffic signal control under over-saturated traffic conditions.

LITERATURE REVIEW

The first known attempt to apply fuzzy logic in traffic control was made by Pappis and Mamdani (1977). They simulated an isolated signalized intersection composed of two one-way streets with two lanes in each direction without turning traffic. The fuzzy controller reduced average vehicle delay compared to an actuated controller. Kelsey and Bisset (1993) also simulated a simple two-phase signal control of an isolated intersection with one lane on each approach. The fuzzy logic control performed better than both pretimed and actuated control especially when the traffic flow between different directions was highly uneven. Niittymaki and Pursula (2000) also simulated an isolated intersection and found that fuzzy logic controller resulted in shorter vehicle delay and lower % stops especially when the traffic volume was heavy.

Trabia, et al. (1999) designed a fuzzy logic controller for a signalized intersection with left-turning traffic. Traffic volumes and queue lengths counted by detectors were used in a two-stage fuzzy logic algorithm to determine whether to extend or terminate the current signal phase. Fuzzy logic control lead to an average of 9.5% decrease in delay and 1.3% reduction in stops compared to actuated control. Nakatsuyama et al. (1984) applied fuzzy logic to control two adjacent intersections on an arterial with one-way movements. Fuzzy control rules were developed to determine whether to extend or terminate the green signal for the downstream intersection based on the upstream traffic.

All the aforementioned research concepts were tested with simulation. Few field tests of fuzzy logic control have been conducted to date. In 2001, Niittymaki (2001) presented a field test of a simple two-phase fuzzy signal controller. The results showed that the fuzzy logic controller performed better than vehicle-actuated control in terms of delay, % stops and savings in fuel and emissions.

STUDY OBJECTIVE

The research reviewed above generally reported a better performance of fuzzy logic controllers compared to pretimed and actuated controllers. However, most of the reviewed research involved either one-way streets or intersections without turning movements. In addition, fuzzy rules were determined mostly by traffic conditions on the subject approaches without taking into account the traffic conditions on competing approaches. The objective of this study is to design and evaluate a fuzzy logic algorithm to control over-saturated intersections of two-way streets with left-turning movements and compare it with pretimed control and actuated control using microscopic simulation.

FUZZY SET AND FUZZY LOGIC

Fuzzy set theory is suitable for systems that involve imprecise and vague information. The fuzzy set theory was first introduced by Zadeh in 1965 as a mathematical method for representing vagueness in everyday life (Zadeh 1965). Fuzzy logic (Zadeh 1973) is a mathematical representation of human concept formulation and reasoning. Fuzzy logic has been applied to practical problems with controls and decisions which involve the imprecise human reasoning process.

Fuzzy Set

In contrast to the classical set theory, fuzzy sets admit intermediate values of class membership. A fuzzy set is represented by a membership function which expresses the degree that an element of the universal set belongs to the fuzzy set. The most commonly used range of membership values is the unit interval [0,1].

Fuzzy Logic

Introducing a concept he called "approximate reasoning," Zadeh (1973) successfully showed that vague logical statements enable the formation of algorithms that can use vague data to derive vague inferences. Fuzzy logic makes it possible to compute with words, which enables complex analysis reflecting the human thinking process. Each fuzzy logic system can be divided into three elements: fuzzification, fuzzy inference and defuzzification (Klir and Yuan 1995). Input data are most often crisp values. Fuzzification maps crisp numbers into fuzzy sets. The fuzzifier decides the corresponding membership grades from the crisp inputs. The resulting fuzzy values are then entered into the fuzzy inference engine which is based on a fuzzy rule base which contains a set of If→Then fuzzy rules. A commonly used defuzzification strategy for continuous membership functions is the centroid (center of area) method.

METHODOLOGY

Signal control is basically a process for allocating green time among conflicting movements. Alternatively, signal control is a process for determining whether or to extend or terminate the current green phase. The proposed fuzzy logic controller (FLC) works in the same way but it is significantly different from actuated control. Actuated control extends green time based on an extension interval, a maximum green time and the vehicular actuations, but no examination of the conditions on conflicting movements and no optimization are involved.

The proposed fuzzy logic controller determines whether to extend or terminate the current green phase based on a set of fuzzy rules. The fuzzy rules compare traffic conditions with the current green phase and traffic conditions with the next candidate green phase. The set of control parameters is:

QC = Average **q**ueue length on the lanes served by the **c**urrent green, in veh/lane.

QN = Average **q**ueue length on lanes with red which may receive green in the **n**ext phase, in veh/lane.

AR = Average **a**rrival **r**ate on lanes with the current green, in veh/sec/lane

$\mathbf{T_{MIN}}$ = minimum green time for each phase, in sec.

$\mathbf{T_{MAX\_TH}}$ = maximum green time for through lanes which can vary for different approaches, in sec.

$\mathbf{T_{MAX\_LT}}$ = maximum green time for left-turn lanes which can vary for different approaches, in sec.

The fuzzy logic controller determines whether to extend or terminate the current green phase after a minimum green time of T_{MIN} has been displayed. If the green time is extended, then the fuzzy logic controller will determine whether to extend the green after a time interval Δt. The interval Δt may vary from 0.1 to 10 sec. depending on the controller processor speed; Δt = 3 sec. in this study. If the fuzzy logic controller determines to terminate the current phase, then the signal will go to the next phase. If not, the current phase will be extended and the fuzzy logic controller will make the next decision after Δt and so forth until the maximum green time is reached. The decision making process is based on a set of fuzzy rules which takes into account the traffic conditions with the current and next phases. The general format of the fuzzy rules is as follows:

If {QC is X_1} and {AR is X_2} and {QN is X_3} Then {E or T}.

where,

X_1, X_2, X_3 = natural language expressions of traffic conditions of respective variables

E = Extension of green phase

T = Termination of green phase

QC and QN are divided into four fuzzy sets: “short,” “medium,” “long” and “very long.” AR is divided into three fuzzy sets: “low,” “medium” and “high.” The input data (traffic conditions) are first fuzzified using the proposed fuzzy sets for QC, QN, and AR. Then the fuzzified input data are entered into the fuzzy inference system which is composed of a set of fuzzy rules. The max-min composition method (Klir and Yuan 1995) is applied for making inferences.

SIMULATION TOOL

A simulation software, Intersection Control Simulator (ICS), was developed based on the intersection simulator NIT (Li et al. 2004). NIT was originally developed to evaluate a new adaptive control strategy TACOS (Li and Prevedouros 2004). ICS is a microscopic, stochastic, interval-oriented traffic simulator programmed in C++. ICS was designed to be able to simulate the intersection operation under pretimed, actuated, and FLC control.

ICS scans the traffic system and summarizes measures of effectiveness (MOE) at each time interval. MOE for each lane are necessary to evaluate the performance of different signal control strategies. Link-based, approach-based and intersection-wide MOE can be derived from lane-based MOE using weighted aggregation. The lane-based MOE estimators of ICS include a throughput estimator, a speed estimator, a delay estimator and a queue estimator. A comprehensive test showed that ICS is a valid tool for testing various intersection control strategies (Li et al. 2004; Li and Prevedouros 2004).

CASE STUDY

FLC was evaluated against pretimed and actuated control strategies. The MOE used in the evaluation were (1) network delay, (2) network speed, (3) % stops, (4) network time in queue, and (5) network throughput-to-demand ratio.

The geometric configuration of the case study signalized intersection consisted of four approaches, with three lanes on the north- and south-bound approaches: one through and right lane, one through lane and one left turn lane, and four lanes on the east- and south-bound approaches: one right turn lane, two through lanes and one left-turn lane. Left-turn lanes were made long enough to accommodate left-turning traffic queues. Traffic volumes were varied from 20% to 100% of the highest volume. The 100% traffic volume level represents a condition where two conflicting movements have a volume-to-capacity ratio greater than 1.0, thus, the intersection is substantially over-saturated.

The pretimed signal timings were optimized with TRANSYT-7F using “delay and stops” as the objective function. Vehicle extension intervals of 2, 3 and 4 sec. were tried for the actuated control and the value that produced the best performance in terms of speed and delay was used to compare actuated control with FLC. In addition to presence detectors at the stop line, all lanes have passage detectors 15 m (50 ft.) upstream the stop line. These detectors place calls when the phase is active.

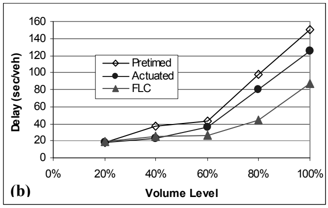
Because of the stochastic nature of simulation models, each simulation run may produce different results. For this study, 20 simulation runs were performed for each scenario, resulting in a total of (5+2)×5×20 = 700 simulation runs. The average values of delay, speed, % stops, time in queue, and throughput-to demand ratio from the 20 simulation runs were averaged to obtain the final MOEs.

FLC is compared with pretimed and actuated control as summarized in Figure 1. The summary table in Figure 1(a) shows average proportional differences of FLC's MOE over pretimed and actuated control for the test intersection. FLC produced the best performance for all MOE examined.

(a)

% Difference

	Delay	Speed	% Stops	Queue Time	Throughput
FLC vs. Pretimed	-42%	21%	-7%	-50%	2%
FLC vs. Actuated	-28%	10%	-2%	-44%	1%



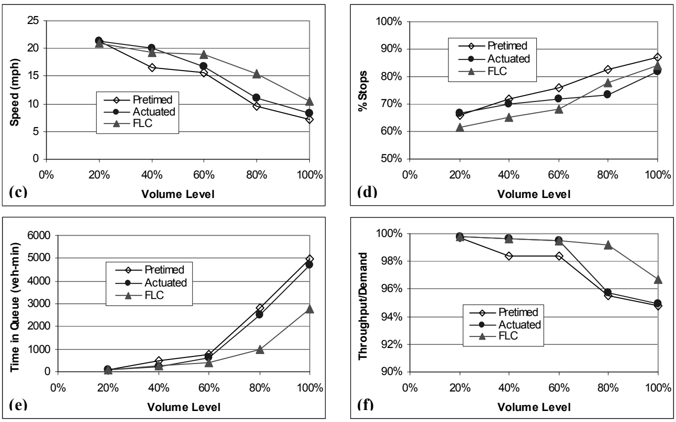


FIGURE 1. Results of case study.

Further inspection of the delay, speed, % stops, time in queue, and throughput-to-demand ratio (Figure 1) show that:

- <u>Delay</u>: FLC produced the lowest delay at the 60%, 80%, and 100% volume levels. The improvements are more significant with increasing traffic volume. Both pretimed and actuated controls produced lower delay than FLC at the 20% volume level.
- <u>Speed</u>: FLC produced the highest speed at the 60%, 80% and 100% volume levels. Both pretimed and actuated controls produced a higher speed than FLC at the 20% volume level.
- <u>% stops</u>: FLC produced a lower % stops than pretimed control at all levels of traffic volume. FLC also produced a lower % stops than actuated control at the 20%, 40%, and 60% volume levels. However, actuated control performed better than FLC at the 80% and 100% volume levels.
- <u>Time in Queue</u>: FLC produced the lowest time in queue at the 60%, 80% and 100% volume levels. Both pretimed and actuated controls produced shorter times in queue than FLC at the 20% volume level.
- <u>Throughput-to-demand ratio</u>: FLC consistently produced higher throughput-to-demand ratios than pretimed and actuated control at all

volume levels. The improvement is more pronounced at higher levels of traffic volume.

Overall, the FLC strategy produced substantial improvements over the pretimed and actuated control strategies under heavy traffic conditions.

CONCLUSION

A basic fuzzy logic control (FLC) algorithm for full intersections with two-way streets and left-turn lanes was developed. The FLC controller makes the decision whether to extend or terminate the current green phase based on a set of fuzzy rules and real-time traffic information. The microscopic intersection control simulator (ICS) facilitated the evaluation of the proposed FLC strategy. FLC was compared with pretimed and actuated control strategies using a typical intersection with varying traffic volume levels. FLC showed substantial improvements over pretimed and actuated controls for all MOE except % stops under heavy traffic volumes. Overall, the simulation results indicated that FLC has the potential to improve operations at over-saturated intersections.

REFERENCES

Kelsey, R. L., and K. R. Bisset. (1993). "Simulation of Traffic Flow and Control Using Fuzzy and Conventional Methods." *Fuzzy Logic and Control: Software and Hardware Applications*, Prentice Hall, Englewood Cliffs, New Jersey, pp. 262-278.

Klir, G. J., and B. Yuan. (1995). *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Englewood Cliffs, New Jersey.

Li, H., and P. D. Prevedouros. (2004). "Traffic Adaptive Control Integrated with Phase Optimization: Model Development and Simulation Testing." *Journal of Transportation Engineering,* ASCE, Vol. 30, No. 5, 594-601.

Li, H., P. D. Prevedouros, and L. Zhang. (2004). "Development and Testing of the NETSIM, INTEGRATION and TACOS Simulator, NIT." *The 83rd Annual Meeting of TRB*, National Research Council, Washington D.C.

Nakatsuyama, M., H. Nagahashi, and N. Nishizuka. (1984). "Fuzzy Logic Phase Controller for Traffic Junctions in the One-Way Arterial Road." *Proceedings of the IFAC Ninth Triennial World Congress*, 2865-2870.

Niittymaki, J. (2001). "Installation and Experiences of Field Testing a Fuzzy Signal Controller." *European Journal of Operation Research*, Vol. 131, 273-281.

Niittymaki, J., and M. Pursula. (2000) "Signal Control Using Fuzzy Logic." *Fuzzy Sets and Systems*, Vol. 116, 11-22.

Pappis, C. P., and E. H. Mamdani. (1977). "A Fuzzy Logic Controller for a Traffic Junction." *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-7, No. 10, 707-717.

Tarnoff, P. J., and P. S. Parsonson. (1981). *Selecting Traffic Signal Control at Individual Intersections*, NCHRP Report 233, TRB, National Research Council, Washington, D.C.

Trabia, M. B., M. S. Kaseko, and M. Ande. (1999). "A Two-Stage Fuzzy Logic Controller for Traffic Signals." *Transportation Research*, Part C, Vol. 7, 353-367.

Zadeh, L. A. (1965). "Fuzzy Sets." *Information and Control*, Vol.8, 338-353.

Zadeh, L.A. (1973). "Outline of a New Approach to the Analysis of Complex Systems and Decision Processes." *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-3, No. 1, 28-44.

Research on Spatial-Temporal Features of Urban Freeway Congestion

By Yuan Hao,[1] Ying Wang,[2] P.E., M. ASCE, Li-jun Sun,1 and Tian-dong Xu[1]
[1]School of Transportation Engineering, Tongji University, Room 306, 4800 Cao'an Road, Shanghai, 201804, P. R. China; PH (86) 21 6958-8634. email: hao_silvia@126.com, ljsun@mail.tongji.edu.cn, xtdtj163@163.com.
[2] Dewberry, 15 E. 26th Street, New York, NY 10010, USA, PH (212) 685-0900; FAX (212) 685-2345, email: ywang@dewberry.com

Abstract: Relieving traffic congestion on roadway has always been one of the prior concerns of traffic engineering. Thoroughly understanding the factors which cause traffic jams plays the fundamental role in planning, design, construction and management of roadway facilities. This paper focuses on experimental observations and theoretical analysis of urban freeway congestion due to complex features of traffic flow, shock wave and state transition. Real time detected data were collected to calculate shock wave speed and establish vehicle trajectory models. A typical urban freeway segment in the center of Shanghai City is used in this analysis. As a result, conclusions have been drawn and the theory of shock wave speed could be used to help establishing traffic control strategies and to control queue length.

Keywords: urban freeway; capacity gap; state transition; shock wave; traffic control.

Introduction

Previous researches have explained the causes and characteristics of traffic congestion. B.S.Kerner (1997) thought that 25veh/km～40veh/km of density was a possible domain for phase transition from free flow to congested flow. Shy Bassan (2005) used 80km/h as speed threshold for breakdown. Yaping Zhang (2005) considered that traffic congestion was not a bit-by-bit process but a catastrophe. As for traffic wave and disturbance, Edie and Baverez (Taewan Kim et al. 2004) had observed the bottleneck section of Holland Tunnel in the city of New York and found that the density oscillates between 39 veh/km and 87 veh/km and the traffic wave speed was 14~17 km/h. Zheng Wu, (2002) studied the start-up wave and jam-up wave of intersection flow: start-up wave was about 6.08~6.97m/s (21.85km/h~25.09km/h); the equation used to calculate the jam-up wave is $N=5.3264q^{1.1327}$, where, N is the backward transmitting speed of shockwave, q is the upstream traffic flow. Recent experimental observations and theoretical analysis have shown that the features of traffic congestion are more complicated than our former understanding.

This paper is divided into four sections: Section one delineates the freeway segment in the city of Shanghai, China, and specifies collected data for analysis.

Section two analyzes and demonstrates some complex features of traffic flow using real-detected data. Section three provides recommendations for design of traffic flow control strategies. Section four summarizes the main conclusions.

Delineation of Shanghai Nanbei Freeway and data

Figure 1 is a map of the Nanbei Freeway in the center of Shanghai City. This freeway segment contains 30 pairs of inductive loops within 13.5km in length. Inductive loop detectors are installed with a spacing of 250 to 500 meters. They record vehicle counts, occupancies and time mean speeds in each lane at 20-second intervals. The detectors are labeled from "NBXX30" to "NBXX01" along the southbound traveling direction. Ramp metering is not installed on this facility. The lanes are numbered from left to right, as lane 1 to lane 4. Figure 2 shows the plain view of the Nanbei Freeway southbound segment. Data were taken on April 10th, 2006, and April 14th, 2006.

Complex features of traffic flow

1. Capacity gap and state transition

The area containing "turning point" in the fundamental figure (Fig. 7a) is considered as the capacity area. There are gaps at some locations in this area and capacity can not be reached due to the following causes:

① If traffic demand is insufficient, traffic flow will not fall into congestion, so there is no turning point, but a nearly straight line in the free flow area in the fundamental figure.

② If there are influences from on-ramps, downstream bottlenecks, and traffic incidents, capacity gap may occur. Normally, downstream bottleneck is the main factor that results in capacity gap at upstream sections. When traffic congestion occurs at downstream, shockwave will spread to upstream. Drivers at upstream don't have as much reaction time as those at downstream do, so they are forced to decelerate sharply to join the downstream queue with low speed and a capacity gap forms. When queue is cleared away, drivers start to accelerate gradually, but if traffic demand is insufficient, capacity can not be reached.

Sometimes, speed gap can be more easily observed than capacity gap because high volume is easy to attain in a relatively lower speed, but critical speed is not, which depends on drivers' skill and sensitivity.

If there is a capacity gap, state transition will occur.

Traffic flow state within a basic freeway segment can be categorized into five flow states (Fig. 3): state 1 is free flow; state 2 is quasi-free flow; state 3 is critical flow; state 4 is congested flow; state 5 is jammed flow. Without external influence and with the increase of the traffic demand, traffic flow varies in a sequence from state 1→state 2→state 3→state 4→state 5 (i.e. arrow *a* in Fig. 3). Under this condition, traffic density increases gradually, and when it reaches a critical value, traffic congestion, which is called "active congestion" occurs. When there is a severe perturbation in a "sensitive" flow (like the top of state 2), traffic density increases sharply, and speed "fall" occurs. This phenomenon is named as "state transition", and this kind of congestion as "passive congestion". "State transition" is hard to predict, for all the parameters change suddenly before reaching critical values. Similar phenomena appear when shockwave spreads to upstream stations from its origin. It has been observed that traffic flow jumps from state 2 to state 4 (i.e. arrow b in Fig. 3) and transfers to state 3 or state 5, and finally goes back to state 2. In flow-occupancy figure, there is usually a "gap", called capacity gap, at state 3 and the top of state 2 in these zones. See Fig. 4.

2. Shock wave

Shock wave is a term of traffic wave, which is formed at the boundary between upstream traffic flow with high speed and downstream traffic flow with low speed. Shock wave speed is the slope of secant in the fundamental figure that connects a point in the free flow side to a point in the congested flow side (Fig. 7 a). Shock wave speed w can be obtained from equation (1).

$$w = (Q_1 - Q_2)/(k_1 - k_2) \quad \ldots\ldots\ldots\ldots\ldots \text{(1)}$$

Where, Q_1, Q_2 are upstream and downstream traffic volumes respectively; k_1, k_2, are corresponding traffic densities; Shock wave spreads downstream if w is positive, and upstream if w is negative, it stands still if w is zero.

Shock wave usually occurs when state transition occurs. Figure 5 presents typical shock waves recorded on April 10^{th} and 14^{th}, 2006.

Occupancy can be detected by inductive loops directly, while density can not. However, the relationship between occupancy and density (Transportation Research Board, 1999) can be obtained from the equation: $occ=(L+d)k$. Where, occ is occupancy in percentage， L is the average length of a vehicle in km, and d is the length of a loop detector in km, k is the density in veh/km.

(L+d) can be calculated by the following equations. For stable traffic flow condition, $q=kv$,,where $k=occ/(L+d)$, so $(L+d)=v.occ/q$. Usually, the vehicle composition of traffic flow does not change with time, but in absolute free flow and congested flow, (L+d) is very different from the normal value. Most of the time, (L+d) is reliable under 10% to 30% occupancy. With this method, shock wave speed at four stations in figure 5 can be calculated:

a): w_1 = -15.76 km/h; b): w_2= - 4.81 km/h; c): w_3 = -18.19 km/h; d): w_4 = -6.7 km/h

Figure 6 shows the vehicle trajectories in this freeway section, the time interval between two neighboring lines is one minute. NBXX30 corresponds to zero meter in the space dimension (position). In figure 6-a), 1~211 corresponds to 6:20:00~7:30:00 in the time dimension; in figure 6-b), 1~241 corresponds to 11:50:00~13:10:00 in the time dimension. The two figures are drawn by using a "fictional car" method and real-time data from inductive loop detectors (Tiandong Xu, et al. 2007). The space-time plots are pointed every 20 seconds, which record actual travel time of vehicles.

From figure 6, the shock wave speed can be calculated by equation (2), where, $x1$, $x2$ are two locations on a shock wave line (red lines in figure 6); $t1$, $t2$ are the corresponding time.

$$w = (x1 - x2)/(t1 - t2) \quad \text{……………..} \quad (2)$$

Shock wave speeds w_1, w_2, w_3, w_4 are calculated as follow:
w_1 = -18.15 km/h; w_2= - 3.61 km/h; w_3 = -21.6 km/h; w_4 = -4.77 km/h.

The calculated shock wave speed from equation (1) in the Flow-Density plane is called "point speed", and the transition speed between stations obtained from equation (2) in the Space-Time plane (i.e. vehicle trajectory) is called "space-mean speed". There is a slight difference between these two speeds, because the "point speed" of shock wave varies among stations where the inductive loops are located.

Identification and control of traffic congestion

The purpose to study traffic congestion and characteristics of traffic flow is to alleviate and control traffic congestion. Traffic control measures can be designed to improve traffic efficiency by thorough understanding of traffic flow features.

1. **Identification of congestion headstream**

Generally speaking, local traffic control measure, like ALINEA, is suitable for isolated bottlenecks, where traffic jam affects only a limited area around the bottleneck. However, since the distance between on-ramps on urban freeway is relatively short, the congested flow often spills back and joins up to form "gridlock". Therefore, identifying the headstream and primitive causes of traffic congestion plays especially important role in actualizing a traffic control measure.

2. **Traffic Control Measures**

The program in congestion management is queue avoidance and containment, therefore, it is important to know about the behavior of bottlenecks and the spatial extent of queues. This can be achieved through a combination of active control measures such as ramp metering and passive measures such as route guidance.

A queue often forms from an active bottleneck and spills kilometers back and it may even block an upstream divergent point, such as an off-ramp. As a result, the vehicles that don't need to pass through the bottleneck may also be affected. If upstream traffic demand is restricted, queues could be slowed down and would not block the off-ramp. If the off-ramp is blocked, information should be sent to drivers for another route choice beforehand. This illustrates the importance of queue containment and the need for techniques for predicting the distance spanned by queues.

Shock wave speed is related to the upstream traffic demand. Lower traffic demand can decrease the upstream propagating shock wave speed, and shock wave stands still if Q_1 is equal to Q_2 according to equation (1). Set L as the distance between an active bottleneck and upstream off-ramp; T as the period to stop the shock wave; $Q_1(t)$ as upstream traffic flow varying with t ; $k_1(t)$ as upstream traffic density varying with t ; and t=0 once shock wave occurs, function $\int_0^T (Q_1(t) - Q_2) \big/ (k_1(t) - k_2)\, dt < L$ and $Q_1(T)$=Q_2 hold, since we want to control the queue within length L and within period T by shrinking upstream traffic flow bit by bit. By controlling the time of upstream traffic demand, the queue speed can be lowered so as to avoid blocking the off-ramp (figure 7).

Conclusion

The conclusion of this paper comes to 3 points: (1) Full traffic capacity can not be reached in some locations due to influences of insufficient traffic demand, external disturbance, and downstream bottlenecks. (2) State transition occurs usually when capacity decreases, and two types of congestions are defined based on how they are formed. (3) Shock wave can be calculated in two ways. By restricting upstream traffic demand, shock wave speed can be lowered to constrain queue length. This method is useful in design of traffic control strategy.

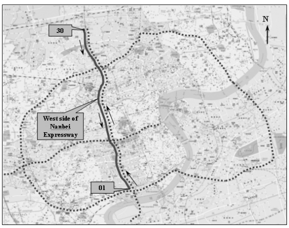


Figure 1 Map of Nanbei Freeway

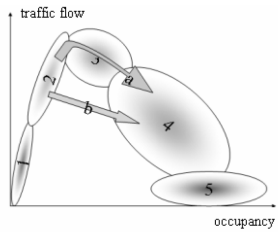


Figure 3 Traffic states

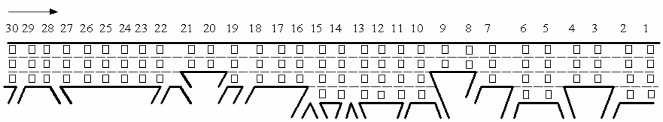


Figure 2 The plan view of the freeway segment

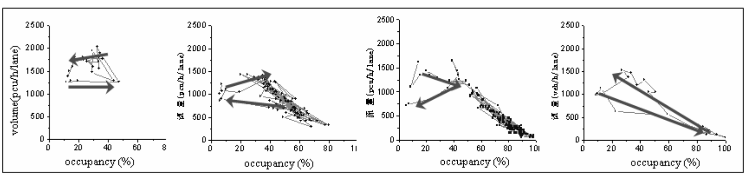


Figure 4 Styles of state transition

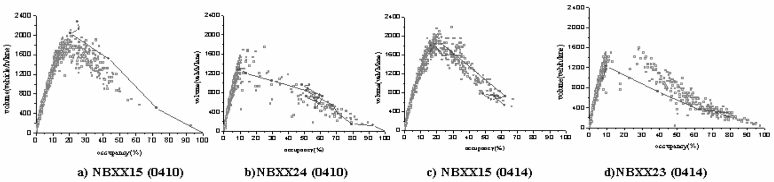
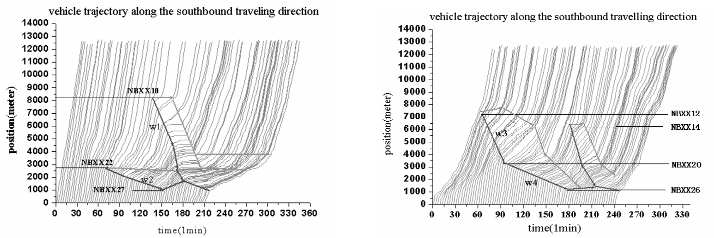


Figure 5 Shock wave



a) 6:20:00~7:30:00, 10th, April,2006 b) 11:50:00~13:10:00,14th,April,2006

Figure 6 Vehicle trajectory in space-time figure

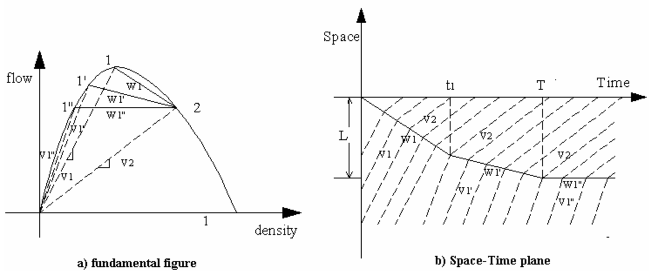


Figure 7 Lower shock wave speed by shrinking upstream demand

References

Bassan, Shy, and Faghri, Ardeshir (2005). "Experimental Investigation of Spatial Breakdown Evolution on Congested Freeways", *The 84th TRB Annual Meeting* (CD-ROM)*,* Washington D.C., USA.

Kerner, B.S., and Rehborn, H. (1997). "Experimental Properties of Phase Transitions in Traffic Flow", *Phys. Re v.Lett.,* 79(20). 4030-4033.

Kim, Taewan, and Zhang, H. Michael (2004). "Gap Time and Stochastic Wave Propagation", *Proc., 2004 IEEE Intelligent Transportation Systems Conference,* Washington, D.C., USA, October, 88-93.

Transportation Research Board, (1999). "Traffic Flow Theory" Chapter 2, 1-9.

Wu, Zheng (2002). “Statistics Analysis on Vehicle Flow Starting Wave Speed of Zhoujiazui Road”, *Journal of Traffic and Transportation Engineering*, 2(1), 67-73.
Xu, Tiandong, Hao, Yuan, and Sun, Lijun (2007). “Travel Time Prediction of Urban Expressway in Unstable Traffic Flow”, *Proc., International Conference on Transportation Engineering 2007,* ASCE, Chengdu, China, Volume 3, 2205-2210.
Zhang Yaping (2005). “Research of Traffic Characteristics and Capacity of Basic Sections on Urban Expressway”, Harbin Institute of Technology, Harbin, China, 53-57.

Urban Expressway Real-Time Traffic State Estimation and Travel Time Prediction within EKF Framework

Tiandong Xu[1], Osama Tomeh[2],Lijun Sun[1]

Abstract: In order to meet the need of traffic route guidance and traffic monitoring system in urban areas, this paper introduces a method to estimate and predict traffic condition and travel time between arbitrary locations by utilizing the Extended Kalman Filtering (EKF) framework. The basic concept is to first predict future traffic conditions using macroscopic dynamic models that are integrated with EKF based on the fixed traffic detectors in urban areas, and then predict actual travel time using "fictitious car" method. Finally, a case study in Shanghai is presented. The results demonstrate acceptable applicability and precision of the method. This study may lead to better traffic control and traffic route guidance, and improve the accuracy of broadcasted traffic conditions.

Keywords: Extended Kalman Filtering (EKF); macroscopic dynamic traffic flow models; traffic state; "fictitious car" method; travel time

Introduction

Dynamic traffic state estimation and prediction, especially dynamic urban travel time prediction are important components of Intelligent Traffic Systems. Reliable estimated traffic condition information is an indispensable part of effective traffic control and management, and can be used in dynamic route guidance, incident detection, ramp control, information propagation on VMS, and so on. Urban areas in many cities have been equipped with some traffic detection devices, such as inductive loop detectors. Estimating traffic conditions on a network level in real time using limited traffic detectors has been a pressing problem requiring urgent attention. Traffic flow models have been introduced as a powerful tool to tackle this problem. However, each type of these traffic flow models has its respective deficiencies (Daganzo,1995, Lebacque, 1999). In this paper, we propose the combination Kalman filtering technique with macroscopic dynamic traffic flow model as a technique used to treat the deficiencies of these traffic flow models. To this end, a state-space model and a traffic state estimator are established to estimate real-time traffic conditions on urban Expressways. The "fictitious car" method is then used to predict travel time.

Methodology

1. Macroscopic dynamic traffic flow model

Research has shown that the Papageorgiou model is reliable in describing the dynamic variation of traffic flow (Papageorgiou,1990). For simplification purposes, the model is expressed in discrete time and space form and illustrated by the following equations:

Density equation:

$$\rho_i(k+1)=\rho_i(k)+\frac{T}{\Delta i}[q_{i-1}(k)-q_i(k)+r_i(k)-s_i(k)] \quad \cdots\cdots\cdots\cdots (1)$$

Speed equation:

$$\begin{aligned} v_i(k+1)&=v_i(k)+\frac{T}{\tau}[(V(\rho_i(k))-v_i(k)]+\frac{T\xi}{\Delta i}v_i(k)[v_{i-1}(k)-v_i(k)] \\ &-\frac{\gamma T}{\tau\Delta i}\frac{\rho_{i+1}(k)/\lambda_{i+1}-\rho_i(k)/\lambda_i}{\rho_i(k)/\lambda_i+\kappa}-\frac{\delta_1 T}{\Delta i}\frac{r_i(k)v_i(k)}{\rho_i(k)+\kappa\lambda_i}-\frac{\delta_2 T}{\Delta i}\frac{s_i(k)v_i(k)}{\rho_i(k)+\kappa\lambda_i} \quad \cdots\cdots (2) \\ &-\varphi\frac{T}{\Delta i}\frac{(\lambda_i-\lambda_{i-1})}{\lambda_i}\frac{v_i^2(k)\rho_i(k)}{\rho_{cr}} \end{aligned}$$

Wherein, speed-density relation $V(\rho_i(k))=v_f\exp[-\frac{1}{a}(\frac{\rho_i(k)}{\lambda_i\rho_{cr}})^a]$.

Flow equation:

$$q_i(k)=\alpha\cdot\rho_i(k)v_i(k)+(1-\alpha)\rho_{i+1}(k)v_{i+1}(k) \quad \cdots\cdots\cdots\cdots (3)$$

Spot speed and space-mean speed relationship:

$$v'_i(k)=\alpha v_i(k)+(1-\alpha)v_{i+1}(k) \quad \cdots\cdots\cdots\cdots (4)$$

Where, i is segment number; k is cycle number; $\rho_i(k)$ is space-mean density； $v_i(k)$ is space-mean speed; $q_i(k)$ is traffic flow passing through segment i ; $r_i(k)$ is on-ramp flow; $s_i(k)$ is off-ramp outflow; Δ_i is the length of segment i ; T is cycle-time; λ_i is lane number of segment i ; v_f is free-flow speed; ρ_{cr} is critical density; a is shape factor; α is weight parameter， $0\le\alpha\le 1$; τ is time lag of speed; γ denote expectation factor; δ_1 is impact factor of on-ramps; δ_2 is impact factor of off-ramps; κ is modifying coefficient; φ is impact factor of lane changing; $v'_i(k)$ is spot speed.

2. State-space model

Traffic density and space-mean speed in macroscopic dynamic traffic flow models are represented by state variables as $x(k)=(\rho,v)_{(k)}$, as referenced in equations (1) and (2). Traffic volume and spot speed, detected by inductive loops, are regarded as observation variables, $y(k)=(q,v')_{(k)}$ as shown in equations (3) and (4). Hence, the state-space model of the dynamic traffic system can be expressed as:

$$\begin{aligned} x(k+1)&=(\rho,v)_{(k+1)}=f[x(k),w(k)] \\ y(k+1)&=(q,v')_{(k+1)}=h[x(k+1),v(k+1)] \end{aligned} \quad \cdots\cdots\cdots\cdots (5)$$

Wherein, traffic state vector x is the traffic state of the considered expressway, $x=[b^T \ c^T \ d^T]^T$. Wherein, $b=[\rho_1 \ v_1 \ \rho_2 \ v_2 \ \cdots \ \rho_{\mathrm{N}} \ v_{\mathrm{N}}]^{\mathrm{T}}$ is traffic state variables of every expressway segment; $c=[q_0 \ v_0 \ \rho_{\mathrm{N+1}} \ r_1 \cdots r_{\mathrm{N}} \ s_1 \ \cdots \ s_N]^{\mathrm{T}}$ is the boundary variable of the expressway system; $d=[\mathrm{v_f} \ \rho_{\mathrm{cr}} \ a]^{\mathrm{T}}$ is the most important model parameter. Vector y is a vector for observations including traffic volume and spot speed. Vector w is the traffic state noise vector, while v is the noise vector for observations. $f(\bullet)$ is a non-linear vector function with n dimensions, while $h(\bullet)$ is a non-linear vector function with m dimensions.

3. Design of traffic state estimator

In order to treat the high-order non-linear nature of traffic state-space model, the extended Kalman filtering theory can be used to design traffic state estimator (Kalman,1960). The traffic state estimator on an urban Expressway can then be designed as:

$$\hat{x}(k+1/k)=f[\hat{x}(k/k-1),0]+K(k)[y(k)-h(\hat{x}(k/k-1),0)] \ldots\ldots\ldots \quad (6)$$

Wherein, $\hat{x}(k+1/k)$ is the estimation of traffic state $x(k+1)$, based on observed information $y(k)$. $K(k)$ is the gain matrix, which can be obtained by calculating the quadratic Taylor equation of $f(\bullet)$ and $h(\bullet)$ at $\hat{x}(k/k-1)$. The principle of traffic state estimator is shown in figure 1.

4. Travel time prediction theory and algorithm

1) Travel time concept and prediction principle

Travel time on Expressway network refers to two kinds—*instantaneous* travel time and actual travel time. The space-time concepts of *instantaneous* and actual travel time are shown in the left part of figure 2. In reality, vehicles travel at the speed that is determined by the traffic condition at changing position and varying time, and actual travel time is the accumulated travel time in each segment. Therefore, the "fictitious car" method is designed to predict actual travel time（Xu, 2007）. The basic principle of this method is to predict traffic state of every segment in short-term based on the traffic state prediction model. Assuming that a "fictitious car" sets out from the start point with an average speed of all the vehicles that set out at the same point simultaneously, this speed is updated at the boundary between two segments with fixed time cycle (e.g. 10s), and then its travel time in each segment with changing speed is accumulated.

This is illustrated in the right side of figure 2. This method encompasses the influence of traffic flow perturbation, and thus it is suitable for calculating travel time in unstable traffic flow.

2) Traffic state prediction

In order to predict traffic state, we combine equations (1), (2), (3), (4), but neglect the impact of un-predictable noise, then formulate these state-space models as shown in the following equation:

$$b(k+1) = u[b(k), c(k), d(k), 0] \qquad (7)$$

The traffic state-space prediction model is:

$$\hat{b}(k+n/k+1) = u[\hat{b}(k+n-1/k), \hat{c}(k+n-1/k), \hat{d}(k+n-1/k), 0] \qquad (8)$$

Wherein, n is the length of predicting period. Initial value of traffic state prediction, $\hat{b}(k)$, $\hat{c}(k)$, $\hat{d}(k)$ can be calculated by equation (6). Research conducted by (Wang, 2006) shows that estimated model parameters are relatively stable with time. Hence, we assume that $\hat{d}(k)$ is constant in short prediction cycles. And $\hat{c}(k)(k+n-1/k)$ can be induced by historical data or the prediction of boundary value by trend extrapolation.

Experiment

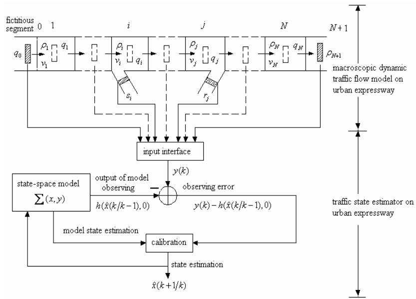
In this paper, we select an expressway section of Yan'an expressway in shanghai. This section includes on/off ramps, three lanes with loop detectors located every, 3.32km. We test the tracking ability of traffic state estimator and its adaptability by using real-detected traffic data (6:30-10:00, July 1st, 2005). The length of every segment is shown in figure 3.

Parameters of macroscopic dynamic traffic flow model of the expressway are then estimated (Xu, 2007). With the help of traffic state estimator(Equation 6), we can estimate the density, speed of the whole research section, then contrast them with the real density and speed, as shown in figure 4. From these figures, we can see that traffic state estimator has better tracking abilities on the whole.

Based on the traffic state prediction model (Equation 8) and the "fictitious" car method for predicting actual travel time, we select a 2min sliding step size and predict the travel time for vehicles completing the full journey. The results can be seen in figure 5. As can be seen, the data have good consistency—the lines are almost superimposed in free flow. In contrast, the difference is relatively large in congested flow, with relative error below 10% (except in several cycles).

Conclusion

We combine extended Kalman filtering technology with macroscopic dynamic traffic flow model to establish traffic state estimator. Then real-detected traffic data with limited traffic detectors is used to modify estimated value in real time to fit in the dynamic variance of traffic state, and as a result the real-time traffic state of the whole is estimated. Experiment shows the precision and reliability of traffic state estimator. Finally we manage to update the traffic state estimation model into traffic state prediction model, and design travel time prediction method using the “fictitious” car. The results of the experiment indicate that it is suitable for predicting travel time under unstable flow.



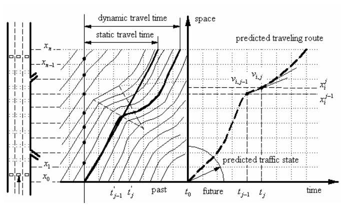


Figure.1 Dynamic traffic model and traffic state estimator of urban freeway

Figure.2 Sketch map of travel time prediction using “fictitious car” method

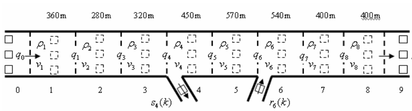
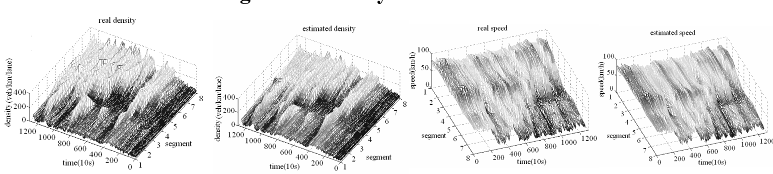
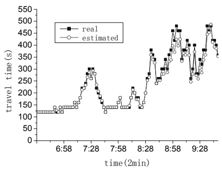


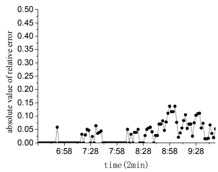
Figure.3 Freeway stretch for test



a) Contrast of density **b) Contrast of speed**

Figure.4 Contrast of the results between real data and estimated data





a) Contrast of travel time b) Relative error

Figure.5 Contrast of the results between predicted data and real data

Authors Contact Information

Tiandong Xu[1], Osama Tomeh[2], Lijun Sun[3] ,Yuan Hao[4]

[1]School of transportation Engineering, Tongji University, Room 306, 4800 Cao'an Highway, Shanghai 201804, China; Phone: (86) 21 6958-8634. E-mail: xtdtj163@163.com
[2]Aecom Consult Inc.3101, Wilson Blvd. Suite 400 Arlington, VA 22201; Phone: +703.682.5041. E-mail: Osama.Tomeh@dmjmharris.com
[3]School of transportation Engineering, Tongji University, Room 306, 4800 Cao'an Highway, Shanghai 201804, China; Phone: (86) 21 6958-8634. E-mail: ljsun@mail.tongji.edu.cn
[4]School of Transportation Engineering, Tongji University, Room 306, 4800 Cao'an Road, Shanghai, 201804, P. R. China; PH (86) 21 6958-8634. E-mail: hao_silvia@126.com

Reference

Daganzo, C.F. (1995). "Requiem for Second-order Fluid Approximations of Traffic Flow". *Transportation Research*, Vol.29B, No.4, 277-286.

Lebacque, J.P., Lesort, J.B. (1999). "Macroscopic Traffic Flow Models: A Question of Order". *Proceedings of the 14Th International Symposium on Transportation and Traffic Theory*, Jerusalem, Israel, 3-25.

Papageorgiou, M. (1990). "Modeling and Real-Time Control of Traffic Flow on the Southern part of Boulevard Peripherique in Paris: PART1 Modeling". *Transportation Research Part A* (24), 345-359.

Wang, Y.B., Papageorgiou, M. (2006). "RENAISSANCE—A unified macroscopic model-based approach to real-time freeway network traffic surveillance". *Transportation Research Part C* (14), 190-212.

Kalman, R.E. (1960). "A new Approach to Linear Filtering and Prediction Problems". *Trans.of the ASME J.of Basic Engineering*, 82(1), 34-35.

Xu, T.D.,et.al (2007). "Travel Time Prediction of Urban Expressway in Unstable Traffic Flow". Proc., *International Conference on Transportation Engineering 2007*. ASCE, 2205-2210.

GIS-BASED TRAFFIC SAFETY MANAGEMENT FOR HIGHWAY INTERSECTIONS

Xiang Qiaojun [1], Pan Fuquan [2], Lu Jian [3], Zhang Guoqiang [4], and William C. Eidson, P.E., PTOE[5]

[1]School of Transportation, Southeast University, 2 Si Pai-lou, Nanjing, Jiangsu, 210096, China, PH(86)25-83790776; email: xqj@seu.edu.cn
[2]School of Transportation, Southeast University, 2 Si Pai-lou, Nanjing, Jiangsu, 210096, China, PH(86)25-83790776; email: panfuquan@sina.com
[3]School of Transportation, Southeast University, 2 Si Pai-lou, Nanjing, Jiangsu, 210096, China, email: jianjohnlu@sina.com
[4]School of Transportation, Southeast University, 2 Si Pai-lou, Nanjing, Jiangsu, 210096, China, email: guoqiang.zhang@163.com
[5]Christopher B. Burke Engineering, Ltd., 9575 W. Higgins Rd., Ste. 600, Rosemont, IL 60018, USA, email: weidson@cbbel.com

Abstract: Intersection traffic safety management includes basic intersection information, safety evaluation, safety problem diagnosis, and countermeasure recommendations. In this paper, a new approach to improve the efficiency of safety evaluation is proposed: GIS-based model of safety level-of-service. Field ratings were used to initially screen intersection information. After entering intersection information into a computer model, the system automatically evaluates the safety of intersection, identifies existing safety problems and recommends prioritized countermeasures based on a benefit-cost ratio model. Initial results show the system improves the efficiency of evaluating and selecting intersection safety improvements.

Key words: *highway intersections; GIS; safety level-of-service; benefit-cost ratio*

1 Introduction

Highway intersections are accident-prone locations. Statistics show approximately 30%-50% of road traffic accidents occurred at or near intersections (Liu 1997; TRB 2003). With the recent campaign in China to “protect lives,” the public has become more interested in traffic safety, resulting in the need for more efficient intersection traffic safety management. In China, there is an opportunity to expand the current traffic safety data management practices, which rarely use GIS software (Fang 2006). The benefits of using GIS include dynamic data display, presentation of spatial relationships, and quick queries of relevant data.

The GIS-based traffic safety management method described in this paper integrates the fundamental elements of intersection safety management: basic intersection information, safety evaluation, safety problem diagnosis, and countermeasure recommendations. After entering basic intersection data into the software, the system automatically evaluates intersection safety using a non-accident model based on a theoretical safety level-of-service. The system finds existing safety problems according to a safety problem model and recommends prioritized countermeasures using a benefit-cost model, affording highway agencies increased efficiency managing networks and evaluating potential intersection improvements.

2 Intersection safety evaluation

Because many developing countries, such as China, lack complete intersection crash histories for the desired two to three year analysis periods, accident-based safety evaluation is often imprecise and unreliable. As an alternative, traffic conflict analysis is often used to evaluate intersection safety, though problems may exist that remain unidentified. Actual conflicts can be determined by manual or video observation, although both methods rely on human observers to subjectively identify each conflict, which is costly and can result in reduced data reliability.

Given the limitations of these methods, this paper describes a new safety level-of-service evaluation method. In theory, traffic conflict points are the leading cause of unsafe intersections conditions: their presence directly results in crashes. The safety level-of-service approach utilizes intersection traffic conflict points to establish an ideal basic model, which is then modified by other minor influencing factors, such as traffic signs and pavement markings, to establish the overall evaluation model. Since different conflict points may result in differing severity of crashes (including different levels of damage or injury), the method considers both the number of traffic conflict points and the relative severity of each conflict point. Since bicycle and walking modes account for a large percentage of trips in countries such as China, traffic conflicts in the model include not only multiple vehicle conflicts, but also conflicts between vehicles and bicycles, and between vehicles and pedestrians. The procedure for evaluating safety level-of-service is shown in Figure 1.

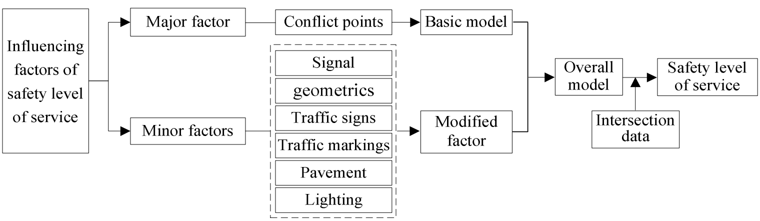


Figure 1 Procedure for safety level-of-service evaluation

The evaluation models are shown in equations (1) to (3).

$$EI = PRI \times MF \tag{1}$$

In equation (1), *EI* is the intersection risk index, or the safety level-of-service evaluation index; *PRI* is the intersection potential risk index, determined by equation (2). *MF* is the modified factor, determined by equation (3).

$$PRI = \sum_i \sum_j W_i \times (NCP_{ij} \times SCP_{ij}) \qquad (2)$$

In equation (2), i is the conflict point category (multiple vehicles, vehicle-bicycle, vehicle-pedestrian); W_i is the weight of the i category; j is the specific type of conflict point in the i category; NCP_{ij} is the number of conflict points of the j^{th} type in the i^{th} category. SCP_{ij} is the severity of the j^{th} type of conflict point in the i^{th} category.

$$MF = \sum_s \beta_s \times \left(1 + \frac{\sum_t w_{st} \times R_{st}}{100}\right) \qquad (3)$$

In equation (3), s is the index for each minor influencing factor; β_s is the weight for the s^{th} minor influencing factor; w_{st} is the weight for the t^{th} sub-influencing factor of the s^{th} minor influencing factor; R_{st} is condition rating of the t^{th} sub-influencing factor of the s^{th} minor influencing factor.

Based on the *EI* values, safety level-of-service is defined using six categories from “A” to “F.” “A” represents the best safety and “F” the worst safety. The methodologies and criteria used for safety level-of-service are detailed by Lu (2007).

3 Safety problem diagnosis and countermeasure recommendation

The purpose of safety problem diagnosis and countermeasure recommendation is to identify existing safety problems, uncover potential hidden conditions, and develop corresponding countermeasures, while providing a specific method to select and maintain safety improvements. Safety problem diagnosis involves initially identifying all possible safety problems for a given intersection based on a causal analysis of accidents, expert consultation, and field survey. The diagnosis requires establishing rating criteria, such as the significance and severity of a safety problem, and designing diagnostic survey forms. The forms are used by field engineers who rate each safety problem using the criteria. Finally, the various safety problems are ranked by the safety problem diagnosis model (Lu 2007). The flow process of safety problem diagnosis for intersections is shown as Figure 2.

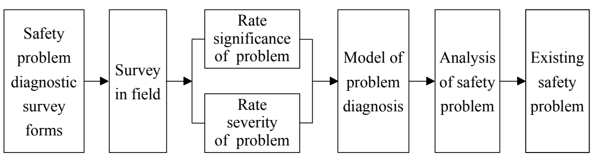


Figure 2 Flow process of safety problem diagnosis for intersections

Often, a given safety problem can be resolved using several different countermeasures, each with varied costs and benefits. The benefit-cost ratio method was chosen to select the most appropriate solution. Since it is difficult to establish exact safety benefit values, the crash reduction factor (Zeeger 1981) can be used as the benefit in the calculation. The benefit-cost ratio is shown as equation (4).

$$P = \frac{Benefit}{Cost} = \frac{CRF}{Cost} \tag{4}$$

Where, P is benefit-cost ratio and the basis for the choice of countermeasure; CRF is average crash reduction factor; $Cost$ is average cost for a particular countermeasure.

4 System composition

This GIS safety management system was built in the Visual Basic programming language to integrate the function groupware of GIS-MapObjects, Active X controls and database methods, using control ADO, SQL, True DBGrid Pro7.0, and ListView. The system databases include the GIS database (graphical data), intersection information database (intersection data and criteria), and the system inner database (user data; various model and report data). The main modules of the system include map operation, information query, safety level-of-service display, report (countermeasure) generation and model parameter setting. The logic framework is shown in Figure 3.

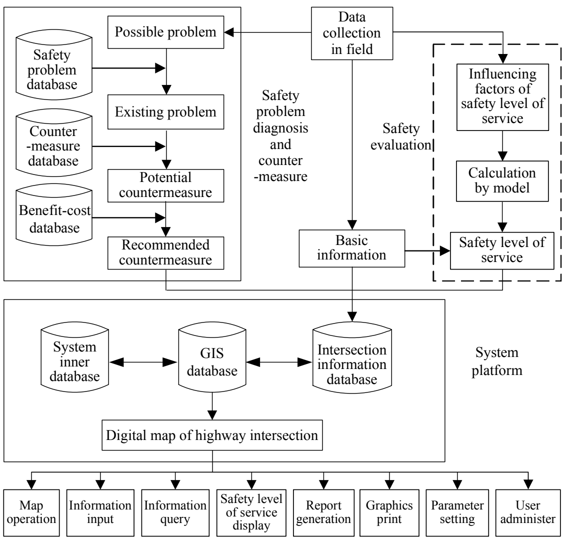


Figure 3 Logic framework of the safety management system for intersections

5 Function realization

The safety management system includes eight primary functions:

(1) Map operation: Digital map of highway network and intersections, including standard map controls, such as zoom, pan, and distance measure.

(2) Information input: Intersection data input allows for subsequent analyses.

(3) Information query: Reports safety level-of-service, existing safety problems, countermeasures, and spatial locations of intersections. User selects by text or map.

(4) Safety level-of-service display: Using intersection data in the database, the system automatically evaluates the safety of intersections and presents the safety level-of-service. Color symbols are used to distinguish different safety levels.

(5) Report generation: Based on intersection data in the database, the system can report and print existing safety problems and recommended countermeasures.

(6) Graphics print: Print intersection map, reports, and data windows.

(7) Parameter setting: Advanced users can modify the system inner database parameters to adapt the system for local conditions.

(8) User administer: System security and control limits access to maintenance functions to Advanced users with proper credentials.

6 Application

The example digital system map, shown in Figure 4, includes arterial highway intersections in the Jiangsu province of China, with color symbols denoting different safety levels-of-service. At present, system use has been limited to the authors' research on techniques to manage intersection traffic safety.

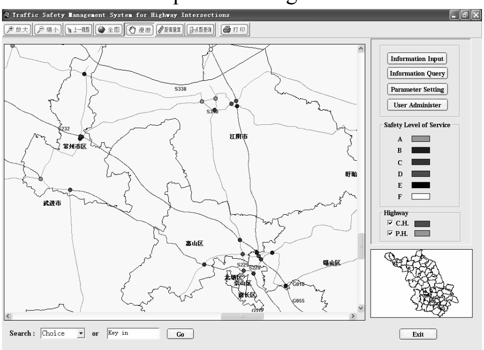


Figure 4 Main interface of the system

After the data is entered, the system automatically analyzes existing safety problems, levels-of-service, and countermeasures. Figure 5 shows a sample report output summarizing existing safety problems and corresponding countermeasures for the intersection of country highway G312 and province highway S122 in China.

Figure 5 Existing safety problems and countermeasures of the example

7 Conclusions and Recommendations

This paper describes the concept of safety level-of-service to evaluate traffic safety management for intersections using GIS technology and the Visual Basic programming language. The initial results show the system can quickly evaluate intersection safety, find existing safety problems, and recommend countermeasures, thereby increasing the efficiency of highway department personnel. It should be noted that safety level-of-service is a new safety evaluation method, so further research is needed, especially to correlate the evaluation results with the traffic accident method. In addition, the graphical user interface in the computer modules may benefit from further refinement and optimization.

8 Acknowledgments

This research was supported by National Natural Science Foundation of China (No: 50522207) and China Department of Transportation (No: 200431822333).

References

Fang Shou-en, Yang Zhen (2006). "Design of GIS-Based Road Safety Information Management System." *Journal of Tongji University*, 34(5): 629-633.

Liu Xiao-ming, Duan Hai-lin (1997). "Methods of traffic safety evaluation at grade intersection." *Chinese Journal of Ergonomics*, 3(1): 49-53.

Lu Jian, Yuan Li, Zhang Guo-qiang (2007). *Diagnostic Approach for Safety Performance Evaluation and Improvements of Highway Intersections*, TRB 2007 Annual Meeting, Washington, D. C.

Lu Jian, Xiang Qiao-jun, Pan Fu-quan (2007). *Research on Traffic safety Technique for Highway Intersections in China*, Southeast University, Nanjing, China.

Transportation Research Board (TRB) (2003). *Guide for Addressing Unsignalized Intersection Collision*, TRB, Washington, D. C.

Zegeer, C. (1981) *Highway Safety Improvement Program (HSIP).* Report FHWA-TS-81-218, Federal Highway Administration, Washington, D.C.

Urban Planning for Sustainability: Ankara's Planning Experience in Creating Sustainable Urban Form and Transport

E. BABALIK-SUTCLIFFE[1]

Abstract

Urban transport problems resulting from high levels of mobility and car usage are major challenges facing most urban areas worldwide. While past decades saw various transport policies and measures to restrict and manage car usage in the cities, it has been increasingly recognized that urban form, development patterns, and density levels affect travel behavior and car usage, and that therefore urban planning can be a major tool in solving transport and traffic problems. Sustainable development debates support this line of thought, emphasizing the most appropriate urban form for creating sustainable urban transport systems. Certain planning approaches, such as intensification of urban development and corridor development that support public transport usage, together with higher density and diversity in development, are promoted as sustainable alternatives to the prevailing lower-density urban sprawl. This paper assesses the effectiveness of such planning approaches, focusing on the planning experiences of Ankara, the capital of Turkey.

1. Introduction

Overcoming car-oriented urban development patterns has become a major challenge for most cities in the world. Urban sprawl, low-density residential development, and dispersal of non-residential activities, such as commercial centers and offices, are recognized as major urban problems of the day, resulting in ever-increasing levels of car-usage. While transport policy measures that promote restrictions on car usage and improvements in transit and non-motorized modes have become widespread, there is increasing awareness that these problems can be solved through land-use policies that can influence travel behavior and attitudes towards car usage. Neighborhood design approaches to influence travel behavior received significant emphasis, with much research on designing transit-oriented or pedestrian-friendly development and less car-dependent neighborhoods (Calthorpe 1993; Katz 1994).

Arguments for 'sustainable urban development' also reinforce these debates, highlighting certain urban patterns as being able to create more desirable traffic outcomes in terms of sustainability, i.e. reduced car usage and increased transit and walking trips. Although the compact city model has been the focus of much discussion, functioning and management become difficult as cities grow in population and size. Compactness and intensification strategies are seen as being

[1] Middle East Technical University, Faculty of Architecture, Department of City and Regional Planning, 06531 Ankara, Turkey; PH 90 312 210 2263; FAX: 90 312 210 7965; e-mail: ebaba@metu.edu.tr

inappropriate, particularly for cities in developing countries with very high population levels and already compact forms due to poor transit services and the need to stay close to central amenities. This compact form is often problematic in such cities where there is a lack of infrastructure and urban management structures to make the model work (Williams 2004). Alternatives to the compact model have been proposed, such as corridor development that can enable a high-capacity transit service, and multi-centered urban areas that function as a network of self-sufficient settlements (Williams et al. 2000).

In addition to these 'macro' models of urban development, it is argued that high-density development and a mix of different land-uses can help reduce the need for travel, particularly by car: higher densities and mixed-use development can result in reduced distances between activities, and hence walking, cycling and public transport can become effective alternatives. Designing areas in a pedestrian-friendly and transit-oriented way is also vital in changing travel demand in favor of transit and walking, hence bringing together density, diversity and design in creating a built environment that is not car-dependent (Cervero and Kockelman 1997). These debates helped bring urban planning to the forefront of the transport policy agenda; and 'planning for less travel' (Banister 1999) became an increasingly important strategy.

The planning experience in Ankara and the resulting urban transformation are worth analyzing in this context. Since the 1970s, planning studies in Ankara have aimed at transforming the compact form and fringe development pattern into a controlled decentralization along two corridors, incorporating strategies for high density development and a mix of land-uses. The resulting urban form, home-work patterns and traffic levels along these two corridors provide opportunities for an analysis of the effects of corridor development, which can contribute to the above arguments.

2. Transformation of Ankara: from compact to transit corridor development

Ankara, the capital city of Turkey, is a metropolitan area with a population of 3.5 million people. The city experienced historically high-density development and has a rather compact form. While the compactness had provided advantages for the functioning of the city, after the 1970s it became a problem in the face of rapid population growth. Urban growth at the fringes resulted in continuous spatial growth in the 1970s, and although densities were still relatively high, the urban area was becoming difficult to manage and to serve effectively with transit.

Planning studies in the 1970s proposed two linear corridors of development for the future urban form of Ankara along the western and south-western corridors. The strategy comprised the decentralization of residential areas, commerce, industry, and other workplaces, creating a mix of different land-uses. From the perspective of 'sustainable urban development models' debate, it can be suggested that the approach was a 'corridor development' with a strong focus on providing a mix of land-uses in order to help relieve the pressure on the central city (Fig.1). Both corridors were planned to accommodate residential areas together with working areas. The western corridor was proposed to develop with industrial estates and a number of commercial sub-centers. The south-western corridor was proposed to accommodate state offices, university complexes and relatively fewer sub-centers. Urban rail systems were proposed for both corridors to connect the new residential areas, offices and commercial centers with each other and with the city centre.

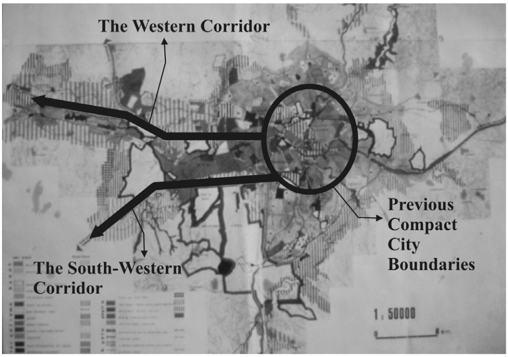


Figure 1: Ankara Urban Development Plan – The Two Development Corridors

3. The resulting urban and transport pattern

Today the western and the south-western corridors of Ankara have become the two main axes of the city, having been the subject of most of the new spatial development since 1985, as proposed in the plans, particularly along the western corridor. The two corridors have different implementation histories. The metro line along the western corridor opened in late 1997. Construction of the south-western line began in 2005, after being postponed for some time, but is yet to open.

The western corridor housed various government-led projects, from large-scale mass-housing projects financed by the government to light and heavy industrial estates. The mass housing projects and possibly the existence of the metro line resulted in higher density in residential development. The creation of commercial cores, such as new town centers, could not be realized. Instead, large-scale shopping centers have opened along this corridor. Notwithstanding the lack of new commercial cores, a certain level of mixed-use development was achieved with a fairly significant level of residential-workplace mix.

The southwestern corridor, on the other hand, developed predominantly in residential character. Rather than government-led housing projects, market-led residential development took place along the corridor, but remained significantly lower than along the western corridor. This is believed to be a consequence of the market-oriented development and the more car-dependent urban structure, which is further reinforced by the lack of a high-capacity transit system along the corridor. State office decentralization took place, although only to locations that were still close to the city center and thus not strongly integrated with the residential areas that developed further along the corridor, and similar to the western corridor no new town centers could be created.

Urban growth, employment, and traffic trends in the two corridors are presented in Table 1, which shows that both corridors experienced significant population growth, but that the western corridor is four times more populated than the south-western corridor, accommodating 43 % of the total urban population. The population size in the south-western corridor is less; however, growth rate is quite high.

Table 1 Population growth, employment patterns and traffic figures in two corridors

	WESTERN CORRIDOR (mixed-use, higher density, with metro)	SOUTH-WESTERN CORRIDOR (mostly residential, lower density, without metro)
Population		
Population in 2000	1,386,000	344,000
Population as a percent of total urban population	43 %	11 %
Population growth since 1985	115 %	210 %
Population growth since 1990	39 %	78 %
Employment		
Employment as a percent of total city employment - industry	50%	10%
Those working along the same corridor	51%	6%
Those working at the city centre	40%	73%
Traffic		
Share of cars in vehicle flow	78%	89%
Share of car usage in passenger trips	16%	35%
Share of metro usage	5%	No metro in the corridor

The difference in workplaces of the residents in the two corridors is quite striking (Table 1). Half of the population living along the western corridor work along this same corridor (Senyapili et al. 2002). This is an extremely significant rate, verifying that a mix land-use strategy can be a powerful factor in changing home-work patterns and reducing the need to travel to the city center. This is further reinforced by the finding regarding the other corridor, which has a limited level of mixed-use development, from where 73% commute to the city center.

The rate of private vehicles in total vehicle flow is high for both corridors; particularly for the south-western corridor. Car usage in passenger trips may appear low when compared to those of industrialized countries; however, it is important to evaluate these levels in the context of Turkey and Ankara. In Ankara average car usage is around 21%. The western corridor, which developed in higher densities, mixed-use, and with the metro system, has a much lower car usage rate when compared to the city average. The south-western corridor, in contrast, has a much higher car usage rate than that in the city, revealing the outcome of car-dependent, lower density (and higher income) residential development along this corridor.

The socio-economic characteristics of the two corridors may also account for the differences in travel patterns. Then again, the socio-economic differences also seem to be the result of the different development strategies along the two corridors: Government-led housing projects and industrial development in the western corridor resulted in a residential area profile that was middle-income and more transit dependent. Market-led residential growth and state office development along the south-western corridor, on the other hand, attracted higher income residents with high car-ownership levels, enabling and promoting a more car-oriented urban pattern.

A final point about the travel patterns is that metro usage along the western corridor is not as high as would be expected. Some 80% of trips are made by bus and minibus transit systems, and only 5% by the metro. In this paper, density and diversity have been discussed as the main approaches behind sustainable urban development, but it is also important to design these mixed-use residential and non-residential areas in a pedestrian-friendly and transit-oriented way in order to change travel behavior. Unfortunately, this approach did not exist when the new development areas were being planned in Ankara. Almost all of the metro stations were designed in such a way that they could not be easily or directly accessed from the new developments. The design of most new developments in the city has been shaped by accessibility by road; and it would appear that this has resulted in a preference for road-based public transport systems rather than the metro.

4. Conclusion

The planning experience, and particularly the implemented corridor development strategy, shows that promoting a mix of land-uses can indeed result in better access between the home and workplace, consequently changing home-work patterns. Both the high rate of 'living and working in the same corridor' and the lower level of car usage in the western corridor support arguments for the traffic benefits of mixed-use and high-density development. While the metro may have attracted more development, resulting in higher densities, usage is still limited, suggesting that had the metro system been better integrated with new developments, benefits for traffic could have been vast.

Another important aspect of the Ankara experience is the difference in the two corridors in terms of the level of 'success' in creating the desired mixed-use pattern when implementing the corridor development strategy. Both corridors were planned to accommodate a higher density and diversity of development; however, the two resulting urban patterns differ vastly from each other. The success in attaining a more mixed-use development along the industry decentralization corridor may be due to the fact that manufacturing and industry generally decentralize first to take advantage of land cost and availability, and better traffic conditions. In addition, government involvement in housing and industrial estate development also seems to have helped the process. All these aspects played an important role in creating a middle-income resident profile that was more likely to use public transport. The south-western corridor, on the other hand, reveals a less successful process in terms of creating a mix of land uses with office-housing integration. Once again, this may be due to the fact that state offices, which were proposed along this corridor, tended to decentralize last due to the advantages of central locations, such as the economies of urbanization and frequency of face-to-face contacts. In addition, having this corridor developed predominantly by the private sector also played a role: the market, not surprisingly, aimed at high-income residents that wished to live in larger suburban houses, thus resulting in a rather sprawled urban form. The resulting profile of residents also made sure that the area did not attract much working places. The trend in this corridor is quite similar to those in industrialized country cities, where high-income residential development tends to exclude major service areas and offices that may jeopardize the secure image of the area.

It is not intended to suggest here that market-led development is to be prevented; but it is clear that such developments target more car-oriented patterns, and without strict planning control and a more pro-active role by the local government the desired urban patterns are hard to achieve. In the light of this finding, it is important to consider the possibility that creating a high level of mixed-use development and more transit-friendly patterns in areas favored by high-income residential growth may not be viable. Unfortunately, it is perhaps even more important to change urban and travel patterns in such areas, where car ownership and car-usage tend to be much higher, threatening the attainment of objectives for sustainable urban transport.

References

Banister, D. (1999) “Planning more to travel less: land use and transport.” *Town Planning Review*, 70(3), 313-338.

Calthorpe, P. (1993) *The next American metropolis: Ecology, communities and the American dream*, Princeton Press, New York.

Cervero, R. and Kockelman, K. (1997) “Travel Demand and the 3Ds: Density, Diversity and Design.” *Transportation Research D* 2, 199-219.

Katz, P. (1994) *The New Urbanism: Toward an architecture of community*, McGraw-Hill, New York.

Senyapili, T., Berkman, G., and Uzun, N. (2002) *Transformation in ownership pattern in the use of inner city housing stock vacated due to suburbanization: case study Ankara*, Middle East Technical University (METU) Research Fund Project, no. 2001-01-02-02-01.

Williams, K., E. Burton, and Jenks, M. (2000) *Achieving sustainable urban form*, E&FN Spon.

Williams, K. (2004) “Can Urban Intensification Contribute to Sustainable Cities? An international perspective.” *City Matters*, Official Electronic Journal of Urbanicity.

Multiple Regression Analysis on the Safety of Transportation and Social Economic Development

QU Zhiming, Ph.D.[1] Grant G. Schultz, Ph.D., P.E., [2], and Bashir AlMalik, Ph.D.[3]

[1]Associate Professor, Department of Civil Engineering, Hebei University of Engineering, 199, South Guangming Street, Handan City, Hebei Province, 056038, P.R. China; PH 013613207797, 0310-8576536; e-mail: chinaqzm@163.com
[2]Assistant Professor, Department of Civil and Environmental Engineering, Brigham Young University, 368 Clyde Building, Provo, UT 84602; PH (801) 422-6332; FAX (801) 422-0159; e-mail: gschultz@byu.edu
[3]Civil Engineering Dept., School of Engineering, King Saud University, Riyadh, Saudi Arabia balmalik@gmail.com

1 Abstract

Statistical analyses have shown that the death rate per 100,000 persons in traffic and transportation related deaths can be correlated to economic indices including per capita Gross National Product (GNP). The death rates along with 14 social economic indices in 28 sample countries were chosen to evaluate the relationship between transportation safety and economic development. Using multiple regression analysis, several models were developed to analyze the effect of economic factors on transportation safety. The results of the models developed showed that the death rate per 100,000 persons tended to increase when predicting the development of national transportation safety. As a result, advice is provided to macroscopically readjust and control the main social economic factors which include changing economic growth, enhancing economic strength, increasing technologic devotion and education training, and strengthening safety control to realize stable improvements in traffic and transportation safety.

2 Introduction

Traffic and transportation safety has been shown to be correlated with social economic development during varying economic periods. Huang indicates that in different stages of economic development, several different features and development trends can be observed (HUANG 1986). Additional research using multiple regression analysis identified multiple functions related to economic development and related trends. Through regression analysis and primary components, the relationship between traffic and transportation safety and the subsequent changes during different stages of the economic development of a society can be determined. At the same time, the inherent laws between transportation safety and development can also be revealed (HE 2005, WANG 2003). Utilizing mathematical models, the safety of the transportation system is studied as a function of the current social macroeconomic development. The main factors in this analysis are evaluated indicating that transportation development tends to be predicted in order to provide important theoretical support for departments developing a plan and reporting measures.

The transportation system is the basis of a strong national economy and way of life, providing a strong economic environment for industry development and regional growth. The specialized division of labor, the regional advantage and the reasonable economic structure are promoted through such a system, which greatly improves social benefits and provides for a sustainable economic environment. In recent years, fast paced economic development has greatly increased the growth of traffic and transportation, which has imposed restrictions on national economies. As a result, a greater focus has been placed on the development of the transportation system to meet the needs of economic growth.

3 Trends in Traffic and Transportation Safety

One of the primary indices to assess transportation safety conditions is that of transportation related deaths per 100,000 persons (WANG et al. 2001). Figure 1 graphically illustrates the number of deaths and injuries per person for the years 1978 through 2005 in China. This figure shows an increase in deaths and injuries over time, particularly after 1994. Figure 2 provides a graphical illustration of the number of deaths and injuries per 100,000 persons as a function of gross domestic product (GDP) in units of renminbi (RMB) from 1978 to 2005. When comparing the two figures it is apparent that the number of deaths and injuries increases after 1995. This corresponds to the time period when the GDP is greater than 5,000 RMB. From the data in the two figures it is theorized that as GDP increases, the number of vehicles increases, and the death and injury rate associated with these vehicles increases as well. The two states classified in the transportation safety analysis are the positive dependent relationship between the death rate per 100,000 persons and the average GDP, while the other is the negative effect of this relationship. The increase in death and injury rate (transportation safety) and the increase in social economic development appear to be closely related.

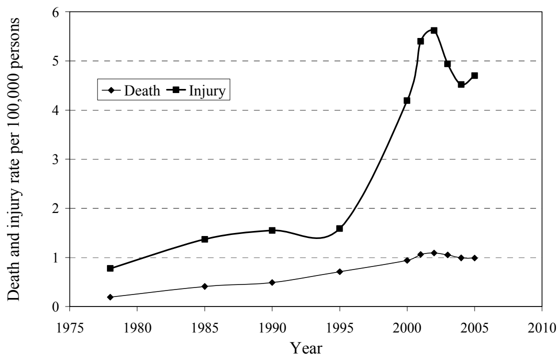


Figure 1 Death and injury rates from 1978 to 2005

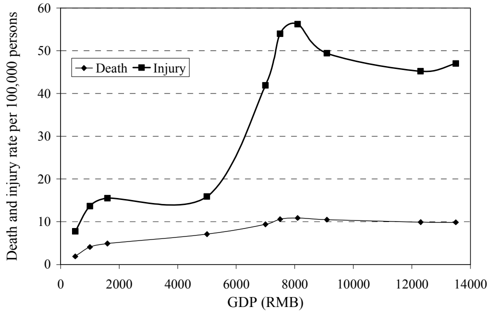


Figure 2 Death and injury rates as a function of GDP (RMB)

4 Multiple Regression Analysis of Transportation Safety and Social Economic Development

4.1 Indices of social economic development from sample countries

To evaluate social economic development as a function of transportation safety, 18 indices of 5 different types were chosen for analysis. Of the 18 indices chosen, only 14 were determined to be applicable to illustrate the exocentric rules between the different stages of transportation safety development and social economic development. The indices and their corresponding variables by type are outlined as follows:

(1) Indices of comprehensive development such as GDP (x_1) and per capita GNP (x_2).

(2) Indices of social structure including percentage of agricultural products in GDP (x_3), tertiary industry product (x_4), percentage of export of cargo and service in GDP (x_5), percentage of urban population in national population (x_6), percentage of public education in GDP (x_7), and percentage of scientific research devotion in GDP (x_{14}).

(3) Indices of population quality including enrollment rate of middle school students (x_8), population growth rate (x_9), and infant death rate (x_{10}).

(4) Quality of life indices such as doctors per 1,000 persons (x_{11}).

(5) Indices of social stability including unemployment rate (x_{12}) and percentage of nonagricultural employment (x_{13}).

To provide generality and contrast in the data an analysis of more than 200 countries was considered (WANG et al. 2001). Of the 200 countries, 28 countries with differing economic development patterns were chosen for the analysis including 10 high income countries, 7 moderate and high income countries, 8 low and moderate income countries, and 2 lower income countries. The data is derived from the literature including statistical data by authorized organizations (ZHU and WU 2001, International Labor Bureau 1998 and 2002).

4.2 Regression analysis between traffic transportation safety and indices of social economics

Stepwise linear regression was utilized in the analysis to determine the variables that were most important when evaluating the relationship between transportation safety and indices of social economics. Using stepwise linear regression an optimized equation was determined wherein the independent variables affecting the dependent variables are selected as those that are the most statistically significant in the regression analysis. Through regression analysis between social economic development and transportation safety in 1990 and 2000 for the 28 sample countries, the main factors influencing transportation safety and the degree to which they influence such safety were determined.

(1) Through regression analysis between death rate per 100,000 and 14 social economic indices ($x_1, x_2, \ldots, x_{14}$) in 28 countries in 1990, the relationship is as outlined in Equation 1 (YU and REN 1999).

$$Y = 7.596 - 0.356x_2 - 0.896x_3 + 1.254x_6 + 0.412x_7 + 0.565x_8 + 0.299x_{12} - 2.894x_{13} \quad (1)$$

Table 1 shows the regression coefficients for the 1990 model. The coefficient of determination or R^2 value for the model was 0.89851 indicating that the model closely fits the data analyzed. Additional statistics included $F = 5.1203$ and $d(D,W) = 3.2654$.

(2) Through regression analysis between death rate per 100,000 persons and 14 social economic indices ($x_1, x_2, \ldots, x_{14}$) in 28 countries in 2000, Equation 2 is developed.

$$\begin{aligned} Y = {} & 14.0251 - 0.7012x_2 - 0.3356x_3 + 0.9541x_4 + 0.5018x_6 - 0.265x_8 + 0.499x_{12} \\ & - 0.8542x_{13} + 5.211x_{14} \end{aligned} \quad (2)$$

Table 2 shows the regression coefficients for the 2000 model. The R^2 value for the model was 0.6598 indicating that the model closely fits the data analyzed. Additional statistics include $F = 4.5564$ and $d(D,W) = 2.1357$.

Table 1 Regression coefficient check of sample model in 1990

	Independent variable	Partial correlation coefficient	*t*-statistic	Level of significance	Standardized partial regression coefficient
$\gamma(y, x_2)$	GNP	-0.3659	1.2564	0.1125	-0.3425
$\gamma(y, x_3)$	Percentage of agricultural product in GDP	-0.6201	3.2144	0.0036	-0.9874
$\gamma(y, x_6)$	Percentage of urban population in national population	0.6235	3.2699	0.0041	1.2121
$\gamma(y, x_7)$	Percentage of public education in GDP	0.5102	2.0655	0.0510	0.3985
$\gamma(y, x_8)$	Enrollment rate of middle school students	0.2659	1.9512	0.0851	0.4649
$\gamma(y, x_{12})$	Rate of unemployment	0.3595	1.5014	0.1951	0.5031
$\gamma(y, x_{13})$	Percentage of nonagricultural employment	0.2545	4.6621	0.0002	-2.3654

(3) Contrasting the results of the 1990 and 2000 data it can be deduced from the regression equations outlined that there is a close relationship between transportation safety and social economic development. In contrast, the primary factors influencing transportation safety are the enhancements of national economic strength, which is

the basic reason for an increase in safety. The correlation coefficient between GNP and safe transportation in 1990 was -0.356, while this correlation increased in 2000 to -0.7012. Although these correlation coefficients are both negative, the level of influence increased from 5th in 1990 to 2nd in 2000.

Table 2 Regression coefficient check of sample model in 2000

	Independent variable	Partial correlation coefficient	t-statistic	Level of significance	Standardized partial regression coefficient
$\gamma(y, x_2)$	GNP	-0.6215	3.1125	0.0025	-0.6645
$\gamma(y, x_3)$	Percentage of agricultural product in GDP	-0.7821	1.6654	0.3156	-0.4567
$\gamma(y, x_4)$	Tertiary industry product	0.5789	2.0548	0.03349	0.3215
$\gamma(y, x_6)$	Percentage of urban population in national population	0.2399	4.0125	0.00215	0.7012
$\gamma(y, x_8)$	Enrollment rate of middle school students	-0.5698	2.5589	0.09012	-0.2011
$\gamma(y, x_{12})$	Rate of unemployment	-0.3251	1.8712	0.2515	-0.5144
$\gamma(y, x_{13})$	Percentage of nonagricultural employment	-0.5123	3.1101	0.05879	-0.2174
$\gamma(y, x_{14})$	Percentage of scientific research devotion in GDP	0.5621	2.6548	0.1215	4.9981

5 Regression analysis of national traffic safety and social economic development

By analyzing the indices of social economic development and transportation safety (National Statistic Bureau 2002), the factors influencing transportation safety and the degree of influence can be described. Using multiple regression analysis, the relationship between traffic death rate and economic development can be determined as outlined in Equation 3.

$$\begin{aligned} Y &= 5.9908 + 0.9945x_2 - 1.1214x_3 + 0.5612x_5 - 2.0155x_6 - 1.0012x_{14} - 0.2258x_7 \\ &+ 0.8412x_{13} \end{aligned} \tag{3}$$

Equation 3 can be checked using F and $D.W.$ checking. The partial regression coefficients are checked using t-statistics, while and the fit of the data can be checked with the coefficient of determination, $R^2 = 0.89034$. The path analysis shows that the direct influence percentage of the dependent variables such as x_6, x_3, and x_2 are 40.12%, 18.26%, and 16.23% respectively, while others are not more than 10%. The most influential indirect percentage on death rate is x_6 at 68.12%, followed by x_2 at 14.18%, and x_3 at 11.17%. The summary influential percentage on traffic death rate is x_3 at 25.24%, followed by x_{13} at 19.98%, and x_2 at 14.19%. The industrial structure has proven to have a large impact on transportation as the nation is currently in a stage of initial industrialization.

In light of the study of historical laws of traffic death rates and contrasting the regression analysis results while using regression prediction, grey theory, trend analysis prediction, and difference self-regression sliding average analysis, the last method is able to be utilized to simulate and predict the death rate over the past several years and into the future. As a result, the model of ARIMA (1,1,1) can be deduced, the results of which are provided in Table 3.

Table 3 Model of ARIMA (1,1,1)

Z(T+L)	Standard error	T
0.00543	0.651283	0.4468594
1.006899 Z(T+L-1)	0.8985845	1.1025681
-1.10008588 Z(T+L-1)	0.8690012	-0.8088561

According to the death rate from 1978 to 2003, the death rate for 2004 to 2013 can be computed and predicted as illustrated in Figure 3. It can be seen from this model that the simulation is able to provide a relatively accurate prediction and fit to the model. From the prediction, the traffic death rate tends to increase in the future.

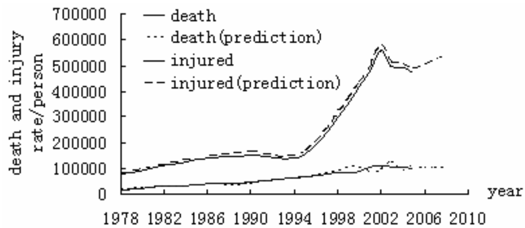


Figure 3 Historical analysis and future prediction of death rate and injury

6 Conclusions

By comparing the relationship between traffic death rates per 100,000 persons and per capita GDP of 28 sample countries in the years 1990 and 2000, it was concluded that safety tends to improve with social economic development.

Through regression analysis, the increase of national traffic safety was determined to be closely related to per capita GDP. Transportation safety is affected by a variety of factors such as industrial structure, employment, science technology education, and economic development.

References

HE Manxi. (2005). Analysis of relationship between transportation and economic development of Zhejiang. Mathematics in economics, 22(2): 162-167.

HUANG Yuanlin. (1986). Traffic transportation and economic development. The symposium of 5th national systems engineering conference.

International Labor Bureau. (1998 and 2002). Year of labor statistics.

National statistics bureau. (2002). The annals of national statistics 2002. Beijing: publishing house of Chinese statistics.

WANG Qingyun. (2003). Inherent laws between transportation and economic development. System transportation: 4-7.

WANG Xianzheng, HUANG Shengchu, JIANG Hanxin. (2001). Coal mine safety and control system in USA. Beijing: coal mine industry.

YU Xiulin, REN Xuesong. (1999). Multiple regression analysis. Beijing: publishing house of Chinese statistics.

ZHU Qingfang, WU Hanguang. (2001). Index system of society. Beijing: publishing house of social science in China.

Statistical Cause Analysis and Potential Countermeasures For Commercial Motor Vehicle Collisions in China

Yiyi Wang
Beijing University of Technology, Beijing, China
No.100 Pingleyuan Street, Chaoyang District, Beijing, China,
E-mail: yiyiwang@emails.bjut.edu.cn

Xiaoduan Sun, Ph.D., Professor and P.E.
University of Louisiana, Specially invited professor at Beijing University of Technology, E-mail: xxs0699@louisiana.edu

Daniel J. Melcher, P.E.,
Vice President, Armstrong Forensic Engineers, Inc.
E-mail: danielmelcher@gmail.com

ABSTRACT

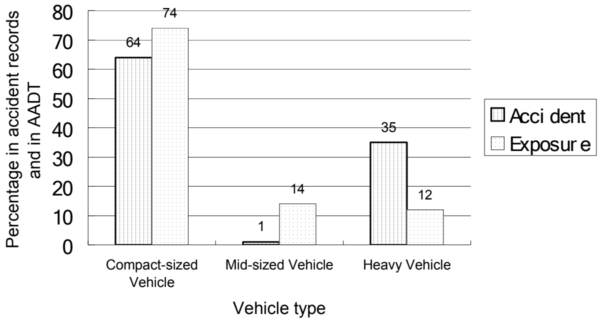
The rate of incidents involving commercial vehicles is a particular concern in China, accounting for an estimated 35.8% of all crashes in 2004 and over 30% in 2005[1]. Commercial motor vehicle safety issues in China were studied by collecting data from the Jingjintang freeway, which serves as the primary high-speed link between the cities of Beijing and Tianjin, China. For this study, heavy vehicles refer to vehicles with a gross weight of at least 20,000 pounds or vehicles designed to transport at least 40 passengers. This vehicle type is a surrogate for most Commercial Vehicles. Mid-sized vehicles are those with a gross weight between 10,000 to 20,000 pounds, or vehicles designed to transport 20 to 39 passengers. Light/compact-sized vehicles are those smaller than the mid-size.

Commercial vehicles are over-represented in collision involvement, and crashes that involve commercial vehicles have a higher severity level. Data collected from a survey of commercial vehicle drivers was analyzed for statistical relationships between collision probability and causes related to driver behavior, driver training, and vehicle mechanical condition. Higher crash risk corresponds to lower driver experience, increased driver tendency to load a vehicle in excess of design capacity, operation of older or defective vehicles, and driver tendency to travel at excessive speeds. Based on the known factors influencing commercial vehicle crash probabilities, a number of potential safety countermeasures are suggested for further study in China.

COLLISION INVOLVEMENT RATES

Figure 1 illustrates 2004 data on the percentage of total accident involvement relative to the percentage of total traffic volume exposure for each vehicle type on the Jingjintang freeway. While accounting for only 12 percent of average annual daily traffic (AADT), heavy vehicles were involved in 35 percent of police-documented accidents. Only 64 percent of crashes involved light/compact sized vehicles despite comprising 74 percent of the AADT. Only 1 percent of collisions involved mid-sized vehicles despite accounting for 14% of the AADT. Since heavy vehicles are over-represented in collisions, a targeted study of the causes and safety countermeasures for this crash type may provide the greatest benefit.

Figure 1. Accident involvement and exposure by vehicle types.



VEHICLE TYPE AND COLLISION SEVERITY

Given the weight, occupant number, or hazardous payload inherent in commercial vehicles[2], any collision involving this type of vehicle can lead to significant losses, injuries, or fatalities. Crash data from the Jingjintang freeway was collected from 2003 to 2005 and sorted into four severity categories[3]: (1) Property-damage only (PDO); (2) Injury; (3) Fatality; (4) Multiple fatality.

Figure 2. Contingency table for accident severity and vehicle type.

Frequency / Percent / Row Pct / Col Pct	Heavy	Light	Mid	Total
1	322	652	6	980
	13.31	26.95	0.25	40.51
	32.86	66.53	0.61	
	39.80	40.90	37.50	
2	419	891	7	1317
	17.32	36.83	0.29	54.44
	31.81	67.65	0.53	
	51.79	55.90	43.75	
3	66	50	2	118
	2.73	2.07	0.08	4.88
	55.93	42.37	1.69	
	8.16	3.14	12.50	
4	2	1	1	4
	0.08	0.04	0.04	0.17
	50.00	25.00	25.00	
	0.25	0.06	6.25	
Total	809	1594	16	2419
	33.44	65.89	0.66	100.00

Frequency Missing = 1

Due to sparse counts in certain cells of the contingency table, Fisher's Exact Test was used for analysis of the available aggregate collision data. There was a strong association between accident severity and collision type under a 0.05

significance level. Figure 2 presents the contingency table and Figure 3 presents the analysis results. Light and mid-sized vehicles are most prone to PDO collisions while heavy vehicles are involved in a disproportionally large number of fatal accidents and multiple-fatality collisions.

Figure 3. Statistical analysis results.

Chi-square test

Contingency table Statistic	DF	Value	Prob
Chi-Square	6	69.0197	<.0001
Likelihood Ratio Chi-Square	6	36.6619	<.0001
Mantel-Haenszel Chi-Square	1	4.7274	0.0297
Phi Coefficient	0.1689		
Contingency Coefficient	0.1666		
Cramer's V	0.1194		

Fisher's Exact Test

Table Probability (P)	3.766E-13
Pr <= P	1.745E-07

Effective Sample Size = 2419
Frequency Missing = 1

VEHICLE OVERLOADING

An estimated 30 percent of truck accidents in China involve vehicles carrying a load in excess of their rated capacity[4]. The excess weight reduces braking and steering performance to avoid hazards, and can lead to brake component wear, overheating, brake fade, and ultimately brake failure. When collisions do occur, the increased mass corresponds to more severe repercussions. Guardrails and other barriers may be unable to absorb and restrain the energy from an overloaded truck impact. Overloaded vehicles deteriorate the road pavement surface, emit more pollutants into the environment, and are more prone to spilling their load or overturning.

Overloading has become a common industry practice in China commercial vehicle operations. and it is not unusual for drivers to overload their vehicles by 30 to 50 percent[4].

DRIVER SURVEY DESIGN

In a survey (sample size=107) conducted among truck drivers on Jingjintang freeway in 2005, respondents were asked questions regarding a number of pertinent factors:

- Personal Information: Age, sex, education, place of residence.
- Vehicle Mechanical Specifications: Make and type, acceleration and deceleration performance, and braking performance.
- Driver Working Characteristics: Annual kilometers driven, speed selection, break frequency, vehicle loading, frequency of brake replacement, and accident experience in the last year.
- Driver Behavior and Attitude: Adherence to the speed limit, following rules of the road, and response to stimuli such as getting lost, being cut off by other drivers, or overtaking.

The surveyed drivers had an average of 9 years of driving experience and a mean age of 34 years. 92.5 percent ended their formal education in either junior, technical or high school, with one college graduate and seven respondents who only finished primary school. The drivers replaced their vehicle brakes at an average of approximately every 4 months, with relatively large variation (4.31 months standard deviation).

SURVEY PARAMETERS FOR OVERLOADING

The drivers were asked directly for information on the allowable load and typical actual load of their trucks. To handle the potential dilemma of drivers downplaying their overloading behavior, drivers were separately asked for the frequency of brake component replacement. Over half of respondents provided answers indicating that they did not overload their vehicles despite having a brake component replacement frequency more often than the statistical mean, indicating a high probability of misleading responses. Due to this disparity, brake replacement frequency was used in the logistic crash causation model to represent vehicle over-loading.

CRASH CAUSATION MODEL DEVELOPMENT

A logistic model was developed to estimate the effects of various driver behaviors, working characteristics, and vehicle specifications on crash involvement probability. Although none of the factors were influential under a 5% significance level, a 10% significance level model provided some insight into the factors affecting commercial vehicle crash probability. Figure 4 summarizes the 10% significance level logistic regression model.

MODEL INTERPRETATION

The following variables corresponded to an overall reduction in the probability of commercial vehicle crash involvement:

- Increased driver's experience as a commercial driver
- Increased interval between brake replacement (less overloading)
- Increased exposure (kilometers traveled per year)
- Increased maximum speed in a loaded condition

A driver with more experience has less likelihood to be involved in a collision, and vehicles operated within their design load capacities are less likely to have brake failures and are more likely to succeed in evasive maneuvers. However, both the exposure and maximum speed data initially seem counter-intuitive. When viewed from a perspective reflecting the Chinese motor vehicle environment, both of these variables likely reflect a more recently manufactured vehicle and an improved mechanical condition, as vehicles with mechanical deficiencies and out-of-date technologies tend to drive at lower speeds when loaded and may be unable to complete as many kilometers of travel in a given year.

The following variables corresponded to an overall increase in the probability of commercial vehicle crash involvement:

- Increased average speed traveled when loaded.
- The combination of increased driver experience with increased annual kilometers traveled.
- The combination of increased annual kilometers traveled with increased brake replacement frequency.

Unlike the maximum speed data, which more accurately reflected vehicle condition, the average speed employed by the driver reflected a tendency to operate at excessive speeds, which corresponded to an increase in crash risk. The confounding variables involving multiples of driver experience, annual kilometers traveled, and brake replacement frequency did not provide a clear insight into safety issues, as each individual variable had an impact reducing crash risk while the combination of the variables increased crash risk. The relatively small sample size (107 drivers), short study period (crashes within the last 1 year), and self-reporting design of the survey may factor in the inconclusive results.

Figure 4. Parameters for logistic model under 10% significance level.

Analysis of Maximum Likelihood Estimates

Parameter	DF	Standard Estimate	Wald Error	Chi-Square	Pr > ChiSq
Intercept 3	1	-2.7123	1.8195	2.2222	0.1360
Intercept 1	1	1.7521	1.5776	1.2334	0.2667
drage	1	-0.2019	0.1055	3.6620	0.0557
exposure	1	-0.2043	0.0787	6.7404	0.0094
replacebrake	1	-0.3914	0.2041	3.6771	0.0552
loadave	1	0.0521	0.0229	5.1724	0.0229
loadmax	1	-0.0355	0.0124	8.1706	0.0043
drage*exposure	1	0.0137	0.00795	2.9809	0.0843
exposure*replacebrak	1	0.0300	0.0153	3.8698	0.0492

CAUSATION SUMMARY

In this study, a logistic model concluded that lower collision probabilities resulted from more experienced drivers, vehicles loaded within their design capacities, vehicles in good mechanical condition with modern designs, and roadways with higher design standards. In contrast, drivers that consistently operated their vehicles at excessive speeds were more likely to be involved in a collision. A small percentage of commercial vehicle drivers are responsible for an inordinate amount of fleet crash occurrence[5], making the identification of high-risk individual driver characteristics crucial for reducing collision rates.

While this study reflects general trends in the commercial vehicle safety arena, the relatively small sample size (107 drivers), short study period (crashes within the last 1 year), limited geographic scope (a single freeway section), and self-reporting survey design limits the extent of applicability. Recommended future research into these topics may benefit from an increased sample size, an expanded time period for collision history, broader geographic coverage, in-vehicle instrumentation to obtain naturalistic data, and a focus on the differences between fatal collisions, non-fatal collisions, and near-crash conflict frequency.

COMMERCIAL VEHICLE SAFETY COUNTERMEASURES

Given the scope of the commercial vehicle safety challenge in China and the wide range of causes, a full spectrum of countermeasures should be evaluated for future implementation. Appropriate technical, regulatory, sociological, and enforcement strategies can be pursued to address known issues, requiring cooperation between drivers, motor carriers, regulatory agencies, enforcement groups, and governmental policy makers. Some proposed concepts for further study include:

1. Driver training: Systemized and institutionalized training for commercial vehicle drivers, including standards for licensure and effective enforcement, can help reduce errors and poor driving judgment.

2. Driver enforcement: Identifying and eliminating repeat offenders from the driving population can quickly address the elevated safety risk caused by a small number of dangerous drivers.
3. Driver health: Periodic mandatory physical check-ups should identify and remove any driver in imminent danger of a catastrophic health problem.
4. Driver fatigue: A combination of a regulatory framework for drivers and companies, effective enforcement of standards, and the implementation of reasonable work schedules for commercial vehicle drivers has been shown to reduce the number of fatigued or distracted drivers on the road[6].
5. Vehicle design and maintenance: A system of initial vehicle registration requirements and ongoing mechanical inspections can help eliminate mechanically defective vehicles from the fleet and make sure that critical vehicle components are maintained in a safe condition. A clear and comprehensive regulatory framework, appropriate incentives and disincentives, and effective enforcement are required.
6. Roadway design: In areas with high volumes of commercial vehicle traffic or high crash histories, geometric design and traffic control strategies can be implemented to reduce the number of conflicts involving large vehicles.
7. New technologies: In-vehicle crash avoidance technology, and onboard vehicle data recorders can directly reduce crashes and provide naturalistic data to help researchers. Technologies can also be used in coordination with driver and mechanical enforcement efforts. Disadvantages include privacy issues, affordability, and the overall acceptance by stakeholders. Economic incentives and governmental involvement may be needed to push the frontier of safety technology. Appropriate study must be completed to determine effectiveness of new technologies.
8. Alternative combinations of transportation modes: When measured by vehicle-mile or freight-pound-mile, highway transportation has a higher accident rate than air, marine or rail freight transport. When viable, shifting to alternative modes or intermodal combinations may be considered in order to improve efficiency and safety.

As further research continues and policies are developed to respond to identified issues, a great opportunity exists to significantly reduce the number of collisions, injuries, and fatalities on the roadway in China by focusing on commercial vehicle crash causes and potential countermeasures.

REFERENCES

---

[1] http://news.xinhuanet.com/legal/2006-01/12/content_4042310.htm
[2] http://www.ct.gov/dmv/cwp/view.asp?a+805&q=244794
[3] http://www.shanghai.gov.cn/shanghai/node2314/node4128/node14793/node14796/userobject30ai5964.html
[4] http://auto.sohu.com/2003/11/16/91/article215649183.shtml
[5] *Conference Proceedings*, "Future Truck and Bus Safety Research Opportunities", Transportation Research Board, 2005, pp. 2-3

Combining the Resources of the Existing State-wide Roadway Databases and Interactive Highway Safety Design Model (IHSDM) to Improve the Safety of Two-Lane Rural Highways

Mitsuru Saito[1], Kaitlin Chuo[2], and Thomas G. Jin[3]

[1] Professor, PhD, PE, F. ASCE, Department of Civil and Environmental Engineering, Brigham Young University, 368 Clyde Building, Provo, Utah 84602, USA; PH: (801) 422-6326; FAX: (801) 422-0159; email: msaito@byu.edu
[2] Graduate Research Assistant, Department of Civil and Environmental Engineering, Brigham Young University, 368 Clyde Building, Provo, Utah 84602, USA; PH: (801) 422-2685; FAX: (801) 422-0159; email: kaitlinchuo@gmail.com
[3] PhD Student, Department of Civil and Environmental Engineering, Brigham Young University, 368 Clyde Building, Provo, Utah 84602, USA; PH: (801) 422-2685; FAX: (801) 422-0159; email: tgjin2004@hotmail.com

Abstract

In order to systematically analyze the trends in crash occurrence, Utah Department of Transportation (UDOT) has developed over the years a state-wide crash database and a GIS-based Data Almanac to search through the crash database as well as other databases including annual average daily traffic (AADT). The Data Almanac is a powerful search engine for extracting necessary crash data. Using the crash data provides one way to describe what has happened to the sections of highways that the analysts are evaluating. However, evaluating what might happen when improvements are made requires going one step further. The Interactive Highway Safety Design Model (IHSDM) developed by the Federal Highway Administration (FHWA), may be used to help the engineers conduct crash prediction and design consistency evaluation and intersection safety review as part of highway safety audits. This software provides an opportunity to design, operation, and safety engineers for evaluating sections of two-lane highways from various aspects and identifying most cost effective improvement alternatives to such audited sections.

Introduction

Two-lane rural highways compose 77% of the nation's highway systems. Although vehicle miles of travel (VMT) wise they do not carry as much traffic as freeways and other major multi-lane highways, their share in the fatal crashes accounts for 44% (FHWA 2006). Head-on collisions and run-off-the-road crashes are some of the major crashes on two-lane rural roads.

It has been difficult to systematically evaluate the integrity of two-lane rural highways from various design and safety aspects. FHWA recently completed a suite of software programs named IHSDM to help the engineers conduct crash prediction and design consistency evaluation and intersection review and other related reviews of two-lane rural highways (FHWA 2006). This software provides a new method for design, operation, and safety engineers to systematically evaluate two-lane highways with high crash occurrences from various aspects of engineering in order to identify improvement alternatives that would be most cost effective to the analyzed sections. The desire is to proactively evaluate the need for improvement rather than reactively respond to the crashes that have occurred. Such evaluation results can be used as an integral part of safety audits of two-lane, two-way rural highways.

The purpose of this paper is to present a way to evaluate the safety of two-lane, two-way rural highways using IHSDM with the help of the existing resources available at UDOT.

Components of IHSDM

The current release of IHSDM (as of June 2007) for two-lane rural highways has five evaluation modules that provide different measures of the expected safety performance of an existing or proposed design. These modules are: (1) Policy Review, (2) Crash Prediction, (3) Design Consistency, (4) Intersection Review, and (5) Traffic Analysis. A sixth evaluation module, Driver/Vehicle, is being developed, but is not part of the current release (FHWA 2006). The reader may want to visit IHSDM's website for detailed discussions of each module (see FHWA 2006 in the reference section for its website). The present study focuses on modules 2, 3, and 4.

Objectives of the Study

Three objectives of the study were (1) Test the capability of IHSDM in locating crash prone segments and evaluating the level of design integrity of segments of two-lane rural highways, (2) Propose a method for incorporating IHSDM in the procedure for evaluation of safety and design integrity of two-lane rural highways, and (3) Prepare educational materials for an IHSDM training workshop. This paper presents the results of the work done to fulfill the first objective of the study.

Analysis Procedure

IHSDM requires horizontal and vertical alignment data, including valid station numbers. However, UDOT does not have plans and profiles of most of its two-lane, two-way highways built many years ago. Hence, if alignment data are not available, it is necessary to create surrogate horizontal and vertical alignments that reflect the current geometric conditions. This was the case of our study; it was necessary to create surrogate horizontal and vertical alignments using a highway geometric design CAD program. Sending a survey crew to get alignment data from the existing highways is costly. As we searched for the way to create surrogate alignment data, we found that UDOT had developed an image-based highway inventory system that covered practically entire sections of the highways managed by UDOT. This web-based program is called the Roadview Explorer (UDOT 2007b) and was used to conduct this study, together with UDOT's crash database. Figure 1 shows a screen shot of the Roadview Explorer. Latitutde, longitude and altitude data are shown at the bottom right corner of the screen shot.

Figure 1. Screen Shot of the user interface of the Roadview Explorer.

Obtaining Alignment Data. In order to convert GPS data available from the Roadview Explorer program (longitude, latitude, and altitude) into CAD readable data (northing, easting, and elevation), a feature of a software program called the Wathershed Modeling System (WMS), available at the author's institution, was used (EMRL 2007) and alignment data were created by using InRoads (Bentley 2006). The Roadview Explorer website is open to the public and its website can be found in the reference section (UDOT 2007b). The data includes milepost, latitude, longitude, altitude, and a photograph.

Obtaining Crash Data. UDOT's Data Almanac (Anderson, Glazier, & Perrett 2005; UDOT 2007a) was used to extract crash data, utilizing the Almanac's search and filter routine. By selecting the start and end year of analysis, the route number, the start and end mile post values, and a user defined filter, the user is able to extract necessary crash data. The website found in the reference section is currently available to authorized users only.

Sample Case Study

In this paper, highlights of the results of one case study are presented. The region engineer of UDOT Region Three recommended a segment of US-40 from MP (mile post) 35 to MP 45 for this study because this segment has experienced a large number of crashes. Though most of US-40 is a two-lane, two-way rural highway, this route is one of the few east–west routes crossing the Rocky Mountains to connect Colorado and Utah. The terrain type of the study section varies from level to rolling terrain, and the segment carried 4,060 AADT in 2006. The section has several curves. During the field visit, it was observed that the pavement condition of the study site was good, lane markings were clear and visible, and traffic signs all seemed functional.

As mentioned previously, the alignment data were created using the GPS data obtained from Roadview Explorer and converted into data by WMS to make them readable by InRoads. The stations created by InRoads were converted to milepost values to be entered into IHSDM.

As the purpose of this study was to pinpoint the locations where the crash rates are higher, evaluating the Crash Prediction Module (CPM) was our primary focus. To test the sensitivity of the IHSDM results, two alternative results were obtained: one evaluated with crash history and the other without crash history. After entering all the data of the studied section that were available and estimating the rest by internal models, IHSDM returned its predicted crash trend of the section.

Before analyzing the results, one must realize that the outcome from the CPM of IHSDM is simply an estimate resulting from the crash prediction models included in the software and that the model is not prophesying what will actually happen in the future. What is important is a general trend in the outcome. Figure 2 shows a graphical way of showing the results of the CPM. As seen in the figure, results from the model for the two cases have similar trends when compared to the crash history. The results shown in Figure 2 are presented with the number of crashes. For safety comparisons it is necessary to convert the number of crashes into a crash rate such as rate per million (or 100 million) vehicle miles of travel (MVMT or 100 MVMT).

The critical data required by IHSDM to run CPM are geometric alignments, speed limits, traffic volume, road side conditions, etc. Thus, it is logical to extrapolate that these are the elements that the software's crash prediction models take into account when calculating future crashes. The outcome from the analysis without crash history is purely based on its embedded crash prediction models. The result of the analysis with crash history is, of course, influenced by the actual crash history.

A closer look at the crash history can help us deduce the reason why the two results (with and without crash history) are somewhat dissimilar. It turned out that 60% of the crashes from the crash history were caused by collision with wild animals, which is not related to the road design at all. Collisions with wild animals scatter along the 10-mile stretch. Crashes other than the ones with wild animals, which accounted for the remaining 40% of the total crashes, were considered to be caused possibly by human errors, weather, and poor geometric design.

Crash occurrence of the two directions of the study section was then compared to pinpoint where crash "hot spots" might be located. Figure 2 shows the locations of the non-animal crashes in the eastbound and westbound directions of the study site. As shown in Figure 3, the westbound direction has significantly more crashes than the eastbound. This indicates that the geometric design of the westbound US-40, MP 35 to MP 45, needs more attention than the eastbound. Concluding the discussion made above, locations of hot spots may be estimated. In the westbound direction, the curve segments in the mid portion of the study site seem to be hot spots, while in the eastbound direction the curves closer to the 35 milepost seem to be accident prone.

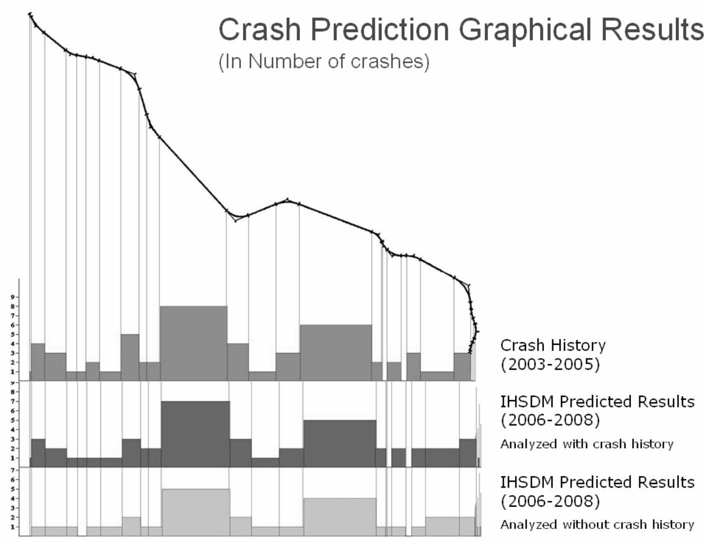


Figure 2. Graphical results in number of crashes.

Conclusions

This paper presented how state-wide databases that already exist can be effectively used for a new study. Roadview Explorer and the Data Almanac are the two useful resources that were available for this study. Costly surveying of the existing two-lane, two-way highways to run the CPM of IHSDM can be avoided by using of the Roadview Explorer's GPS data. Once horizontal and vertical alignments are available, IHSDM becomes a useful tool for safety audits of two-lane, two-way highways and improvement alternatives.

Creating surrogate horizontal and vertical alignments data by InRoads, however, was a challenge. Fitting a horizontal alignment through a narrow stretch of triangulated digital terrain model created by the GPS data of the two directions of the highway provided by the Roadview Explorer took a significant amount of time for a novice InRoads user. If the analysis is focused on dealing with newly constructed highways, IHSDM can be fed by the horizontal and vertical alignment data available from CAD-based highway geometric design software. But, obviously, this is not the case for many two-lane, two-way highways built many years ago, whose design drawings have been long lost. Also, such original drawings would be useless once improvements were made. Hence, unless as-built drawings were kept for all improvements, one must come up with a method to get alignment data from the existing roadway surface. Simplification of this task is the key for integrating IHSDM into a routine safety audit of two-lane, two-way highways.

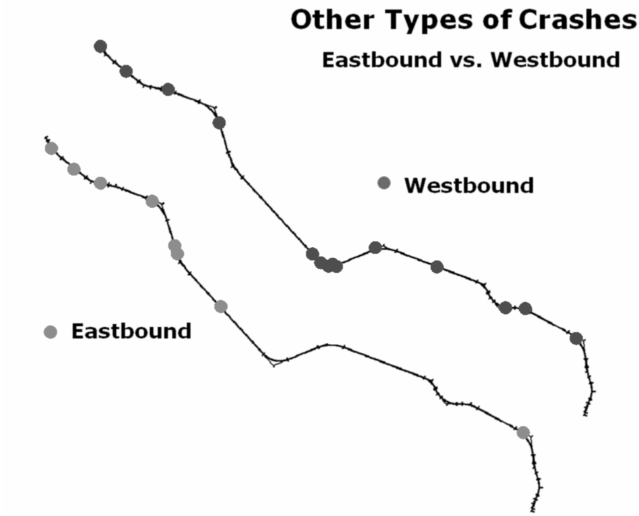


Figure 3. The plot of other-than-animal related crashes: Eastbound vs. Westbound.

References

Anderson, D., C. Glazier, and G. Perrett. (2005). Data Almanac User's Manual. Utah Department of Transportation, R&D Division, Salt Lake City, Utah, USA.

Bentley Inc. (2006). InRoads version 8.04.

Environmental Modeling Research Lab (EMRL). (2007). Watershed Modeling Software (WMS). Brigham Young University, Utah, USA.

Federal Highway Administration (FHWA). (2006). Interactive Highway Safety Design Model (IHSDM), 2006 version. http://www.tfhrc.gov/safety/ihsdm/ihsdm.htm

Utah Department of Transportation (UDOT). (2007a). Data Almanac. http://168.178.120.60/udotAsset/Login/Login.asp

Utah Department of Transportation (UDOT). (2007b). Roadview Explorer 2.0. http://www.roadview.udot.utah.gov/

Unascertained Measurement Evaluation on Maintenance Quality of Highway Pavement

WEN Chang-ping,[1] T. A. Parsons[2], and Bashar AlMalik[3]

[1]Department of Urban Construction, SHAO-Yang University, SHAO-Yang City, HU-Nan Province, China; PH (+86) 739-5391265; email: urbanwen@163.com.
[2]Applied Research Associates, 112 Monument Place, Vicksburg, MS, 39180; PH (601) 629-6165; email: tparsons@ara.com.
[3]Civil Engineering Dept., School of Engineering, King Saud University, Riyadh, Saudi Arabia balmalik@gmail.com

Abstract: Unascertained measurement theory (UMT) is used to create an evaluation method to determine the condition of cement concrete pavement (CCP). The proposed method allows for composite evaluation of the CCP that is consistent with the actual condition of the CCP. The resulting measurement model is designed to strictly adhere to UMT, the measurement functions are rigorously constructed, and a confidence criterion is introduced due to the ordered division the evaluation space. This paper also defines a method to determine index weight using similar weight by similar figures that provides an objective standard and avoids using the subjective opinions of experts when defining index weight. The index weight instead reflects the relative importance of the index in assessing of pavement maintenance quality. The proposed method is explained in detail by means of an example to encourage its use.

Key words: *Pavement, maintenance quality, evaluation, unascertained measurement, attribute recognition*

1 Introduction

The maintenance of highway pavement is one of the most important and most expensive tasks of day-to-day pavement preservation and management. The effectiveness of pavement preservation programs must be monitored to determine the optimum allocation of pavement maintenance resources. It is therefore necessary to objectively evaluate the maintenance quality of highway pavement to determine if the pavement preservation activities are effective.

Technical specifications JTJ073-96 and JTJ073.1-2001 define a composite index for evaluating maintenance quality of CCP. A composite index is one that combines multiple indices into a single index. The indices used to evaluate the maintenance quality of CCP and their assocuated criteria are listed in Table 1.

The evaluation of CCP according to these criteria is a multi-index and multi-level problem that can be addressed using UMT (LIU Kai-di, et al, 1999; LIU Kai-di, et al, 2000). To demonstrate the applicability of UMT to the evaluation of CCP using these criteria, five sections of a major highway were selected as an example. The inspection data for each of the highway sections are shown in Table 2.

2 Unascertained Measurement Model

The unascertained measurement model consists of an object space X and its associated attribute space U. The object space X is composed of sections x_i that can be mapped into attribute space U. Attribute space U is divided into four classes (C_1, C_2, C_3, and C_4) that correspond to pavement grade, which are

Project supported by Hunan Provincial Natural Science Foundation of China (No. 07JJ6080)

defined such that $C_1 < C_2 < C_3 < C_4$ and such that the class of any section x_i can be known. Each section x_i is mapped into the appropriate class in attribute space U using three evaluation indices j, $I_1(PCI)$, $I_2(\delta)$, $I_3(SRC)$, according to the criteria in Table 1. The value of *jth* evaluation index of *ith* section is denoted as x_{ij}.

Table 1 Evaluation indices and criteria.

Indices	Grade			
	C_1 (Poor)	C_2 (Common)	C_3 (Good)	C_4 (Better)
Pavement condition index (PCI)	PCI<50	50≤PCI<70	70≤PCI<85	85≤PCI
Deviation of pavement evenness (δ)	δ>4.5	4.5≥δ>3.5	3.5≥δ>2.5	2.5≥δ
Skid resistance condition (SRC)	SRC<0.28	0.28≤SRC<0.35	0.35≤SRC<0.40	0.40≤SRC

Table 2 Inspection data for example CCP.

Section	Length (m)	PCI (I_1)	δ (I_2)	SRC (I_3)
1	350	23.8	3.9	0.44
2	450	93.6	3.4	0.49
3	480	77.6	3.9	0.50
4	600	87.7	3.8	0.39
5	760	98.9	2.0	0.38

2.1 Unascertained Measurement Functions

Let $\mu_{ijk} = \mu(x_{ij} \in C_k)$ denote the unascertained measurement of x_{ij} with attribute C_k that satisfies the constraints $0 \le \mu(x_{ij} \in C_k) \le 1$ and $\mu(x_{ij} \in U) = 1$. For this example (i=1, 2, 3, 4, 5; j=1, 2, 3; k=1, 2, 3, 4), the matrix $(\mu_{ijk})_{5\times4}$ can be constructed according to Eq. 1 and is referred to as the evaluation matrix of unascertained measurement of x_i.

$$(\mu_{ijk})_{5\times4} = \begin{pmatrix} \mu_{i11} & \mu_{i12} & \mu_{i13} & \mu_{i14} \\ \mu_{i21} & \mu_{i22} & \mu_{i23} & \mu_{i24} \\ \mu_{i31} & \mu_{i32} & \mu_{i33} & \mu_{i34} \\ \mu_{i41} & \mu_{i42} & \mu_{i43} & \mu_{i44} \\ \mu_{i51} & \mu_{i52} & \mu_{i53} & \mu_{i54} \end{pmatrix} \qquad \textbf{Eq. 1}$$

The unascertained measurement functions $\mu_{ijk}(t) = \mu(x_{ij} \in C_k)$ are then constructed using the criteria from Table 1 as follows (CHEN Qian-sheng, 1997):

$$\mu_{i11}(t)=\begin{cases}1 & t<47.5\\ \dfrac{62.5-t}{15} & 47.5\le t\le 62.5\\ 0 & t>62.5\end{cases},\qquad \mu_{i12}(t)=\begin{cases}0 & t<47.5\\ \dfrac{t-47.5}{15} & 47.5\le t\le 62.5\\ \dfrac{77.5-t}{15} & 62.5<t\le 77.5\\ 0 & t>77.5\end{cases},$$

$$\mu_{i13}(t)=\begin{cases}0 & t<62.5\\ \dfrac{t-62.5}{15} & 62.5\le t\le 77.5\\ \dfrac{92.5-t}{15} & 77.5<t\le 92.5\\ 0 & t>92.5\end{cases},\qquad \mu_{i14}(t)=\begin{cases}0 & t<77.5\\ \dfrac{t-77.5}{15} & 77.5\le t\le 92.5\\ 1 & t>92.5\end{cases};$$

$$\mu_{i21}(t)=\begin{cases}1 & t>5.0\\ t-4.0 & 4.0\le t\le 5.0\\ 0 & t<4.0\end{cases},\qquad \mu_{i22}(t)=\begin{cases}0 & t>5.0\\ 5.0-t & 4.0\le t\le 5.0\\ t-3.0 & 3.0\le t<4.0\\ 0 & t<3.0\end{cases},$$

$$\mu_{i23}(t)=\begin{cases}0 & t>4.0\\ 4.0-t & 3.0\le t\le 4.0\\ t-2.0 & 2.0\le t<3.0\\ 0 & t<2.0\end{cases},\qquad \mu_{i24}(t)=\begin{cases}0 & t>3.0\\ 3.0-t & 2.0\le t\le 3.0\\ 1 & t<2.0\end{cases};$$

$$\mu_{i31}(t)=\begin{cases}1 & t<0.275\\ \dfrac{0.325-t}{0.05} & 0.275\le t\le 0.325\\ 0 & t>0.325\end{cases},\qquad \mu_{i32}(t)=\begin{cases}0 & t<0.275\\ \dfrac{t-0.275}{0.05} & 0.275\le t\le 0.325\\ \dfrac{0.375-t}{0.05} & 0.325<t\le 0.375\\ 0 & t>0.375\end{cases},$$

$$\mu_{i33}(t)=\begin{cases}0 & t<0.325\\ \dfrac{t-0.325}{0.05} & 0.325\le t\le 0.375\\ \dfrac{0.425-t}{0.05} & 0.375<t\le 0.475\\ 0 & t>0.475\end{cases},\qquad \mu_{i34}(t)=\begin{cases}0 & t<0.375\\ \dfrac{t-0.375}{0.05} & 0.375\le t\le 0.425\\ 1 & t>0.425\end{cases}.$$

2.2 Index Weight

Index weight, or the importance of an individual index when calculating a composite index, must be accounted for in the unascertained measurement model. The proposed model uses index weight to objectively calculate similar weight, which is a measure of the importance of an index to the unascertained measurement model. The model calculates similar weight by similar figures (CHEN Qian-sheng, 1997; PANG Yan-jun, et al, 2001), thus avoiding subjective "expert opinion" when determining the relative importance of a particular index within the unascertained measurement model. The method for computing similar weight is shown using the example data set in the following paragraphs.

(1) The example data set uses three equally weighted indices to calculate the composite index. For equally weighted indices, compute the unascertained measurement matrix, $(\mu_{ik})_{5\times4}$, according to $\mu_{ik} = \frac{1}{3}\sum_{j=1}^{3}\mu_{ijk}$.

(2) Compute the similar coefficient r_j according to Eq. 2.

$$\begin{aligned} r_j &= \frac{1}{5}\sum_{i=1}^{5}(\mu_{ij1}, \mu_{ij2}, \mu_{ij3}, \mu_{ij4}, \mu_{ij4})\cdot(\mu_{i1}, \mu_{i2}, \mu_{i3}, \mu_{i4})^{\mathrm{T}} \\ &= \frac{1}{5}\sum_{i=1}^{5}\sum_{k=1}^{4}\mu_{ijk}\cdot\mu_{ik} \end{aligned} \quad \textbf{Eq. 2}$$

(3) Compute the similar weight w_j according to Eq. 3.

$$w_j = \frac{r_j}{\sum_{j=1}^{3} r_j} \quad \textbf{Eq. 3}$$

2.3 Composite Attribute Measurement

The unascertained measurement function, Eq. 1, can then be used to compute the unascertained measurement of single index $\mu_{ijk} = \mu(x_{ij} \in C_k)$. This is then combined with the similar weight according to Eq. 4 to determine the composite attribute measurement $\mu_{ik} = \mu(x_i \in C_k)$ of a single section.

$$\mu_{ik} = \mu(x_i \in C_k) = \sum_{j=1}^{3} w_j \mu_{ijk} \quad \textbf{Eq. 4}$$

2.4 Attribute Recognition

Attribute recognition is the final step in mapping a section x_i into a class (grade) k_0 in the attribute space *U*. The confidence criterion (λ) is employed in attribute recognition analysis according to Eq. 5 or Eq. 6 for an attribute space defined such that $C_1 < C_2 < C_3 < C_4$, as in the proposed model (CHEN Qian-sheng, 1997). If the criteria of Eq. 5 or Eq. 6 are met, then section x_i belongs to class C_{k_0}. For purposes of the example, assume λ =0.5.

$$k_0 = \min\left\{k : \sum_{l=1}^{k}\mu_{xi}(x_i \in C_l) \geq \lambda, 1 \leq k \leq p\right\} \quad \textbf{Eq. 5}$$

$$k_0 = n - \min\left\{k : \sum_{l=0}^{k}\mu_{xi}(x_i \in C_{p-l}) \geq \lambda, 0 \leq k \leq p-1\right\} \quad \textbf{Eq. 6}$$

3 Evaluation Procedure

The evaluation procedure using UMT is demonstrated in the following paragraphs using the example data.

(1) Compute the unascertained measurement and composite attribute measurement of each section for each index. The result of the computation for section 1 is shown in Table 3. The unascertained measurements of the other sections may be obtained for each index by similar means. Use these data to construct the unascertained measurement matrix, $(\mu_{ik})_{5\times 4}$, as shown in Eq. 7.

Table 3 Attribute measurement of No.1

Indices	Surveyed Value	Unascertained Measurement of a Single Section			
		C_1	C_2	C_3	C_4
PCI	23.8	1	0	0	0
δ	3.9	0	0.9	0.1	0
SRC	0.44	0	0	0	1
$\mu_{ik}=\frac{1}{3}\sum_{j=1}^{3}\mu_{ijk}$		0.333	0.300	0.033	0.333

$$(\mu_{ik})_{5\times 4}=\begin{pmatrix} 0.333 & 0.300 & 0.033 & 0.333 \\ 0 & 0.133 & 0.200 & 0.667 \\ 0 & 0.300 & 0.364 & 0.336 \\ 0 & 0.267 & 0.407 & 0.327 \\ 0 & 0 & 0.300 & 0.700 \end{pmatrix} \quad \textbf{Eq. 7}$$

(2) Compute the similar coefficient, $r_j=(2.417 \quad 1.748 \quad 2.058)$, and the similar weight, $w_j=(0.388 \quad 0.281 \quad 0.331)$.

(3) Compute the composite attribute measurement matrix $\mu_{ik}=\mu(x_i\in C_k)$ according to Eq. 4 as shown in Eq. 8.

$$\mu_{ik}=\mu(x_i\in C_k)=\sum_{j=1}^{3}w_j\mu_{ijk}=\begin{pmatrix} 0.388 & 0.253 & 0.028 & 0.331 \\ 0 & 0.112 & 0.169 & 0.719 \\ 0 & 0.253 & 0.414 & 0.333 \\ 0 & 0.225 & 0.412 & 0.363 \\ 0 & 0 & 0.298 & 0.702 \end{pmatrix} \quad \textbf{Eq. 8}$$

(4) Assign sections x_i to classes C_k using attribute recognition. Assuming the confidence criterion $\lambda=0.5$, determine the lowest grade for which the cumulative composite attribute measurement of a section is greater than 0.5, thus satisfying Eq. 5. For the example data, sections 1 through 5 would be assigned to grades C_2, C_4, C_3, C_3 and C_4, respectively.

4 Conclusions

This paper presents an unascertained measurement evaluation model for CCP. The model is rigorously constructed using the principles of unascertained measurement theory and simple in application. The evaluation method allows computation of a composite condition index using quantitative computation rather than qualitative analysis. The evaluation index weight is defined using similar weight by similar figures, thus it avoids using subjective "expert opinions" to assign evaluation index weight and objectively reflects the relative importance of each index in the composite index. It is hoped that this study will encourage the application of the proposed model when assessing CCP condition using multi-index and multi-level composite indices.

References

CHEN Qian-sheng. (1997). "Attribute mathematical model and fuzzy mathematical model for quality assessment" (in Chinese), *Application of Statistics and Management*, 16(6), 18-23.

CHEN Qian-sheng. (1997). "Attribute sets and attribute synthetic assessment system" (in Chinese). *Systems Engineering Theory & Practice*, 17(9): 1-8

CHEN Qian-sheng. (1997). "Attribute recognition theoretical model with application" (in Chinese). *Universitatis Pekinnensis, Acta Scientianrum Naturalium*, 33(1): 12-20

LIU Kai-di, PANG Yan-jun, SUN Guang-yong. (1999). "The unascertained measurement evaluation on a city's environmental quality" (in Chinese). *Systems Engineering Theory & Practice*, 19(12): 52-58

LIU Kai-di, PANG Yan-jun, YAO Li-gen, et al. (2000). "The unascertained measurement model of the synthetical evaluation on atmosphere environmental qualities" (in Chinese). *Environmental Science*, 21(3): 11-15

PANG Yan-jun, LIU Kai-di, ZHANG Bo-wen. (2001). "The method of determining the objective index weight in the synthetic evaluation system" (in Chinese). *Systems Engineering Theory & Practice*, 21(8): 37-42

State Communication Administration of China. (1996). Technical Specifications of Highway Maintenance. JTJ073-96.

State Communication Administration of China. (2001). Technical Specifications of Cement Concrete Pavement Maintenance for Highway. JTJ073.1-2001.

Investigating Crash Characteristics of Multiple-Lane Highways in Louisiana

Xiaoduan Sun, Ph.D. PE[1]
Murali Krishna Bommareddy[2]

ABSTRACT

There are about 1,520 miles of multiple-lane (excluding freeway) highways in Louisiana that amounts to 9% of all highways on the Louisiana Department of Transportation and Development (LaDOTD) roadway system. However close to 50% of total crashes and 22% of total fatal crashes occur on multiple-lane highways annually. Reducing number of crashes on multiple-lane highways is crucial for LaDOTD.

This paper investigates characteristics of crashes on multiple-lane highways in Louisiana and identifies the top sites for safety improvements by the network screening methods introduced by the on-going Highway Safety Manual projects. Classifying the multiple-lane highways by median (divided and undivided) and area (urban and rural), the results show that urban undivided highway is the most dangerous highway with the highest overall crash rate, and rural divided highways is the most fatal highway with the highest fatality rate in Louisiana. Unlike other types of highways, collisions with other vehicles consist of more than 70% of all crashes. Rear-end crash is the most common type of collision on this type of highways. The results from network screening reveal the characteristics of top ranking sites by four different methods.

Introduction

Although rural multiple-lane highways only consist of 9% of total highways under Louisiana Department of Transportation and Development (LaDOTD) roadway system, about 50% of all crashes and 22% of all fatal crashes in Louisiana occur on these highways annually. As shown in Table 1, more than half of crashes took place at intersections, and urban undivided highway has the highest crashes per mile and crash rate (crashes per one million VMT) in 2004 while rural divided highway has the highest fatal crash rate (higher than the national overall average fatal crash rate of 1.45 in 2004). As expected undivided highways have higher crash rate than divided highways do on both rural and urban settings.

[1] Professor, Civil Engineering Department, University of Louisiana
[2] Graduate Student, University of Louisiana

Table1. Summary of Basic Crash Statistics

	Rural Undivided	Rural Divided	Urban Undivided	Urban Divided
Intersection Crashes	248	967	8,166	17,556
Segment Crashes	169	1,693	4,004	13,360
Total	417	1,694	12,170	30,916
Length (miles)	70.82	529.15	257.43	663.42
Weighted ADT	6,953.26	11,676.05	17,291.61	25,066.21
Crash Rate	2.32	1.18	7.49	5.09
No. of Fatalities	0	53	30	94
Fatality rate	0.00	**2.35**	1.85	1.55

In Louisiana, more than half of these highways (64%) serve as principal arterials and about 28% serve as minor arterials. While providing both mobility and accessibility to the traveling public, multiple-lane highways possess unique crash characteristics. Analyzing these characteristics can shed light on safety improvement effort.

Crash Characteristics

Table 2 lists the percentage distribution of crash types. Unlike two-lane highways where running-off-roadway comprise two-third of all crashes, collision with other vehicles in traffic is the most common type of crashes particularly on urban highways followed by the collision with fixed object. Because of difference in roadside development, the percentage of collision with fixed object is higher on rural highways than that on urban highways, which indicate improving roadside safety is important for rural multiple-lane highways.

Table 2.Distribution of Crash Type

	Rural		**Urban**	
	Divided	**Undivided**	**Divided**	**Undivided**
Running off roadway	*11.69*	*4.08*	*1.66*	*0.94*
Overturning on roadway	*0.79*	*0.72*	*0.24*	*0.12*
Collision with pedestrian	*0.23*	*0.72*	*0.30*	*0.37*
Collision with other vehicle	***69.14***	***70.26***	***92.28***	***92.05***
Collision with parked vehicle	*0.60*	*1.92*	*0.42*	*0.76*
Collision with train	*0.26*	*0.48*	*0.06*	*0.09*
Collision with bicyclist	*0.34*	*1.68*	*0.42*	*0.47*
Collision with animal	*2.03*	*1.44*	*0.21*	*0.03*
Collision with fixed object	*10.90*	*16.31*	*3.55*	*4.63*
Collision with other object	*0.98*	*1.20*	*0.22*	*0.16*
Other non-collision on road	*3.05*	*1.20*	*0.64*	*0.37*

Further investigation on type of collisions reveals that rear-end collision is the most common type of collisions on multiple-lane highways particularly on urban areas, which is largely related to heavy traffic volumes during rush hours. Generally 88% of all rear-end and 84% of sideswipe crashes occur between 7am and 8pm. As shown in Figure 1, undivided highways have slightly lower percentage of rear-end collisions for both rural and urban multiple-lane highways, which may have something to do with the available space at left of travel lane. The percentage of right-angle crashes is the highest on multiple-lane highways comparing to freeway and 2-lane highways. About 73% of right-angle crashes, one of the most dangerous types of crashes, and close to 75% of left-turn crashes occurred at intersections.

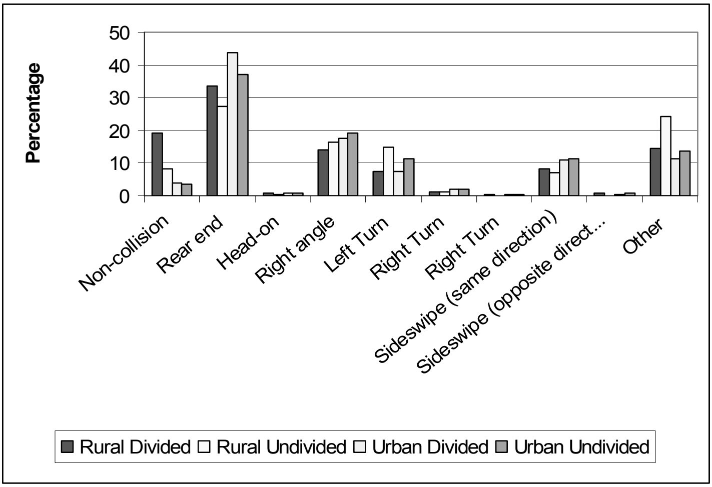


Figure 1. Distribution of Collision Type

Identify the Most Promising Site for Safety Improvement

Due to the financial constrains, it is impossible to implement highway safety improvement at all locations. Effectively promoting highway safety calls for roadway safety management system. A roadway safety management process consists of a series of steps including reviewing an existing transportation network, identifying sites with potential for safety improvement, identifying crash countermeasures, prioritizing potential improvements, and monitoring outcomes after improvements have been constructed. Network screening is the first step in roadway safety management process. Over the years, many screening methods have been proposed. The current on-going Highway Safety Manual has documented the best practice in nine highway safety screening methods with different data requirements and strength (1). The availability of data has restricted our analysis to four screening methods: Frequency, Equivalent Property Damage Only (EPDO), Relative Severity Index

(RSI), and Crash Rate. Instead of using roadway segment with uniform length, we used control sections for the screening analysis simply because they are readily available in a database format. Control section was historically defined with homogenous attributes such as lane width and Average Daily Traffic (ADT). Since the length of control section varies greatly, the ranking from the first three methods are derived by the combination of per mile and per control section. The EPDO method uses a formula to assign weighting factor to crashes by severity (fatal, injury, property damage only) to develop a single severity score per site. The weighting factors are basically computed relative to Property Damage Only. Without any detailed cost data from Louisiana, we used the information from a 2005 Federal Highway Administration (FHWA) published report (2), which results in a weighting number 542 for fatal crashes and 11 for injury crashes. RSI uses an average crash cost by type to identify the sites which has higher than average crash cost. In addition to crash frequency and crash severity distribution at each site, crash type and cost are important input for this method. Again, without reliable data from Louisiana, we used the information from the FHWA report on the cost of crashes. Considering the differences among four different highways classified in this study, the rankings were developed by type of highway. Table 3 gives the top 10 sites (excluding intersections) on Urban Divided Multiple-lane highways.

Different screening methods obviously result in very different rankings as evidenced in Table 3. While no control section was picked up in the top 10 list by all four methods, Control Section 283-08 is ranked the highest by the first three methods because they are all related to crash frequency. High frequency of crashes generally occurs at segments with high traffic volume. The crash rate method is designed to identify relatively short control sections with small ADT because of its. Overall, the RSI method gives out the highest total number of crashes from top 10 sites. As commonly recognized, using crash rate alone to identify problem sites for improvement has significant bias in highway safety analysis. As shown in Table 4, control section 283-80 on Highway 90 has another sub-control section with undivided feature that is ranked highest under urban undivided highway category. This segment has different segment log mile from the segment under urban divided category.

The top one ranking sites in Table 4 further reveal the characteristics of all top one sites identified by four different screening methods. Although the first three methods are all somewhat related to crash frequency, they emphasize different aspect of crash frequency. Top one site from urban undivided and rural divided by EPOD has the highest traffic fatality rate (fatalities per million VMT). The rest of top one sites either have no fatality or fatality rate is small. RSI method picks up the costliest sites where the more costly crashes occurred such as head-on collision that costs 10 times higher than area-end collision on average.

It is important to point out that different highways yield very different values for the top one site. Because the urban divided highways carry large ADT, they have the highest ranking value for all four methods. Particularly, the top crash rate of urban divided highway is more than 15 times higher than that of rural undivided. If screening analysis were performed by all multiple-lane highways together, only urban highways could get attention for safety improvement.

Table 3. Top 10 Sites on Urban Divided Four-lane Highways

Ranking	Frequency				EPOD				RSI				Crash Rate			
	Control section	**Length**	**3-yr. Averge Crashes**	**Average ADT**	**Control section**	**Length**	**3-yr. Averge Crashes**	**Average ADT**	**Control section**	**Length**	**3-yr. Averge Crashes**	**Average ADT**	**Control section**	**Length**	**3-yr. Averge Crashes**	**Average ADT**
1	283-08	1.07	226.33	108,225	283-08	1.07	226.33	108,225	283-08	1.07	226.33	108,225	025-08	0.03	15.33	19,300
2	007-90	0.78	157.67	50,400	007-90	0.78	157.67	50,400	283-09	0.98	157.00	83,867	255-01	0.05	5.00	11,900
3	283-09	0.98	157.00	83,867	007-08	0.43	56.33	47,033	283-08	0.36	60.67	136,100	060-01	0.08	23.00	32,100
4	080-02	2.21	244.33	36,533	283-08	0.36	60.67	136,100	283-09	1.24	124.00	83,867	046-02	0.06	10.00	27,133
5	258-02	0.58	89.33	33,467	808-07	0.23	54.67	44,333	007-90	0.78	157.67	50,400	853-36	0.09	12.67	19,433
6	817-40	0.44	77.67	39,167	283-09	0.98	157	83,867	283-08	0.43	64.00	81,167	855-14	0.07	7.33	19,100
7	006-90	1.07	121.33	41,567	007-02	0.51	33.33	40,533	006-90	1.07	121.33	41,567	417-02	0.06	5.33	21,800
8	283-09	1.24	124.00	83,867	006-90	1.07	121.33	41,567	006-03	2.02	158.33	43,200	283-30	0.09	30.33	56,200
9	283-08	0.36	60.67	136,100	283-09	0.86	41.33	62,467	258-02	0.58	89.33	33,467	077-04	0.08	9.67	22,700
10	808-07	0.23	54.67	44,333	817-40	0.82	77.67	20,967	080-02	2.21	244.33	36,533	010-06	0.08	9.67	23,367
Total		8.96	1,313	657,525		7.11	986.33	635,492		10.74	1,403	698,392		0.69	128.33	253,033

Note: the number of crashes and ADT are the average of three years of each control section (20002-2005)

Discussion

The analysis presented by this study yield following discussions:

1. As they share the highest proportion of crashes comparing to other types of highway in Louisiana, multiple-lane highways deserve special attention and fund for safety improvements.
2. The targeted type of crashes in general should be multiple-vehicle collisions particularly rear-end collisions at urban areas during rush hours. This type of crash reduction can not be effective achieved through highway engineering project alone. Enforcement, driver education and necessary traffic control devices such as warning sign must be promoted together. Attention must be made to the 12% running-off-roadway crashed on rural multiple-lane highways.
3. It is necessary to screening network for potential improvement sites by highway features (divided and undivided) and environment settings (urban and rural). Otherwise the high risk sites on rural highways could hardly be identified.
4. Screening method related to crash frequency works better than crash rate method does if segment length is not uniform. Theoretically, the best screening method involves safety performance function (SPF) because it overcomes the bias caused by the so-called regression-to-the-mean. The first edition of HSM will provide SFP for rural multiple-lane highways and some urban multiple-lane highways.

Table 4. Characteristics of Top One Sites

	Method	Site Attributes			Value of Ranking Method	Crash Data		
		Mile	Control Section	ADT		# of Crashes	# of Injuries	# of Deaths
Urban Divided	Frequency	1.71	283-08	107,688	225 crashes	225	78	0
	EPOD	1.71	283-08	107,688	1,230.84	225	78	0
	RSI	1.71	283-08	107,688	$2,739,821	225	78	0
	Crash Rate	0.03	025-08	18,300	2,162.5	13	3	0
Urban Undivided	Frequency	2.32	283-08	107,688	207 crashes	207	54	2
	EPOD	0.85	060-01	17,700	1,360	18	4	2
	RSI	2.32	283-08	107,688	$881,000	207	54	2
	Crash Rate	0.05	279-04	4,500	1,462.5	6	1	0
Rural Divided	Frequency	2.28	424-05	27,900	49 crashes	49	19	0
	EPOD	4.2	019-03	10,200	266.4	9	2	2
	RSI	2.25	005-07	25,600	$427,882	21	6	0
	Crash Rate	0.02	262-07	10,800	1,902.6	3	1	0
Rural Undivided	Frequency	0.24	056-06	13,600	7 crashes	7	0	0
	EPOD	0.26	262-07	18,000	57.69	1	1	0
	RSI	0.7	007-03	20,606	$661,873	12	3	0
	Crash Rate	0.03	055-30	22,100	137.3	1	0	0

Reference

[1] "Network Screening" a draft Chapter 11 for upcoming Highway Safety Manual, Kittelson & Associates, Inc., 2007

[2] "Crash cost estimates by maximum police-reported injury severity within selected crash geometries", FHWA, October 2005

Relationship between Weaving Length and Traffic Safety in Road Interchanges

Yuan XU[1], Shunying ZHU[2], Hong WANG[2], Juxiang GUAN[1], Qinmei SUN[1], Bing LIU[1], Dan LIU[1], Jing XIA[1], Constantinos ANTONIOU[3]

[1] Master Student, Dept. of Traffic Engineering, Wuhan University of Technology, 1040 Heping Ave., Wuchang, Wuhan, Hubei, China. 430063. E-mail: xyjtgc@126.com
[2] Professor, Dept. of Traffic Engineering, Wuhan University of Technology, 1040 Heping Ave., Wuchang, Wuhan, Hubei, China. 430063. E-mail: zhusy2001@163.com
[3] Research Associate, Dept. of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge 02139, USA. E-mail: costas@mit.edu

Abstract: Weaving areas in road interchanges are dynamic areas, where vehicles change lanes frequently to complete merging and diverging operations in a short distance. As a result, traffic flow is rather unpredictable and volatile, with negative impact on the safety and efficiency of the entire interchange. This paper studies the impact of weaving length on traffic safety and efficiency. An orthogonal experimental design is developed in order to analyze the relationship between traffic conflicts and (i) weaving length, (ii) traffic volume and (iii) ratio of weaving volume to mainline flow. A microscopic simulation model, VISSIM, is used to model the Youyi Avenue interchange of Wuhan Changjiang River Tunnel and extract the selected as the traffic safety indices for each scenario, i.e. traffic conflicts and conflict velocity. Finally, a reasonable weaving length for the interchange that balances traffic safety and efficiency is estimated using the difference analysis method.

Key Words: *Weaving area; Weaving length; Traffic conflict; Traffic safety*

1 Introduction

Interchanges are critical points in the transport network, in terms of road safety, as accidents tend to concentrate around them. Vehicles need to perform maneuvers in a limited area in and around interchanges, in order to perform weaving, merging, diverging and other tasks, often involving more than one lane changes in very short time. These lane-changes result in traffic conflicts, that are not present in general freeway segments, and create turbulence in the traffic flow. The weaving length is defined as the minimum distance which allows drivers to complete the desired lane changes successfully, and has a significant influence to the safety and efficiency of the weaving area. The Highway Capacity Manual (HCM, 2000) describes three main types of weaving configurations (type A, B, and C), with type A being the most common at present. Therefore, the remainder of this paper focuses on type A weaving configurations.

The following assumptions have been explicitly made in the remainder of this paper:

(1) The study focuses on type A weaving areas with four lanes. (2) The design speed is 60km/h. (3) Traffic signs upstream of the intersection are according to traffic regulations and do not interfere with the driving task, and (4) road conditions are ideal (i.e. dry weather and even slope).

2 Literature review

Jillella (2001) found that as the length of the weaving area decreases, ceteris paribus, the intensity of the lane changing and the turbulence in the weaving area increase, while the average speed of the vehicles in the weaving area reduces. Golob (2002) found that more conflicts occurred in type A weaving areas during off-peak periods, in the left-most lane, at night, and during wet conditions. Ahammed (2005) found that an acceleration lane length of 425m was adequate for all geometric and traffic conditions. Sahan (2005) developed a model to compute the length of acceleration and deceleration lanes based on the expected collision frequency. More recently, Xiaojuan, (2006) introduced traffic conflict technology (TCT) to the safety evaluation of weaving areas. Overall, according to the literature weaving length has a significant relationship with traffic conditions in weaving sections.

4 Traffic safety and efficiency model

In this paper the traffic conflicts, the relative conflicts, the conflict intensity and the capacity are considered as the main indices for the analysis of traffic safety and efficiency.

4.1 Traffic conflicts model based on VISSIM simulation

A significant linear relationship between conflicts and accident rate has already been established (Su, 1998), and therefore severe conflicts are often used as a proxy to the accident rate. Because of the complexity of actual conflict data collection, in this paper, the number of traffic conflicts is obtained from a model of the studied interchange, simulated in the microscopic traffic simulator VISSIM.

The main factors that affect traffic safety of weaving areas are (i) weaving length L, (ii) total traffic volume Q in the weaving area, (iii) volume ratio VR (ratio of weaving volume to total volume), and (iv) weaving ratio R (ratio of the minimum volume to the maximum weaving volume). Each of these factors has 7 levels and the selected fractional orthogonal design includes 49 groups of experiments along the four dimensions defined above; these experiments are realized using VISSIM. The obtained relationships between simulated conflicts and each of the main four factors are shown in Figures 1 through 4.

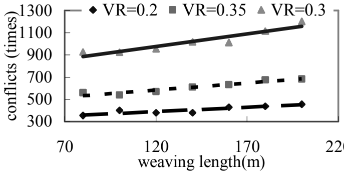


Figure 1. Relationship between conflicts and weaving length

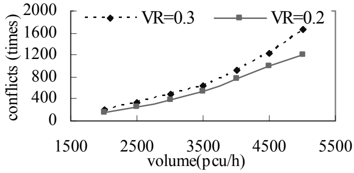


Figure 2. Relationship between conflicts and traffic volume

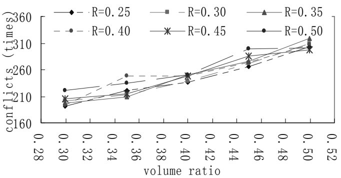


Figure 3. Relationship between conflicts and volume ratio

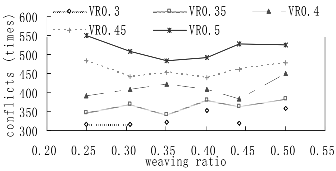


Figure 4. Relationship between conflicts and weaving ratio

Analyzing the figures above, one can deduce that traffic conflicts have a roughly linear relationship with the weaving length *L* and the volume ratio *VR* (see Figures 1 and 3, respectively). The relationship of traffic conflicts with the traffic volume *Q* shows a correlation with a monotonic, slightly concave trend (see Figure 2). Figure 4 suggests that no conclusive relation can be drawn between conflicts and the weaving ratio. These qualitative results are formalized through a multiple regression model (Eq. (1)), estimated in the statistical package SPSS. Based on the information obtained through visual inspection of Figures 1 through 4, the weaving ratio is not considered as an explanatory variable. The resulting model is:

$$TC = 1.134L + 6.99\times10^{-5}Q^2 + 1556.074VR - 704.062 \qquad (1)$$

where *TC* denotes the number of traffic conflicts (times); *L* is the weaving length (m); *Q* is the total volume of the weaving area (pcu/h); and *VR* is the volume ratio (unitless). The fit of the regression is satisfactory, with a coefficient of determination equal to R^2=0.954, while the p values of the joint F-test and individual t-tests are all less than 0.05.

4.2 Other models

The metric of relative conflicts TC_{unit} is defined as the number of traffic conflicts *TC* per unit length, i.e. $TC_{unit}=TC/L$, and reflects the spatial density of traffic conflicts. Increasing values of relative conflicts imply that the probability of an accident occurring is greater, resulting in lower levels of road safety. Conflicts intensity is defined as a measure of the severity of the possible accidents that could result from the conflicts. Arguably, conflicts involving vehicles moving in higher speeds are more likely to result in severe incidents; therefore, the average speed of conflicting vehicles is taken as a proxy to conflict intensity.

The main intent of this paper is to study the relationship between weaving length and safety. In order to limit the degrees of freedom, the total volume,

volume ratio and weaving ratio are given values from a specific interchange of interest (Youyi Ave. Interchange of Wuhan Yangzi River Tunnel), where the total volume is 2005pcu/h, the volume ratio (*VR)* is 0.46 and the weaving ratio (*R)* is 0.24. According to equation (1) and simulation results, the relative conflict model takes the following form $TC_{unit} = \frac{292.76}{L} + 1.134$, where TC_{unit} denotes the relative conflicts (times/m) and L is the weaving length (m). The conflict speed model is given by $\frac{1}{V} = \frac{1}{54} + 0.08 \times 0.972^{L}$, where V is the conflict speed (km/h). According to the Highway Capacity Manual (HCM 2000), $C = 506.83\ln(L) + 716.69$, where C is the capacity (pcu/h).

5 Results and Analysis

5.1 Relationship between Traffic Safety, Efficiency and Weaving Length

In order to simplify the subsequent analysis, the four main indices are normalized to the standard interval, i.e. [0, 1]. The considered range of weaving length is 100m≤L≤200m, in intervals of 10m, i.e. eleven values. Shunying (2005) discussed two normalizing methods: cost type and benefit type. In this research, the conflicts, relative conflicts and conflicts speed are normalized by cost type, while the capacity is normalized by benefit type (see Figure 5).

An intuitive relation between traffic safety and efficiency exists. The probability of accident occurrence increases proportionally to the frequency of conflicts; conversely, the occurrence of an accident leads to capacity reduction.

According to the above analysis, conflicts *TC* are a function of *l, Q and VR*, i.e. *TC~f(l, Q, VR)*, and accidents are related to conflicts; therefore it can be argued that the accident number *N* can be given by *N~ f(l, Q, VR)*. Denoting the average time per accident clearance as *t*, total time lost by accidents *D=N·t~ f(l, Q, VR)*, capacity lost is *D·C~ f(l, Q, VR)*. Defining $\text{efficiency} = \frac{\text{capacity - capacity lost}}{\text{capacity}}$, then traffic efficiency~ $f(l,\ Q,\ VR)$.

It is said that traffic safety and efficiency are all related to both physical indices (length) and operational indices (volume, volume ratio) in the weaving area. The severity of the accidents affects the time it takes to clear them, while the frequency of accidents affects the total time lost due to the incidents, which in turn reflects on the loss of capacity. Clearly, a greater capacity loss corresponds to a lower efficiency.

Figure 5 shows that as the weaving length increases, the normalized values of relative conflicts and capacity increase, while the normalized values of conflicts and conflict speed decrease. Increasing the weaving length can increase the capacity and reduce the relative conflicts; however, the number of conflicts and the conflict speed would increase, and the probability of more frequent and more severe accidents may increase. Increasing weaving lengths correspond to higher construction costs for the interchange.

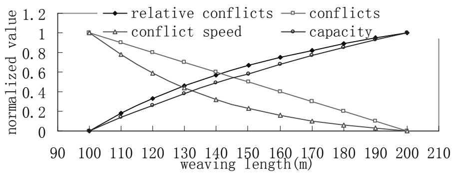


Figure 5. Relationship between weaving length and the normalized values

5.2 Reasonable Weaving Length

In this paper, the normalized values are analyzed using the one-step difference method: the selected range of values is divided in n cases of equal length (each 10m), and the difference between two adjacent cases, $\Delta y_i = y_{i+1} - y_i$, with i=1,2,..., n, is used. It is assumed that the indices change smoothly when $\Delta y_i \leqslant 0.1$, and the improvement of safety and efficiency is small, while engineering cost keeps the original increasing rate, so case i is the best.

In the example of the modeled Wuhan Yangzi River Tunnel, it is found that a weaving length of approximately 140m provides a reasonable balance between safety and efficiency (Figure 5.) For a weaving length of 140m, the normalized values of safety and efficiency change smoothly (with a change rate $\leqslant 0.1$), so it is recognized that safety and efficiency in the weaving area are then balanced. It is therefore indicated that 140m is a reasonable length for this weaving area.

6 Summary and Further Studies

The weaving length is one of the main factors that affect traffic safety in interchanges. The results obtained in this paper show that increasing the weaving length results in an increase in the capacity and the total number of conflicts. It is also found that the measure of relative conflicts decreases, while the expected severity of conflicts (as manifested via the average speed) increases. Furthermore, it is empirically confirmed that a reasonable weaving length for the interchange in the Wuhan Yangzi River Tunnel is 140m.

Lane changing is another form of generalized traffic conflicts. Among other studies, Roess (2006) showed that longer weaving length results in a higher lane changing rate. This finding is consistent with the results in this paper. While this paper has also demonstrated the use of microscopic simulation for the analysis of weaving section characteristics, there are several questions that need to be studied further. For example:

- The effect of the number of lanes, the design speed or the road condition (all of which may affect the weaving significantly) has not been considered.
- The relationship between the occurrence of conflicts in VISSIM and accident rate needs to be further studied, e.g. by a field survey.
- It is necessary to confirm the weighted value in difference analysis.

Acknowledgments

This work was supported by NSFC (NO. 50778142) and the Wuhan Changjiang Tunnel Project Department.

References

Alauddin M. Ahammed, Yasser Hassan and Tarek A. Sayed. (2006). “Effect of Geometry of Entrance Terminals on Freeway Merging Behavior”. TRB 2006 Annual Meeting CD-ROM.

Dr. Roger P. Roess and Mr. Jose M. Ulerio. (2007). “Analysis of four weaving sections: implications for modeling”. TRB 2007Annual Meeting CD-ROM.

Li Xiaojuan. (2006). “Studies on the safety of the High-level Highways Weaving Area”. Master of Engineering Thesis. Fujian Agriculture and Forestry University.

Mohamed Sarhan, Ph.D., Yasser Hassan and Abd El Halim O. Abd El Halim. (2006). “Design of Freeway Speed Change Lanes: Safety-Explicit Approach”. TRB 2006 Annual Meeting CD-ROM.

Ponlathep Lertworawanich, Ph.D. and Lily Elefteriadou, Ph.D. (2007). “Generalized Capacity Estimation Model for Weaving Areas”. Journal of Transportation Engineering. (3): 166-179

Srinivas Jillella. (2001). “A Comparative Study of Weaving Sections in TRANSIMS and Highway Capacity Manual”. Master of Science Thesis. The Faculty of Virginia Polytechnic and State University.

Steven Venglar, P.E. and David Fenno, P.E. (2003). “Weaving Recommendations for Managed Lanes”. TRB 2003 Annual Meeting CD-ROM.

Thomas F. Golob, Wilfred W. Recker, Veronica M. Alvarez. (2003). “Safety Aspects of Freeway Weaving Sections”. TRB 2003 Annual Meeting CD-ROM.

Yihua Zhang and Hesham Rakha. (2005). “Systematic Analysis of Capacity of Weaving Sections”. TRB 2005 Annual Meeting CD-ROM.

Zhang Su. (1998). “China Traffic Conflict Technology”. Chengdu: Southwest Jiaotong University Press.

Zhu Shunying, Wang Hong, Yan Xinping. (2005). “A Kind of F-AHP Method to Macroscopically Evaluate Road Traffic Safety”. Journal of Wuhan University of Technology. 29(5): 697-703

SAFETY BASED SIGNALIZED INTERSECTION LEVEL OF SERVICE

Pan Fuquan, Lu Jian, Xiang Qiaojun, and Deborah S. McAvoy

School of Transportation, Southeast University, 2 Si Pai-lou, Nanjing, Jiangsu, 210096, China, PH(86)25-83790776; email: panfuquan@gmail.com; jianjohnlu@sina.com; xqj@seu.edu.cn; Ohio University, Department of Civil Engineering, Athens, OH 45701 Phone 740-593-1468; email: mcavoy@ohio.edu

Abstract: Intersections are crash-prone locations due to the presence of conflicting vehicular movements. However, safety is not included in the intersection performance measure of level of service (LOS) for signalized intersections according to the current *Highway Capacity Manual*. The research presents the concept of safety level of service (SLOS) and develops model of SLOS for signalized intersections based upon vehicle conflicts, intersection geometrics, signal phasing, pavement markings, signage, pavement condition and ambient lighting. The average control delay of LOS is combined with risk index of SLOS to develop the delay and risk index for the improved LOS performance measure accounted for safety. In light of current LOS criteria and distribution of risk index values, the improved LOS criteria are defined for signalized intersections.
Key words: *level of service; signalized intersections; safety level of service*

1 Introduction

Intersections are important connections between two or more roadways and the resulting conflicting traffic movements make intersections crash-prone locations. Across the USA in 2000, there were more than 2.8 million intersection-related crashes, which accounted for 44 percent of all reported crashes. Similarly, nearly one million injury crashes occurred at or within intersections, constituting 48 percent of all injury reported crashes. Based upon the frequency of crashes at intersections, it seems that traffic safety problems are more common and more severe at intersections than other locations. Unfortunately, the level of service (LOS), which describes the operation of the intersection, is based upon the average control delay of vehicles and does not consider the frequency of crashes or safety of the intersection. While level of service is a theoretical representation for the operation of an intersection, the lack of an intersection safety factor weakens the performance measure and the ability to utilize level of service as an operational predictor. Previous research has shown that intersection capacity can be reduced by as much as 50 percent when a crash occurs at an intersection. Therefore, in order to effectively predict an intersection's operational performance, the service measure should include a safety factor. The research presented in this paper examines how to combine average control delay with a risk index to develop a comprehensive performance measure to evaluate the operational and safety performance at signalized intersections.

2 Safety Level of Service

While the importance of traffic safety at intersections continues to be a growing concern among engineers, the current *Highway Capacity Manual* (HCM) does not include safety in the LOS for signalized intersections. The lack of a safety measure in the LOS does not allow the operations of a signalized intersection to be realistically determined. In order to incorporate safety into the performance measure for intersections, this paper evaluates the concept of a safety level of service (SLOS). The SLOS for intersections incorporates the crash potential of an intersection based upon the intersection geometrics, pavement quality, traffic control, signage and ambient lighting.

2.1 SLOS Factors for Signalized Intersections

SLOS is used to evaluate the safety of an intersection based upon the relationship of the intersection geometrics, traffic control, and traffic crashes. In other works, the SLOS model is based upon factors that influence the intersection safety as well as the intersection operation.

In order to determine which influencing factors should be incorporated into the SLOS model, experts in highway safety and operation were consulted and a field survey was conducted. Using an analytical hierarchy process, factors that were determined to minimally influence intersection safety were removed from consideration in the SLOS model and the remaining influencing factors were classified as either major influencing factors or minor influencing factors. The major influencing factors are conflict points between vehicles include crossing conflict points, merging conflict points or diverging conflict points. The minor influencing factors include traffic control, intersection geometry, signage, pavement markings, pavement condition and ambient lighting.

2.2 Development of SLOS Model for Signalized Intersections

A two-phase signalized intersection has 20 potential conflict points with major conflicts including right angle, left-turn hand-on conflicts which are generally associated with fatalities or serious injury crashes and minor conflicts including merge or rear-end conflicts which are generally associated with less severe crashes. As conflict points are reduced at intersections, the potential for crashes, represented by a lower crash rate, is also reduced (Preston, 2001). In theory, if conflict points do not exist at an intersection, the potential for traffic crashes under normal operating conditions is eliminated. The installation of a traffic signal at an intersection reduces the traffic conflicts to 20 conflict points from 32 conflict points for an unsignalized intersection. The reduction in traffic conflicts is due to the dedicated right-of-way assigned to opposing traffic through the traffic signal, where a minimum two-phase signal is utilized. While the number of conflict points can be decreased through the use of multi-phase traffic signals, the number of signalized intersections is small in comparison to the total number of intersections along the road network. Unfortunately, increasing the number of signal phases also reduces the traffic efficiency and operation at an intersection by increasing the delay experienced by motorists. With increases in delay, the motorists become impatient and may proceed into the intersection at an inappropriate time, such as red light running. These conditions further increase the crash potential at an intersection while the number of conflict points may not be increased.

Therefore, the SLOS model was developed based on the number of conflict points and not on the frequency of traffic crashes or field collected traffic conflict data, as traffic crashes tend to be a rare and random event. However, utilizing traffic conflict points provides a level of comparative certainty and measurability. The calculation of conflict points for signalized intersections used in the SLOS model is not simply the addition of conflict points in each signal phase, but the conflict points are weighted averages calculated according to the split (green interval, yellow interval, plus the all-red interval) in each signal phase.

Additional related data such as the traffic signal (with the exception of phase data), geometrics, signage, pavement markings, pavement condition and ambient lighting have an influence on intersection safety. For example, a green interval that is too short leaves motorists stranded in the middle of the intersection thereby increasing the crash probability of the intersection (Richard A. Retting, 2002). Intersection geometrics such as limited intersection sight distance or an acute crossing angle may also increase the crash potential at an intersection. Signage that is installed properly with consideration of positive guidance play an important role in improving intersection safety. A research study indicated that advanced warning signs could reduce crashes at rural intersections by 30%, and reduce crashes at urban intersections by 40% (Tindale, 2004). Similarly, pavement markings properly installed and maintained can assist motorists navigate through an intersection. Pavement condition also plays a role in intersection safety. It has been found that uneven pavement increased the crashes along the roadway as much as 18 percent. Poor ambient intersection lighting also increases the probability of traffic crashes and their resulting injuries due to the limited visibility conditions.

The SLOS model, based on conflict points and modified with the minor influencing factors, is shown in equation 1.

$$RI = (\sum_k \sum_i NCP_i \times \frac{g_k + y_k}{C} \times SCP_i) \times (\sum_j f_j \times MF_j) \tag{1}$$

Where RI is the risk index of intersection; k is of the number of signal phases; g_k is green interval for the k^{th} phase; y_k is yellow interval for the k^{th} phase; C is cycle length; i is a factor representing the type of conflict point between vehicles, i.e. crossing conflict point, merging conflict point or diverging conflict point; NCP_i is number of conflict points for each i type; SCP_i is severity associated with each i conflict point type, j represents each of minor influencing factors, which include type of traffic control, geometrics, signage, pavement markings, pavement condition and ambient lighting, MF_j is modified factor for each type of j minor influencing factor, f_j a weighting factor for each type of j minor influencing factor which reflects the actions of each type of minor influencing factor.

In the SLOS model, the number of conflict points for each conflict type and modified factor for each minor influencing factor must be determined. The former can be calculated based upon the number of intersection legs and the number of lanes for each leg, and the signal phasing, etc. The latter is calculated by a panel of expert traffic engineers who would determine a score for each minor influencing factor based upon the field conditions using pre-designed judgment criteria. Due to the limitation of the paper length, the details for the calculation of the modified factors have not been provided. Constants that need to be determined are the severity for each type of conflict point and the weighting factor for each minor influencing factor. According to theoretical analyses of conflict points and intersection crash data, crossing conflict point crashes result in angle crashes; merging conflict point crashes result in rear-end or property damage only crashes; and diverging conflict point crashes result in rear-end crashes. So the severity of each type of conflict point can be determined through the economic loss of crashes associated with each type of conflict point. According to crash statistics from 1,700 highway intersections in Wisconsin for the years between 2001 and 2003 (Knapp, 2005), and the research on the economic loss based upon crash type (Council, 2005), the severity for each conflict point, crossing, merging, and diverging, are calculated and shown in Table 1.

Table 1 Severity based upon type of conflict point

Conflict point type	Severity
Crossing conflict point	1.0
Merging conflict point	1.5
Diverging conflict point	3.0

Due to the numerous minor influencing factors and their effect on the SLOS model, the weights f_j are determined by a panel of expert traffic engineers through a field survey. For the SLOS model, a large sample survey was conducted to determine the weights f_j for various minor influencing factors. The weights for the minor influencing factors as determined through the survey are 0.29 for traffic control, 0.17 for geometrics, 0.15 for signage, 0.16 for pavement markings, 0.12 for pavement condition and 0.11 for ambient lighting.

A validation was performed on the risk index for an intersection and a reasonable relationship with the subjective average rating score for the crash potential was found. Unfortunately, due to the limitations of the paper, the process of validation has not been detailed.

3 Current service measure for signalized intersections

The conventional performance measure for signalized intersections is average control delay as detailed in the HCM2000 (TRB, 2001). Average control delay for each movement for each approach is weighted by the number of vehicles for that movement to determine the average control delay per vehicle for the overall intersection. The LOS criteria based upon average control delay for signalized intersections as presented in the HCM 2000 are listed in Table 2.

Table 2 HCM LOS criteria for signalized intersections

LOS	Average control delay (s/veh)
A	≤10
B	>10-20
C	>20-35
D	>35-55
E	>55-80
F	>80

4 Service measure of LOS accounted for safety

Because the LOS performance measure for signalized intersections, as detailed in the HCM 2000, does not include a safety factor, the capacity analysis results do not thoroughly reflect the operational performance of signalized intersections. As discussed earlier, the SLOS model was developed in order to introduce a safety factor to determine the ultimate performance measure for the signalized intersection which will include delay as well as crash potential. Therefore, the HCM's performance measure of LOS has been combined with the SLOS performance measure (*RI*) to develop a new comprehensive performance measure, which will thoroughly represent the operational and safety performance of signalized intersections. The procedure for determining the delay and risk index involves combining the average control delay for an intersection with the risk index. The model is shown below in equation 2.

$$DRI = D + RI \tag{2}$$

Where, *DRI* is delay and risk index for signalized intersections; *D* is average control delay for signalized intersections; *RI* is risk index for signalized intersections.

The delay and risk index for signalized intersections not only evaluates the safety and operational status of the intersection, but the *DRI* can also be used to design signalized intersections. With an emphasis on safety, a signalized intersection may be designed with minimal conflict points, proper signage, the provision of retroreflective pavement markings, and adequate ambient lighting. In order to optimize traffic efficiency in addition to minimizing conflicts, a signalized intersection may be designed with minimal phases in order to reduce vehicular delay.

5 Improved LOS Criteria for Signalized Intersections

According to original LOS criteria and distribution of *RI* values of SLOS, the improved LOS criteria which accounts for intersection safety are defined in Table 3.

Table 3 Improved LOS criteria for signalized intersections

New LOS	Delay and risk index
A	≤20
B	>20-35
C	>35-55
D	>55-80
E	>80-115
F	>115

6 Conclusion and further study

Based upon the existing performance measure criteria for signalized intersections, LOS, which does not include a safety factor, the concept of SLOS was presented and the SLOS model was developed. In order to objectively evaluate the operation of a signalized intersection, the existing performance measure of LOS was combined with risk index for the intersection to develop the delay and risk index, the improved LOS performance measure. The improved LOS criteria were based upon the current LOS criteria and distribution of *RI* values of SLOS. While this research has developed a method for combining the operational and safety performance of an intersection, there are several aspects require further investigation. Therefore, further research should be performed in the following areas: a sensitivity analysis of delay should be conducted, the risk index should be evaluated based upon a variety of minor influencing conditions and the improved LOS criteria should be tested with numerous intersections.

7 Acknowledgments

This research was supported by National Natural Science Foundation of China (50522207) and China Department of Transportation (200431822333).

References

Council, F., et al. (2005). *Crash Cost Estimates by Maximum Police-Reported Injury Severity Within Selected Crash Geometries*. FHWA-HRT-05-051. Federal Highway Administration, Washington, D.C.

Knapp, K. and Campbell, J. (2005). *Intersection Crash Summary Statistics for Wisconsin*. Wisconsin Department of Transportation, Madison, pp.28-38.

Preston, H. and Albrecht, C. (2001). *Traffic Safety Fundamentals Handbook*. Minnesota Department of Transportation -Office of Traffic Engineering, St. Paul. pp. 40-42.

Retting, R., et al. (2002). "Changes in crash risk following re-timing of traffic signal change intervals." *Accident Analysis and Prevention*, Vol. 34, pp. 215–220.

Tindale, S. A., Miller, D. (2004). *Toolbox on Intersection Safety and Design*. Institute of Transportation Engineers, Washington, D.C., pp. 91-128.

TRB. (2001). *Highway Capacity Manual 2000*. TRB, National Research Council, Washington, D.C.

Safety Analysis of Freeway Interchanges

Yulong He
Associate Professor
Transportation Research Center,Beijing University of Technology, Beijing, China
Tel: (8610) 6739-6176 Fax: (8610) 6739-1509
ylhe@bjut.edu.cn

Xiaoduan Sun
Professor
Transportation Research Center,Beijing University of Technology, Beijing, China
Tel: (8610) 6739-1597 Fax: (8610) 6739-1509 xsun@louisiana.edu

Richard C. Coakley, P.E., PTOE
CH2M Hill,135 South 84th Street,Suite 325,Milwaukee, WI 53214,(414) 847-0423
WisDOT (262) 548-6438;Cell (414) 807-7431; Richard.Coakley@CH2M.com

ABSTACT

While investigating Jingjintang Expressway's traffic accidents, it is clear that, crash frequency and rates at interchange areas are more severe than that at other locations. To further explore the relationship between crash rates and characteristics of freeway geometric design features, this study analyzed the existing safety problems at Jingjintang Expressway's interchanges. Particularly we examined crash characteristics at ramps including length of acceleration lanes and deceleration lanes, ramp geometric design, traffic control devices at ramps and locations of toll plazas. The driver behavior and performance of overloaded trucks on ramps were also studied. The results of the study have clearly shown that it is important to have a uniform traffic information system and a consistent traffic control system. Most importantly, the study has concluded that the length requirements in our current design specification for ramp acceleration and decoration lanes are not sufficient for safety, particularly for large vehicles. A set of recommendations are made as a result of this study.

BACKGROUND

Jingjintang Expressway is the connection of Beijing and Tianjin, which is the link between the capital of China and the capital of hebei province. This 142.6 kilometers of freeway is situated in a wide plain with very large curves. It was one of the first freeways built in China in 1989, its abnormally high crash frequency and rates generate much notoriety from the local and national news media. Because of its important geographic location, reducing the crash frequency and severity has become a top priority in the regional sustainable economic development plan and an urgent task for all related agencies.

While investigating Jingjintang Expressway's traffic accidents that occurred from 2002 to 2004, it is clear that the crash frequency and rates at interchange areas are more severe than that at other locations. To further explore the relationship between crashes and characteristics of interchange areas, this paper analyzed the existing safety problems at interchanges areas of Jingjintang Expressway. The main purpose of the study is to identify the root of the problems to enable better and workable safety enhancement strategies.

DATA COLLECTION

Considering a lot of possible factors influencing the occurrence of crashes at interchange areas, three categories of information were included in the paper: expressway and interchange area geometric data, traffic flow and crash data, and traffic control devices.

Geometric Design

Jingjintang Expressway was designed in the early 1980's before China developed its own freeway design criteria. It's necessary to collect revenue from Jingjintang Expressway, most of the interchanges are partial cloverleaf or trumpet configurations. Due to a rigid restriction on right-of-way acquisition，The minimum ramp radius on horizontal curves is 30 meters. The length requirement for acceleration and deceleration lanes on entrance and exit ramps closely follows the Japanese design standard that is shorter than that specified by the AASHTO design policy.

Traffic Flow

Traffic flow data, including speed, vehicle count and classification, were collected from 23 loop detector systems installed along the expressway, which clearly reveals the traffic conditions at five minute intervals. Limited by the detection system, traffic flow data is divided into two kinds, one for cars and another for trucks.

Crash Data

A total of 2829 crashes occurred during 2002-2004, including 144 with fatalities, 1435 with injuries, and 1250 with property damage only.

In addition to gathering the above data, a driver survey was conducted at three rest areas. A total of 218 truck drivers participated in the survey. The main purpose of the survey was to investigate drivers behavior on expressways, i.e., how they understand and obey freeway laws and regulations. A few questions were designed to see how drivers perceive the freeway design criteria. For accurate and reliable results, the survey was carried out by personal interviews.

Traffic control devices data

In order to analyze the factors concerning traffic accidents, the data on traffic signing and marking along the expressway were collected. Signing and Marking follows the standard for all traffic control devices in china. It was published in 1999. It doesn't give detailed requirements on driver information needs such as frequency of reinforcing the message to drivers, i.e. the interval of repeating the same sign.

CRASH ANALYSIS

Table 1 accidents number of interchange areas and Jingjintang expressway

	The number of crashes	Critical crashes	Minor crashes	The number of fatalities	The number of injured	The length of two way （kilometer）
Interchange area	176	8	168	10	45	16.192
Jingjintang expressway	2829	144	2685	190	782	285.38
Percentage	6.22%	5.56%	6.26%	5.26%	5.75%	5.67%

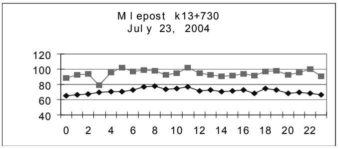
The mileage of interchange areas accounts for 5.67% of Jingjintang expressway, as shown in Table 1. The crash rate of interchange areas is higher than other sections according to the accidents from 2000 to 2003. The number of crashes was 176 in the past three years, it accounts for 6.22% of the total number of crashes. 6.26% of minor crashes and 5.56% of critical crashes occurred within interchange areas. The Percentage of fatalities within interchange areas accounts for 5.26%. Comparing the mileage and crash statistics above, the frequency of crashes is higher than other locations, however, the severity of crashes is similar to other sections of jingjintang expressway.

According to the list of crash rates according to Beijing mile post, it is based on the crash data from 2002～2004. It is obvious that the Majuqiao interchange area is over twice as high as other non–interchange sections. It is necessary to explore the key causes on Majuqiao interchange area. Majuqiao interchange area extended from milepost 12.308 to 13.442. It is a half-cloverleaf interchange. There were 109 crashes that occurred along 1.134 kilometers of the Majuqiao interchange area. The most common type of crashes were rear-end collisions, which accounted for 50.46% of total crashes. Among the rear-end collisions, 44% occurred between car and truck, 16% occurred between trucks, and 35% occurred betweens cars. Roadside crashes accounted for 17.43% of total crashes, second to rear-end type collisions. 68% of roadside crashes were caused by illegal lane changes. Illegal parking and backing within the interchange area contributed 5% of roadside crashes. Illegal operations were involved in up to 21% of roadside crashes. It was a common situation on Jingjintang expressway. The third most common type of crash was sideswipe, which amounted to 11.93% of total crashes.12% of the crashes within the Majuqiao interchange area were caused by illegal backing maneuvers.

DISCUSSION

Highway safety deals with a complex system of three major components: human, vehicle and roadway. When these three elements do not interact in accordance with the environment, crashes occur. Based on comprehensive analysis, it is clear that the combination of poor driver behavior, poor vehicle conditions, overloaded trucks (the main causes of vehicle’s mechanical problem) and somewhat substandard geometric and traffic control design features are the root of the problem as discussed below.

Differential Speed



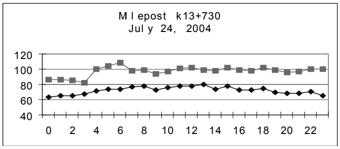


FIGURE 1 **Examples of speed differentials by hour of the day**

There is a large gap between the mechanical performance of cars and trucks in china. Overloaded trucks are currently a widespread problem in China, this results in the high frequency of rear-end collisions particularly for the unusually high number of such fatal collisions. As depicted in FIGURE 1, the average speed of trucks is barely above 80 kilometers per hour, while the speed limit is 110 km/hr in this section of the expressway. The problem with different speeds between cars and overloaded trucks is worsened for freeways on a slope. The overloaded trucks traveling on the ramp upgrades are the cause of many fatal rear-end collisions between slow climbing trucks and fast moving cars, especially at night when visibility is limited and rear lights of truck are covered by dirt.

Interchange Spacing

In order to allow for the maintenance crews and U-turn for the wrong-way drivers, it is necessary to limit the Maximum space and Minimum space of interchange. Because it is a toll road, the drivers have to pay an extra fee for their wrong way movement. Most important is the uncertainty of whether they can make a U-turn that will place them where they need to be for the correct exit.

Although Jingjintang expressway is a high standard highway facility, some drivers do not follow the traffic control devices, which results in crashes. For instance, 12% of surveyed drivers said that they would go back to a missed exit if they suddenly realize that they have just missed the intended exit. The actual number is likely higher than 12% considering that the research team witnessed several such maneuvers during a few very brief on-site data collection periods. Two types of maneuvers were observed that drivers used to go back to a missed exit: backing up or turning around to drive against the proper direction of travel. This travel in the wrong direction was the main cause of 14 head-on collisions that occurred in the last three years.

Traffic Signs

Being an important link connecting Beijing to Tianjin, the expressway carries 40,000 vehicles per day. Advertisement companies set a lot of big and fancy billboards along the expressway. These high and huge billboards distract the driver's attention from traffic signs. Recognizing and paying attention to the signs in a timely and accurate manner presents a challenge to some drivers.

"Road traffic signs and marking" GB5768-1999 have no specific criteria about the interval of traffic signs. For example, Within the Majuqiao interchange area, there were 8 traffic signs set along 1.134 kilometers. The average space is 167 meter between traffic signs. That means drivers have to deal with a traffic sign every five seconds at 110 kilometer/hour. Two of the signs are entrance signs, five of them are guide signs for exits, one of them is service area sign. These seven guide signs have text messages. It takes more time than the graphic symbol signs. Drivers need to make the decision whether it is their intended exit in five seconds.
12% of surveyed truck drivers didn't exactly understand the meanings of some traffic signs. 3% of them didn't understand the meaning of all traffic signs. 5% of them said they ignore the speed limit sign. 3% of them would keep on speeding unless they noticed a speed camera or police. Only 76% of them responded that they obey the speed limits.

length of acceleration and deceleration lanes

Comparing the standard length of deceleration and acceleration lanes in the USA with those on the Jingjintang expressway, the lengths of acceleration lanes are less than the USA standard length. 38% of surveyed truck drivers thought the length of the acceleration lanes on Jingjintang expressway is not enough. 58% of them thought the lengths were adequate. Because the acceleration performance of most trucks in china is well below that of passenger cars, longer acceleration lanes are needed to get up to speed.
32% of surveyed truck drivers thought the length of the deceleration lanes are not enough. 67% of them thought the lengths were adequate. For example, the length of deceleration of Majuqiao interchange is 270 meters, the length of taper is 70 meters, the deceleration is 14.26 m/s^2, this deceleration rate is uncomfortable for the driver and passenger. The suggested deceleration is $2m/s^2$~$3m/s^2$ in china.

SUMMARY

Crash frequency and rates at interchange areas are more severe than that at other locations. To improve Jingjintang expressway safety, it is critical to improve the safety of interchange areas. Among the problems above that affect traffic safety, some can be tackled from a highway engineering aspect alone, such as eliminating billboards along the expressway, locating traffic signs to meet drivers' expectancies and limit the number and spacing of traffic signs according to the ability of drivers within interchange areas, increase the length of acceleration and deceleration lanes on the expressway if it is possible.

The effort to reduce the unusually high number of fatal rear-end crashes must start by eliminating slow moving trucks from the freeways. Strictly enforcing the regulations on overloaded trucks is absolutely critical to the safety of expressway and interchange operations.

Driver education is another urgent task in combating traffic crashes. Safe driver practices result in safe highways. Changing driver behavior can be achieved by many different methods. Persistent and uniform enforcement coupled with commercial driver training has proven to be effective in many developed countries for enhancing large vehicle safety.

ACKNOWLEDGEMENT

The authors wish to express their gratitude to Huabei expressway Company Limited, who shared all geometric, traffic control design and traffic volume data. Appreciation also goes to all the police officers that helped us in collecting all traffic crash information and all graduate students in our research group at Beijing University of Technology for their hard work.

REFERENCES

[1] China Ministry of Communication, "National Freeway Development Plan", 2005

[2] China Ministry of Security, "Annual Traffic Crash Statistics"

Enterprise Architecture Planning for Public Transit

Mimi Hwang, PMP, CISSP[a] and Eva Lerner-Lam, M.ASCE, F.ITE[b]

[a]President, KAO eTechnologies, Inc., Orlando, Florida, USA, mhwang@kaoetech.net
[b]President, Palisades Consulting Group, Inc., Tenafly, New Jersey, USA, elernerlam@palisadesgroup.com

Abstract

As public transport systems continue to modernize, China's implementation of Information Technology and Intelligent Transportation Systems will require a focus on increasingly complex data and technology issues. The success of these systems and the management of the underlying data are crucial to the support of China's current and future public transport services.

This paper describes a business-technology planning approach used by leading-edge transit agencies in the United States (US) called "Enterprise Architecture Planning" or "EAP." EAP enables open-architecture, standards-based technology planning and deployment. By using an EAP approach, public transport agencies in China can define, document and coordinate the information technology requirements and implementation plans for mission-oriented business needs. By so doing, they can support local and national collaboration for increased efficiency and improved services to society.

Keywords: *Enterprise Architecture Planning, Business Architecture, Information Technology, Intelligent Transportation Systems, Advanced Public Transportation Systems, National ITS Architecture, FTA National ITS Architecture Policy*

1 Background

Rapid information technology advancements have dramatically altered the options available to transit agencies to develop and deliver more effective and efficient services. Information Technology (IT) and Intelligent Transportation Systems (ITS) are becoming increasingly important to the transit industry. They help to achieve strategic and tactical business goals, and affect virtually every business process. Technology advances in wireless, data, Internet, geographic information systems (GIS), electronic commerce, and ITS "on-board" and "back office" technologies have all been significant in the last five years, providing China with many new opportunities for service-level improvements based on those advances in technology. ITS systems, such as Automated Vehicle Location (AVL) and Automatic Passenger Counting (APC) systems, are being used to improve transit service delivery, safety, reliability, customer convenience, efficiency, responsiveness and regional service integration.

In the US, national Advanced Public Transportation Systems (APTS) policies have emphasized integration, regional data sharing, and the use of common standards to increase system reliability and reduce the costs of implementation.[1] Consistency with the policies is a requirement of federal grant funding. The US Federal Transit Administration (FTA) monitors conformance with the ITS Architecture policy and non-conformance can impact the timing and flow of a project's grant funding. Figure 1 shows an overview of the major components of the US National ITS Architecture. This high-level framework defines the Travelers, Centers, Vehicles and roadside Infrastructure and the information flows that must be accommodated. The National ITS Architecture also identifies participants, describes functions and activities, describes data flow options, provides a framework for identifying, planning and integrating ITS systems.[2]

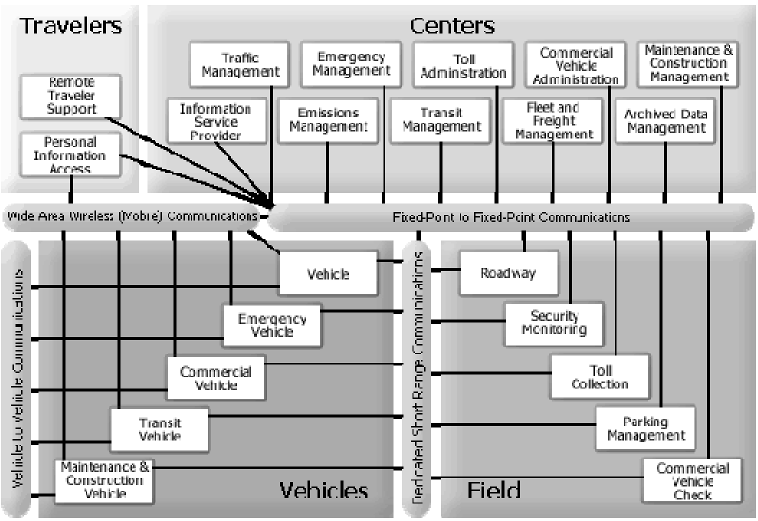


Figure 1
US National ITS Architecture
(www.iteris.com/itsarch/)

---

[1] FTA National ITS Architecture Policy on Transit Projects policy: www.its.dot.gov/aconform/aconform.htm.

[2] The US National ITS Architecture, outreach, deployments and other technical and programmatic information on the National ITS Architecture can be found at www.iteris.com/itsarch/.

2 EAP Methodology

EAP has been endorsed by the US Federal Chief Information Officers' (CIO) Council to promote shared development, interoperability, and sharing of information for Federal agencies and other government entities.[3] EAP development methodology supports US FTA Policy and Regional ITS Architecture goals for integration, regional coordination and cooperative deployments, and cost-effective technology investment. To support federal and regional policies, internal transit agency environments must also embrace these goals.

U.S. FEA Reference Model	**Description**
Performance (Vision/Mission)	Common set of general performance outputs and measures for agencies to use to achieve business goals and objectives
Business	Describe hierarchy of transit operations independent of the departments that perform them. Include the definition of services provided.
Service (Application)	Define IT application (service) that support transit operations and promotes the reuse of components across departments/agencies.
Data and Information	Describe at aggregate level the types of data and information that support program and business line operations and hierarchical relationships among them.
Technology	Describes how technology is supporting the delivery of service components including relevant implementation standards

Table 1 US Federal Enterprise Architecture Reference Model (www.cio.gov)

EAP is the process of defining architectures for the use of information in support of the business and a high-level plan for implementing those architectures. Essential transit business activities are linked to the information needed to support defined agency goals, critical success factors, and performance measures. An IT/ITS strategic planning process based on EAP guidelines will define the detailed systems architectures. It is important to note that EAP does not design systems, databases or networks. The design and implementation work is initiated after the EAP definition process has been completed. An example of an IT/ITS Strategic Planning approach based on EAP guidelines and combined with a structured IT methodology is illustrated in Figure 2.

---

3 See www.cio.gov for architectural framework documents by the US Federal Chief Information Officers' Council.

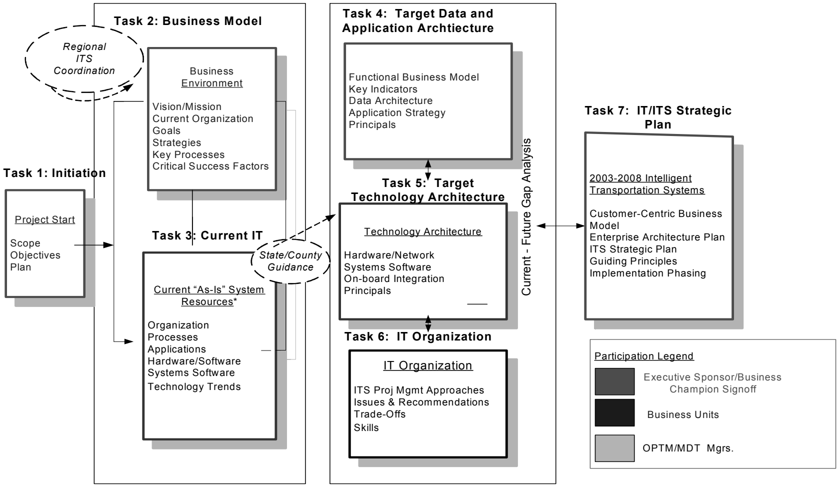


Figure 2 EAP-Based, Miami-Dade Transit IT/ITS Strategic Planning Process

3 Business and Technology Alignment: EAP Framework

Key to the successful deployment of technology is ensuring that the agency's business mission and goals drive the information and technology investments. As Figure illustrates, the EAP Framework for IT/ITS strategic planning identifies the relationship of an agency's goals to its business processes, decision-making, information resources and information technology as its Business, or "Enterprise" Architecture.[4]

The framework begins with the agency's mission and goals. Policies required to meet the mission and goals are adopted by the agency's board. Primary processes are defined to support the policies. Technology design and implementation follow the determination of primary processes. Executive and senior management engagement in the EAP process is essential, as is the involvement of external stakeholders whose interactions with the agency depend on information transmitted between and among them.

4 Transit Business Architecture: Primary Business Processes to Support Transit Mission, Goals and Policies

Figure 3 illustrates six "primary processes" that were identified as part of an EAP effort conducted in 2003 by Miami-Dade Transit. Transit technology was to be

[4] Adapted from *Miami-Dade Transit Information Technology and Intelligent Transportation Systems 2003-2008 Strategic Plan Summary, Volume I*, January 19, 2004, Figure 1-3. p. 1-17.

planned and specified according to these processes in order to help transit meet its *mission, goals and policies*. The primary processes were Service Implementation, Service Management, Customer Information, Safety and Security, Asset Management and Internal Infrastructure.[5]

5 Transit Enterprise Architecture: Data, Application and Technology

Based on transit business processes and goals that are clarified during the EAP-based strategic planning process, a "Target" Enterprise Architecture (EA) can be defined for transit data, applications and technology. The Target Enterprise Architecture provides the initial concept of information systems, including ITS, and the linkages between them. It is made up of key information systems, systems boundaries, enterprise data hubs, data flows, data standards that integrate those information systems and the surrounding hardware/software and networks.

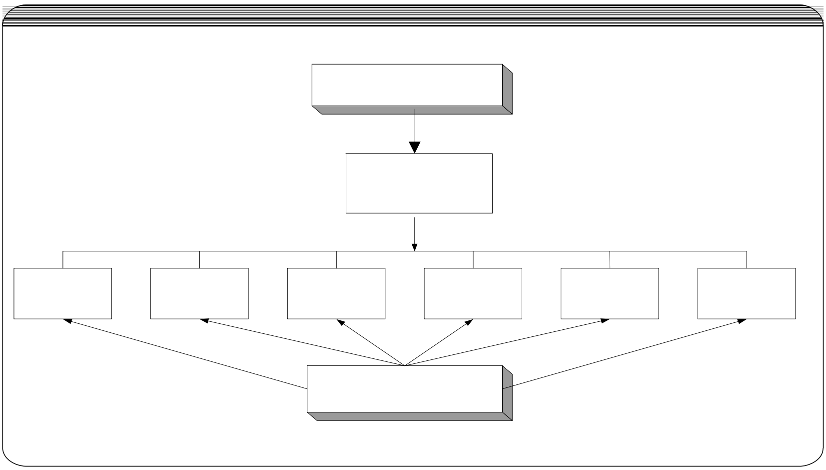


Figure 3 Transit EAP Framework and Strategic Processes

A conceptual view of a Target EA is depicted in Figure 4. Project initiatives can be driven from this framework, and data and application initiatives can be supported and contained within the technology projects. The data and application layers illustrated show linkages to the corresponding Regional ITS Architecture that were developed in response to the US FTA ITS Policy and is based on User Services and Market Packages as defined by the National ITS Architecture.[6]

---

5 Ibid.

6 *Miami-Dade Transit Information Technology and Intelligent Transportation Systems 2003-2008 Strategic Plan, Volume I*, January 19, 2004, page 5-41.

6 ITS Considerations

In the US, transit agencies are applying strong linkages between IT and ITS projects and their operating environment. IT and ITS should be considered together because they both:

- are driven by the business goals and other aspects of "Enterprise" Architecture
- use and impact the Technology Architecture and infrastructure
- need to be integrated together to receive and send needed data
- drive the design and use of the Data and Application Architecture
- should be planned, designed, implemented and operated using the same Enterprise Technology goals, standards and quality assurance procedures
- have users that would like to learn and use the same graphical user interface, data query, analysis and reporting tools, whether they are querying a traditional IT dataset or an ITS derived real-time dataset.

Through the application of Enterprise Architecture Planning guidelines in IT/ITS strategic planning, transit agencies and customers will benefit in a variety of ways.

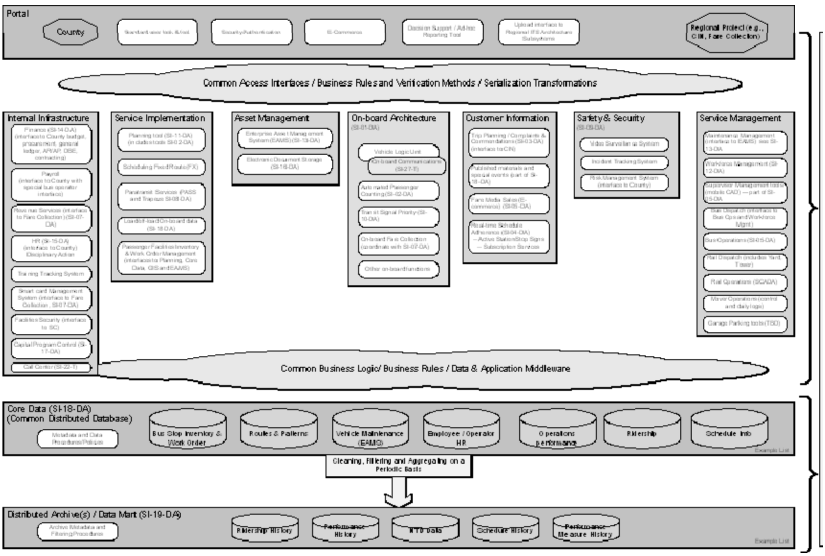


Figure 4 A Transit Enterprise Architecture

References

Federal Chief Information Officers Council. www.cio.gov.

Federal Transit Administration, US Department of Transportation. National ITS Architecture. www.iteris.com/itsarch/

Federal Transit Administration, US Department of Transportation. Conformity Directive. www.its.dot.gov/aconform/aconform.htm

Gonzales, Paul. *Building Quality Intelligent Transportation Systems Through Systems Engineering*. Report No. FHWA-OP-02-046. Department of Transportation Intelligent Transportation Systems Joint Program Office. http://www.itsdocs.fhwa.dot.gov/JPODOCS/REPTS_TE/13620.html

Lerner-Lam, Eva, E. Heimall, M. Hwang, B. Kronenberger, R. McGoff, N. Neuerburg, P. Okunieff and A, White. Miami-Dade Transit Information Technology and ITS 2003-2008 Strategic Plan, Volume I. January 19, 2004.

Lerner-Lam, Eva, M. Hwang, J. Kemp, J. Schiavone. *National Capital Region Workshop on Transit Compliance with the National ITS Architecture—Summary of Proceedings*. Federal Transit Administration, and Washington Metropolitan Area Transit Authority, January 30, 2004.

Spewak, Steven and S. Hill. Enterprise Architecture Planning: Developing a Blueprint for Data, Applications, and Technology. January 1995.

Application of Data Warehouse in Intelligent Public Transportation System

Qu Dayi[1,3], James P Hall[2] , Zhang Xiaojing[1], Wang Dianhai[3]

(1. Qingdao Technological University, Qingdao266033；2.University of Illinois at Springfield；3. Jilin University, Changchun130025)

Abstract: This paper presents the application of data mining technologies for GPS positioning data uploaded by on-board intelligent public transit terminals to a data warehouse. The results provide a reliable reference for the application of increased intelligence to public transportation industry operation and management. This helps management to organize and guide public transportation operations, thus enabling the city public transportation to complete its operational plan and efficiently manage its daily operations by promoting more effective decision analysis.

Keywords：public transport; GPS positioning; data warehouse; cluster analysis

The purpose of the Advanced Public Transportation System (APTS) is to enhance the reliability, efficiency and service of the public transit system by reducing travel times and costs and by providing security for the pilots and passengers. The result is more reliable and convenient public transportation. The system makes use of bus-carried terminals with an advanced global positioning system (GPS), which features high accuracy data collection and easy driver operation. The GPS location data is transferred to the control center by the terminal using General Packet Radio Service/Code Division Multiple Access (GPRS/CDMA) wireless network protocols. The dispatcher of the control center can monitor, organize and dispatch vehicles on a GIS platform. The GPS data are transferred to the control center including arrival and departure mark, station name and elapsed time. Although the GPS data volumes are large, it is organized through a data warehouse that enables analysis using data mining technologies. The study of the periodic variations of the data provide effective data support for public transport operational decision making.

1 Establishment of APTS Data Warehouse

Organization of the data in the APTS data warehouse enables the efficient analysis of the large data resource. The APTS data warehouse is a specialized application database integrating the APTS data pool. It enables data integration and organization of the APTS correlation domain by subject. APTS data management provides information, as required by the application, to support policy-making analysis for public transport system administration. The APTS data warehouse contains such information as enterprises, lines, vehicles, personnel and stops. In addition, the data warehouse includes transit vehicle maintenance history, the enterprise revenue, and vehicle and personnel business operations data (e.g. operating time and distance).

In this paper, we use the data warehouse to analyze the inter station running times and arrival station time-gaps using arrival and departure GPS data,. There are many types of relational database structures such as ORACLE, DB2, and MS

SERVER. Considering the universality and performance, we selected the ORACLE database software to store the raw data and we used ORACLE-integrated tools to perform the data analysis. To create the data warehouse, we first extracted the arrival and departure GPS data and filtered out the irrelevant data (such as longitude, latitude and speed). In addition, since the raw data component was still very large, we created a partition using the field "time" and then established a multi-dimensional model. As shown in Figure 1, using the Erwin logical diagram tool, we identified method IDEFIX (Integration Definition for Information Modeling). The Yellow symbol in the middle represents the primary analysis table with the other symbols representing other relevant dimension tables. We used the multi-dimensional tool, available in ORACLE, to establish a cube for the purposes of (1) identifying subjects (since the data scope is wide, the cube enables representation of one subject; and (2) analyzing data from many aspects. In addition, we established a transformation view for this cube which is updated on a regular basis. The transformation view inquiry data includes: line name, vehicle number (or vehicle trademark), stop name, and arrival and departure mark and time.

2 Data Analysis

Data mining is a powerful tool to analyze large and disperse datasets to provide useful information. The knowledge is expressed in forms of Concept, Rule, Regularity, Patterns and so on. Typical data mining technologies include: correlation analysis, classification analysis and cluster analysis, each of which depends on effective algorithms.

Fishier (in 1958) developed a cluster algorithm of samples, which can obtain the optimal solution, definition diameter $D(i,j)$ representing the class $\{i, i+1, \ldots, j\}$ diameter. The class diameter $D(i,j)$ may have the many kinds of expression methods, the commonly used method which expresses the diameter size with this kind of value and the Sum of squares of average value difference. $b_{n,k}$ means dividing n samples into k.

Where,

$$b_{n,k} : \{i_1 = 1, i_1 + 1, \ldots, i_1 - 1\}, \{i_2 = 1, i_2 + 1, \ldots, i_2 - 1\}, \ldots, \{i_k = 1, i_k + 1, \ldots, n\}$$,

$i_1 = 1 < i_1 \ldots < i_k < n$. The loss function under the $b_{n,k}$ classification is defined as:

$$L(b_{n,k}) = \sum_{j=1}^{k} D(i_j, i_{j+1} - 1)$$

In the equation $i_{k+1} = n+1$, the smaller the loss function value is, the more reasonable the classification. In practical applications, we should determine the precision of the loss function, and deal with the flaw use value and any unusual values in calculation. Only in this way can we obtain a reasonable result. In this article, the writer performed this analysis using inter station running time and the arrival station time-gap as subject, and applying the Fishier algorithm as necessary. This algorithm enables the analysis of time series data and graphical description of the rule.

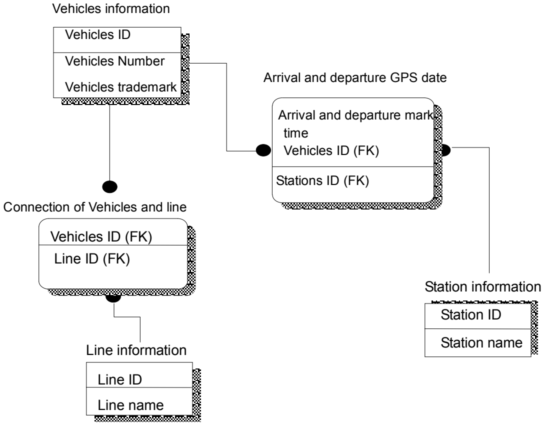


Figure 1 Multi-dimensioned Chart

2.1 Inter stations running time analysis

The inter station running time analysis goal is to determine the running time between any two stations, e.g. the vehicles which arrive at station B station from station A. To calculate this running time we need to subtract the departure time for station A from the arrival time at station B. The parameters include line, arrival and departure, date, time and area. In addition, in order to prevent oversized computational quantities due to data flaws, we also determined the effective running time scope. As an example, we chose city public transportation bus No. 228 to demonstrate this analysis, with separate analyses for Friday and Saturday.

Figure 2 displays a portion of the raw data of bus No. 228's running time on Friday. The x-coordinate represents the transport business time interval, according to the chosen time interval length and the time interval unit (Note: we used 15 minutes as the fundamental unit with multiple increases, such as 30 minutes, 45 minutes and so on). The y-coordinate represents the actual travel time between two stations, which is the average data of the time interval, and in the expression it was used as the data of next time interval. For example, in Figure 2, the average running time of the 6:00-6:15 interval is 35 minutes, so the x-coordinate of the 6:15 point has the y-coordinate data of 35 minutes. As shown in Figure 2, during 6:00-7:15 the change is slower, with running time approximating 36 minutes, until 7:30 when the running time increases.

The change characteristics of the data can be roughly seen from Fig. 2, but we cannot judge the specific scope and the average value in it, therefore it is difficult to determine variation patterns. Figure 3 displays the results of the clustering analysis of the data. The clustering for transportation is generally divided into 7 classes (morning low traffic, morning high traffic and morning flat traffic etc)., However, when there are unusual data, for example in certain intervals when the data is very large, we can suitably increase the classification categories. This is shown in the Fig. 3 that during 6:00-7:15 time period, the average operating time is 36 minutes, increasing to 40 minutes from 7:30-8:30, and then falling to 37 minutes.

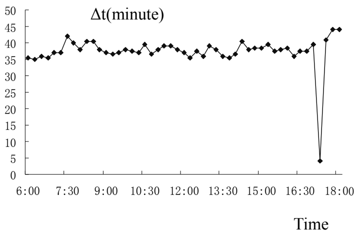


Fig. 2 Raw data of operating time inter stations On Friday

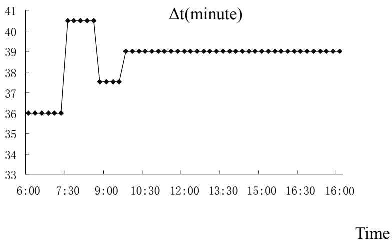


Fig. 3 Cluster results of operating time inter stations On Friday

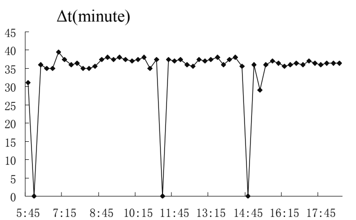


Fig. 4 Primary data of inter stations running time on Saturday

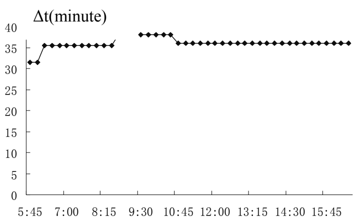


Fig. 5 Cluster results of inter stations running time on Saturday

The method for the operating time analysis on Saturday is identical to that of the previously presented analysis of the Friday data for the No. 228 bus inter stations. The original calculation data for Saturday is displayed in Fig. 4. The "0" point represents lost data in the middle of the timeframe. Please note that the 7:30-8:30 time period of the Saturday data is very similar with the surrounding data and it is difficult to determine peak hours. However, between 9:00-10:00, there is a clear trend moving upward in operating time.

Figure 5 represents the results of the clustering analysis of the Saturday data. Since some raw data was lost, directly using the gathered analysis method is not as appropriate. By using the average value between two time scopes, the interval time can be approximated. As shown in Figure 5, the peak hour has shifted, 7:15~8:30 approximately 35 minutes, 8:45~10:30 up to 37 minutes, which are different from the data distribution on Friday.

As shown in the data above, the peak hour of Saturday is moving backwards when compared with Friday. If we perform an analysis on running time of inter stations, we can check for illegal vehicles and identify late vehicles. We can list all vehicles running time by analyzing raw data and analyzing the peak and low time intervals. The peak-low time interval division has a lot to do with the cycle time along with passenger flow. If the running time is long, it may represent periods of severe traffic congestion. We may also analyze the change resulting from holiday passenger flow.

2.2 Arrival time-gap analysis

The goal of the arrival time-gap analysis is to determine how soon the vehicle will arrive at a particular station. For an example, assume Vehicle A arrives at 6:00 and Vehicle B arrives at 6:10. To calculate the time gap, we need to subtract the arrival time of Vehicle A from the arrival time of Vehicle B, i.e. 10 minutes. Relevant parameters include line, station, date and period. Using the No. 228 bus as an example, we chose the data of arrival time on Friday and Saturday to perform a contrast analysis.

We used the Fishier cluster algorithm to analyze the raw data of the arrival time-gap on Friday. As shown in Fig. 6, the middle data point “0“ represents a data flaw in this period. We can see from the figure that the actual arrival time interval fluctuations are very large. The time gaps are related partially to changes in the departure time and also to traffic congestion conditions. The distribution trend is also hard to be see in Figure 6.

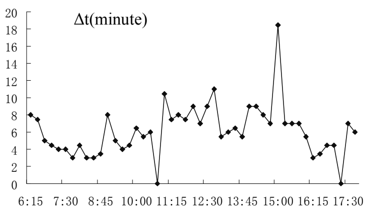


Fig. 6 Raw data of arrival time-gap on Friday

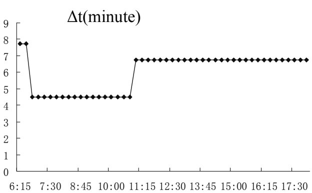


Fig. 7 Cluster result of arrival time-gap on Friday

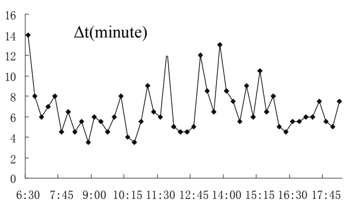


Fig. 8 Raw data of arrival time-gap on Saturday

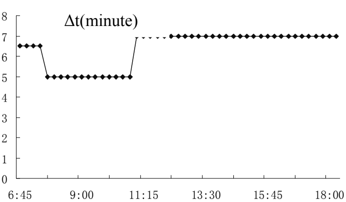


Fig. 9 Cluster result of arrival time-gap on Saturday

Through clustering analysis, we can obtain the result as shown in Fig. 7: it appears that the distribution trends of arrival time-gap is inconsistent with time schedules. The low traffic peak time interval in time schedules is 7 minutes (some are 6 minutes) and interval for peak hour is 4 minutes. But the actual result is that the average low traffic peak time interval in time schedules is 8 minutes, and the average peak time interval is 4.5 minutes. There is a small running time fluctuation between the departure stations and Government Station, but we can analyze the causes through passenger flow and road congestion and then adjust programs.

The original data of the arrival time-gap calculation for the line 228 Government Station on Saturday is shown in Fig. 8, from which we can see large fluctuations, but difficult to see differences from that of Friday. The arrival time cluster analysis results on Saturday are shown in Fig. 9. Generally speaking it is distributed more evenly with the time interval less than Friday. If we combine the running time inter stations on Saturday in the analysis, the result is that though there is a peak time on Saturday, it varies smoothly and lasts a longer time.

The arrival time interval analysis can be used to check vehicle violations, determine which vehicles are not arriving on time, and to list all vehicles' arrival time through analyzing raw data. Furthermore we can analyze the changes in the implementation of traffic travel schedules and the traffic congestion. The longer arrival time interval indicates that there is a larger customer flow at the site and the site will be seen as the main station on this line.

3 Conclusion

The application of information technologies to traditional public transportation can assist management decision making and improve service levels. A bus-carried terminal not only facilitates the work of drivers, but is also of great significance for comprehensive bus operation management. The above analysis is only the beginning, given the widespread use and advancement of technology, GPS data can expand the scope of the analysis. We can make full use of data warehouse and data mining technologies to improve intelligence gathering and scientific management of public transportation.

Reference:

[1] Wang wei, Yang xinmiao, Chen Xuewu etc. "City Mass Transportation System Plan Method and Management Technology". Beijing: Scientific publishing house, in 2002. P.210-223

[2] Han yin, Yang Zhaosheng, Hu Jianming. Dispatch foundation data forecast method research of Intelligence transportation system .Highway traffic science and technology, 2,003 Vol.20 No.3P.140-143.

[3] Zhu Xiaofeng. Data sampling method research. of Intelligence transportation system .Transportation systems engineering and information,2,004 Vol.4 No.1 P.27-29.

[4] Liu Xingjing. Application research of Data warehouse technology in ITS data management. Transportation and computer, 2003 Vol.21 No.4 P.12-15.

[5] Shi Qixin. Application of GPS technology in intelligence transportation system. China guard product information, 2004 No.3 P.19-25.

Transit-Oriented-Development (TOD) Suitable for China

Li Zhiqiang[1], Chen Feng[2], and Richard G. McGinnis[3]

[1]Master, School of Civil Engineering and Architecture, Beijing Jiaotong University; Address: Room 124, Building No.4, Xue Yuan Apartment, No.18, East Road of Beijing Jiaotong University, Haidian Strict, Beijing China, 100044; E-mail: lzq-1122@163.com

[2]Vice-president and Professor, Beijing Jiaotong University; Address: Beijing Jiaotong University, Xizhimen, Shangyuan Village, Beijing China, 100044; E-mail: fengchen@bjtu.edu.cn

[3]Professor of Civil and Environmental Engineering, Bucknell University; Address: Department of Civil & Environmental Engineering, Bucknell University, Lewisburg, PA 17837, USA; E-mail: mcginnis@bucknell.edu

Abstract: Transit-oriented development (TOD) combines transit infrastructure with urban land utilization in ways that optimize the long-term benefits of urban mass transit (UMT). TOD has been successfully implemented in the USA, Europe, Singapore, Hong Kong and many other places; however, its practice in China is so uncommon that few people understand it well. Although it is worthwhile to study the foreign TOD models, Chinese cities cannot simply copy these examples because of how its politics, economic conditions, and regional cultures differ from the countries where TOD has previously been used. Chinese cities should analyze the successful TOD cases to determine how the various aspects of TOD apply to conditions existing in China and ultimately develop TOD designs and procedures appropriate for China.

Keywords: urban rail transit; transit-oriented-development, TOD; city planning; land utilization

1 Transit-oriented development (TOD)

Today the problems of urban traffic, urban land utilization, and pollution are becoming more and more severe. Transit-oriented development is an effective way to combine urban transport construction and land utilization to make our cities more efficient and sustainable.

TOD takes place around the UMT stations where access is maximized. Intensive land uses that require high levels of accessibility are located close to the station and less intensive uses are put farther away from the station. The idea is to use the appropriate development patterns to create large passenger flows that can be handled efficiently since they are close to the UMT stations. In turn these heavy passenger demands help to support the operation of the UMT system.

Since urban rail transit has large capacity, high speed, and low pollution, TOD can take full advantage of these characteristics to create land use patterns that increase the use of public transport, reduce congestion and air pollution, and move the city closer to being sustainable.

2 Requirements for successful implementation of TOD

According to previous experience in the USA, Europe and other locations, implementing TOD successfully requires the following specific actions.

2.1 Ensure that economic conditions and the real estate market are sufficient to justify new development.

The question of whether the real estate market has adequate demand must be considered first, which mainly concerns whether the development of the districts around the urban mass transit (UMT) stations can meet required economic conditions. If the conditions above cannot be met, the planning of TOD is not justified and can potentially do harm to the prosperous development of the city.

2.2 Make a comprehensive plan in advance and handle the relationship between UMT network planning and city planning.

The goal of TOD is to orient passenger flows and new development in the vicinity of UMT so that travel distances to stations are minimized and economic development potentials are maximized. For TOD to be successful UMT construction should happen either before the land development or concurrently with it. Urban planning principles suggest that development around the urban rail transit stations should be integrative, intensive and multifunctional. While designing the urban functional layout, it is important to ensure that development patterns are coordinated with the UMT network.

2.3 Make sure there is adequate availability of developable land and flexible land development policies.

Leaving sufficient undeveloped land along urban rail transit is one of the essential prerequisites for TOD planning. In addition, flexible land development policy can eliminate institutional obstacles, accelerate the integrative development of the land alongside the UMT, and maximize the economic benefit of TOD.

2.4 Increase the capacity and service levels of public transport through policy changes, new financing, and other means.

Public transport must be recognized as the top priority in urban development. In order to carry out TOD planning successfully, the service level and competitive advantage of public transport should get promoted through the effective coordination of all the transit modes in the transport system. The ultimate goal will be that the UMT stations are provided with public transport networks dense enough to meet the passenger demands, and there should be convenient transfer facilities between urban rail transit and bus systems that are rationally planned and joined with residential areas through pedestrian paths and bicycle-ways.

2.5 Plan integrated "TOD villages (TOD communities)" in the literal sense by the center of the UMT stations.（Zheng Mingyuan, 2006 & Ma Qiang, 2004）

Once the above terms and conditions have been met, the principles of high density, diversity, and compatibility should be incorporated into the TOD planning so that the areas around the stations and along the rail lines can get fully developed and integrated TOD communities can be developed properly.

3 Present applications of TOD

The use of TOD is relatively common in the USA, Europe, Singapore, and Japan. Most of these uses of TOD have occurred with large scale urban rail transit projects.

As early as 1947, Copenhagen, the capital of Denmark, began to implement "the figure plan" as the first step of its TOD program, in which the city was planned to be developed in five directions and the urban layout would be just like five fingers. (Feng Jun, Xu Kangming, 2006) Through hard work over the past few decades Copenhagen has achieved integrated TOD, as the development of its surrounding areas was carefully coordinated with its complex railway system. In the city center pedestrians and bicycles facilities have been integrated with the public transport system so effectively that Copenhagen has been able to maintain the style and features of its Middle Ages city center while still being able to handle modern-day traffic flows. Meanwhile, Copenhagen, with one of the highest incomes per capita, has the lowest number of cars per capita, proof that public transport can play an important role in urban transportation.

In recent years, Hong Kong and Guangzhou in China have begun to implement TOD. By incorporating TOD into its urban development in recent years, Hong Kong has been able to develop its needed public transport with the least acreage of land possible. Hong Kong achieved these successful TOD projects by coordinating its real estate development with the financing and expansion of its urban rail transit. Today Hong Kong has a modern, multimodal public transport system with canonical-based traffic management. By using business-like practices in its rail transit construction and operation, Hong Kong has become the only city in the world that generates a profit from its urban rail transit. (Zheng Jiefen, Liu Hongyu,2002)

Guangzhou's Line 3 and Line 4 (north section of Nansha station) of its UMT system used ideas taken from successful TOD experiences in America, Japan and Hong Kong. The government has tailored development patterns to match the different conditions of the individual UMT stations and has positively advocated the development of bus (subway) zones. Also, the designs of the station entrances and exits incorporate the optimum pattern of land uses and facilities. Ultimately, TOD will be an important aspect of these new communities. (Zhang Baoxian, 2006)

4 Discussion of TOD Suitable for China

TOD has been used successfully in many cities abroad; however, its practice in China is far less common. Although the TOD experiences overseas richly deserve consideration, because of the differences in politics, economic conditions, and regional culture, Chinese cities should not simply copy foreign TOD methods. Instead, Chinese cities should analyze the successful TOD cases to determine how the various aspects of TOD apply to conditions existing in China and ultimately develop TOD designs and procedures appropriate for China.

Given the state of development and spatial layout of most Chinese cities, planning TOD for the long term is proper and feasible. However, currently many of these cities lack the prerequisite conditions for successful TOD planning. If TOD is to be incorporated successfully into Chinese cities, the following changes must occur.

4.1 Accelerate the construction of public transport systems, especially urban rail transit.

Today, Chinese cities have common problems of high population density and antiquated public transport systems. In all the large cities, the proportion of passengers carried by public transport is mostly less than 10%, and Beijing is unique in that its proportion is about 30%. This situation is contrary to the idea of TOD. Therefore it is essential that policies be changed to emphasize public transport and simultaneously discourage use of private cars.

Considering current conditions in Chinese cities, developing large-capacity, high-speed and low-pollution urban rail transit makes a lot of sense. If Beijing and Shanghai could complete the construction of their large-scale urban rail transit networks simultaneously with TOD, these two urban transport systems could demonstrate the full potentials of TOD. Then other Chinese cities could use these systems as examples for their own UMT/TOD expansion projects.

4.2 Modernize the existing land development policies and urban mass transit planning/management procedures and improve associated rules and regulations.

First, Chinese cities cannot take full advantage of the successful experiences of TOD abroad. To some extent, antiquated land development policies and urban mass transit planning/management methods are obstacles. Chinese cities need to make the policy and management changes that would allow them to adopt the process used abroad in which the planning for UMT projects considers investment, system design, system operation, and land development as an integrated process. This approach aids in optimizing resources during each step in the UMT development process, and in the end, life cycle costs are minimized. If at the same time, TOD is considered for the lands in and around the stations, additional benefits from the appreciation of real estate adjacent to the UMT system should occur. Only in this way, can TOD achieve its full economic potential.

Second, China has not enacted laws on subsurface development and utilization which restricts the high-density and comprehensive land development required for TOD. These legal deficiencies must be addressed if TOD is to be successful in China.

4.3 Control the timing and sequencing of TOD construction.

The timing of TOD construction is crucial. If it occurs too early there will be insufficient passenger flows for efficient operation of the UMT system which will lead to large operating deficits and other wastage. If the construction occurs too late, the land development along the line may be so dense that TOD can no longer be planned and implemented because of the lack of sufficient vacant land.

Taking UMT Line No.13 in Beijing as an example, although TOD was not considered in its original planning, its construction actually stimulated new urban development in the north of Beijing city that closely resembles TOD. The overall shape of Line No.13 looks like the letter “U”. Since its completion and start of operations in October 2002, the eastern and western sides of the rail line have been always busy. Its passenger flow peaks in the morning and evening when people are commuting to and from work. By contrast, the passenger flows of the northern section of the line (bottom of the “U”) are much smaller which forces the government to provide large on-going subsidies. Many people have suggested that it would have been better to have constructed Line No.13 in stages. That is to say, at the beginning, only the eastern and western sides of the rail line where sufficient passenger flows existed should have been constructed. Then, when the passenger flows in the northern section reached required levels, the final section should have been built. (Liu Jinling, Zhang Yong, 2004)

4.4 Consider coupling TOD with SOD if possible.

In railway planning, there are two approaches to designing the networks and associated land development. SOD (Service-oriented development) aims at utilizing the large capacity of rail transit to reduce traffic congestion in the city center and therefore costs a lot and takes a long time to implement. The goal of TOD (Transit-oriented development) is to induce passenger flows through new land development coordinated with rail transit expansions. Often TOD is used to create a series of densely developed villages (or new towns) away from the center city to reduce center city congestion through decentralization. TOD generally costs less and can be implemented in a shorter time than SOD, but its near-term benefits are generally lower.

Each of these approaches has its own advantages and disadvantages, and their reasons for implementation are also different. However, the biggest difference is that SOD is based around central urban transit while TOD is focused on fringe or suburban transit. As a result, because of the temporal differences in peak flows between central city and suburban areas, the two approaches derive benefits from passenger flows at different periods of time. Thus, when planning railway networks, it is important to coordinate TOD with SOD if possible so each of these two approaches can be integrated to optimize the long-term benefits of the overall transit system.

5 Conclusions

Actually, many regions in China have tried or are pursuing TOD. The subways of Guangzhou and Beijing are examples.

TOD has been considered in the present plans for the Yi Zhuang and Da Xing subways in Beijing. However, due to the limitations of out-dated public transport planning and management, China still has not been successful in implementing TOD.

But through the assistance of the Eleventh Five-Year Plan and the reform of urban rail transit planning, it is hoped that more and more Chinese cities will decide to incorporate TOD in their UMT expansion plans.

References

Feng Jun, Xu Kangming. (2006). “Research on TOD Pattern in Copenhagen.” *Urban Transit*, (2006).

Liu Jinling, Zhang Yong. (2004). “Analysis on the Influence of the Passenger Flow by the Land Utilization in the Northern Section of the Line No.13 in Beijing.” *Traffic Forum*, (2004).

Ma Qiang. (2004). “The Study of the TOD in North America in Recent Years.” *Urban Study Abroad*, (2004).

Zhang Baoxian. (2006). “The Development of TOD Pattern in the Rail Transit in Guangzhou-Taking Line No.3 and No.4 for Example.” *Constructing Technology*, (2006).

Zheng Jiefen, Liu Hongyu. (2002). “The Urban Rail Transit in Hong Kong and Its Integrative Development of Land Resource.” *China Rail Science*, (2002).

Zheng Mingyuan, (2006). “The Urban Development in the Times of Rail Transit.” *China Railway Publish House*, Beijing, (2006).

Transit-Oriented Development Strategies and Traffic Organization

ZHANG Ru-hua
Traffic Engineering Department of Shandong University, Jinan 250061, China; Email: ruhua_zhang@163.com

Abstract: Transit-oriented development (TOD) may soon become mandatory practice in China due to the rapid growth of traffic congestion in densely populated urban areas. Three core elements of TOD include transit, specific land development techniques and strategic organization of transportation systems. Well organized transportation systems provide transit, automobiles, cyclists and pedestrians with a contemporaneous environment and an attractive alternative to traffic congestion. Two key components include access management and nodal development.

To investigate the three core elements of TOD, the Beiyuan Bus Rapid Transit (BRT) corridor in Jinan, China was taken as the study case. The study showed that it is most efficient to set dedicated pedestrian corridors at each end of the transit stations which are spaced on average 200 meters apart and set dedicated vehicle corridors between the pedestrian crossings. This planning strategy serves to delineate the separation of human and vehicle. An analysis model was developed to demonstrate that for each trip one minute can be saved using set dedicated cross pedestrian corridors at both ends of a station as compared to one end. The research followed the stipulations that traffic organization is consistent with BRT station design, the adjacent stations are considered together in one transportation system which means traffic organization is continuous along the corridor, and the existing land development is protected as possible through setting parallel vehicle corridors.

I. Introduction

There are two main development strategies within a city: Car-Oriented Development (COD) and TOD. COD has the tendency to encourage and/or contribute to urban sprawl, low efficiency of land use, severe traffic congestion, serious air pollution and a heavy burden on infrastructure, causing a series of unconquerable social, economic and environmental problems. Considering these problems, new city designers proposed the TOD theory with emphasis on integrating public transportation and high density land use.

TOD is ideal in urban areas for coexistence of the pedestrian and vehicle. The land development technique in transit-oriented cities is the product of organized transportation systems and sustainability. This land development technique is emphasized in developing countries as well as in existing countries. The main characteristics of TOD include well-organized, mixed land use, the coordination of land use and public transportation, nodes of public activity linked by transit, good walking and bicycling environments and enforcement of corresponding policies and systems.

This paper recognizes development strategies in China are below optimal and states the significance of the TOD strategy, analyzes the TOD traffic organization technique and proposes the conceptual traffic organization scheme as seen in Jinan, China.

II. The Significance of TOD Strategy

TOD is the basic means of solving and mitigating traffic problems. In contrast, COD contributes to the vicious cycle of reconstructing infrastructure – construction of traffic facilities, capacity reached and exceeded resulting in traffic congestion, expansion of traffic facilities, expansion of cities and so on until severe traffic congestion prevents further development, as seen in Figure 1. Once COD is initiated, it is impossible to convert to TOD.

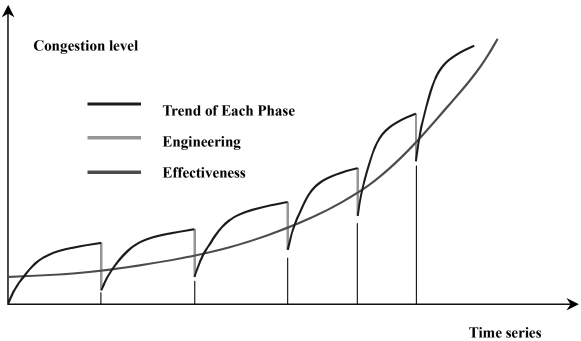


Fig. 1 the relation between traffic capacity and congestion

TOD allows transportation systems to perform its original function in urban development. Without the uninhibited flow of traffic "the whole city can't function properly and the city has lost its original serving function." Since the 1970's we, globally, have been in the post car traffic (PCT) era in which the main purpose is to promote multi-transportation modes and to return to the traditional human society through optimizing the transportation modes and land use to create "human livable" rather than "car livable" cities.

TOD encourages reform of strategies to conservation in order to guarantee sustainable development. This mainly involves conservation of energy, water, raw materials, land and mines and recycling. TOD also encourages conservation through the intensive use of traffic facilities and land resources and discourages the reliance of economic development on land loss.

TOD prevents the expansion of urban scale and the exponential spreading of suburbanization much unlike COD. Between the 1970's and 1990's the urbanization of population and industries was developed in America but at the cost of transportation, environment, land use and the quality of life. With traffic congestion at its worst, Americans dream of what Peter Calthorpe stated - "Land and energy can be used efficiently; traffic can be greatly reduced; living expenses can be affordable; the old and the young can be easily accessible and working people don't need to bear the tiresomeness of long-distance travel from home to work…"

TOD supports sustainability and healthy development while COD promotes excessive consumption of land and energy and greatly impacts the living environment. Sustainability consists of the security of crops, energy and the environment. The negative impacts of COD on sustainability are too great for a county with a population of 1.3 billion and relatively scarce resources. The ratio of farming land per capita is 1.41 acres in China as compared to 25.9 acres in Canada, 13.1 acres in Russia, 10.9 acres in America and 2.83 acres in India. TOD is the only option for China to focus on sustainability and future development which should be "restricted to a designated area."

TOD is a reliable strategy in the process of rapid urbanization and motorization. In such an aggressive development era, it is impossible to correct a wrong choice of development strategy. The Chinese government has begun to emphasize on public transportation. For two consecutive years, 2004 and 2005, the Construction Ministry has promulgated "proposal on giving priority to public transportation in city development" and asks every city to put it into practice. TOD concepts should be included in national policies and once written will evolve into a way of life.

III. Research on TOD Strategy

The essence of TOD can be summarized in three elemental parts – public transportation as the premise, land development as the guarantee, and traffic planning as the knot.

Public transportation should serve as the skeleton of city transportation. Transit with large capacity, such as orbital traffic or bus rapid transit (BRT) should serve as the backbone, supplemented by all grades of public transportation lines to form the organic and synchronized public transportation network. Transit stops are the connecting tissue between the public transportation system and land use and should be optimized according to different grades and functions. Priority should be given to public transportation in road resources allocation and traffic management control and the remaining road resources can be allocated to private cars with a certain usage fee.

Using the TOD strategy land development involves two levels of planning – the first being city-wide planning and the second giving focus on the areas influenced by transit stops. The goal of the TOD strategy is mixed and intensive land use. Mixed development does not mean the same in land use form and land use quality. The distinctive arrangement of land uses should be made in the microcosmic sense, attaching to transit lines and stops. Decisions on land use quality and intensity should be made based on surveys about walking distances from the homes of transit users to the transit stops, walking conditions and service level. To increase the efficiency within the area influenced by the public transportation system and to minimize limitations on land development density, it is essential to prohibit development outside the designated area covered by the public transportation system. Existing cities should make full use of the opportunity of reconstructing to carry out TOD and developing cities should initiate TOD.

The organization of transportation modes should consist of public transit, pedestrians, bicyclists and public automobiles with priority being given to transit which carries the foundation of conservation construction and sustainable development. The organization of transportation modes will be analyzed more specifically in the next section.

IV. Organization of Transportation Modes in TOD

The basic principles of transportation organization include giving primary priority to mass transit and secondary priority to pedestrians and bicyclists, limiting the use of cars and separating pedestrians from automobiles. The basic methods of TOD can be summarized as three coexisting networks and four parts in agreement as seen in the chart below.

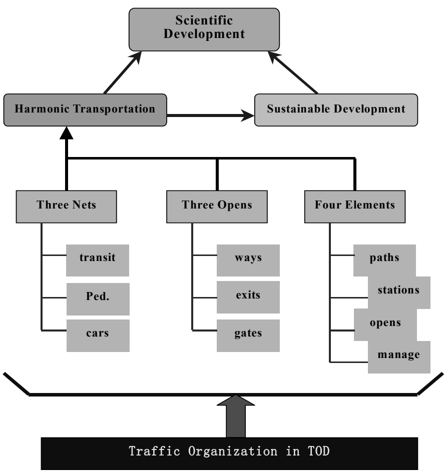


Fig.2 Concept of Traffic Organization in TOD

The coexistence of three networks refers to the cooperation of the public transportation network, pedestrian and bicyclist network and automobile network. The reasonable organization of the pedestrian and bicyclist network is key to implementation of the TOD strategy. A safe and attractive pedestrian environment and high accessibility to transit stops will reduce bus transit time, effectively curb the demand for vehicles, improve people's reliance on public transportation, give support to the high density land development and enhance the land use value.

Four parts in agreement refers to the comprehensive planning and coordination of passages, stations, access management and management control. Passages include those of transit, pedestrians/bicyclists and automobiles; stations include transit stops, parking lots and public open spaces; access management involves the management of feeders and exits and gates along the way; traffic management control involves demand, organization and signalization. All of these aspects should be systematically optimized in terms of transportation planning, design and management.

V. Case Study

The Beiyuan Street BRT Corridor in Jinan, China was chosen as the target of study. With consideration given to the geological structure of Jinan and the advantages that a BRT system brings, implementation of a BRT system was suitable for the city of Jinan. Jinan became the first model city with a BRT system. The project was sponsored by American Energy Foundation in 2005 and Beiyuan Street was designated as the first BRT street which covers about 12 kilometers with 20 stops linking 608 meter segments. Jinan is set about the BRT network which covers the entire city. This paper intends to design the transportation organization scheme with TOD principles and methods of the coexistence of three networks, the four parts in agreement and the separation of pedestrians and vehicles.

The standard length of a BRT platform in the study is 132 meters and the distance between crossings on either side is 192 meters as shown in Figure 3. The parallel lanes along the crossings are dedicated to pedestrians and bicyclists and the passages between crossings are dedicated to motor vehicles including transit as shown in Figure 4. According to the model, establishing parallel passages on both sides will save on average one minute of travel time within 500 meters compared to establishing a parallel passage on one side. To protect established development and the integrity of TOD and to reduce the number of exits for motor vehicles toward the BRT corridor, the roads for collecting and distributing motor vehicles should be built at least 50 meters away from the corridor on both sides. The parallel passages dedicated to pedestrians and bicyclists should be built 100 to 200 meters away from the corridor on both sides. Figure 5 demonstrates this separation of pedestrians and vehicles and convenient connection to BRT platforms.

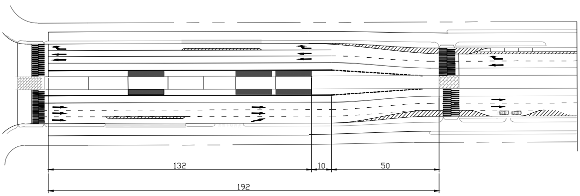


Fig. 3: The Design of BRT Platform

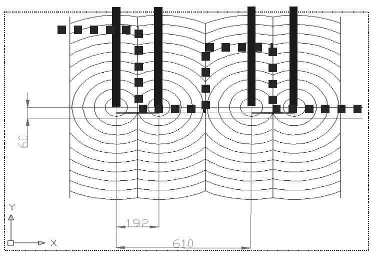


Fig. 4 Traffic Organization Concept of BRT Corridor

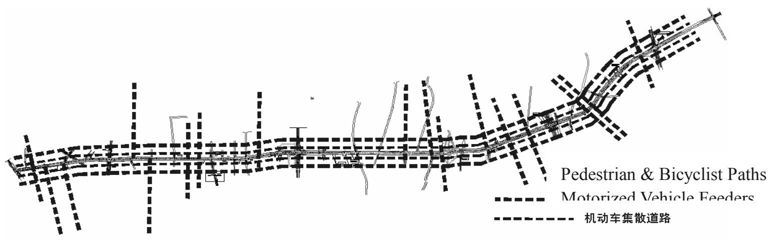


Fig. 5: The Traffic Organization Scheme Design of BRT Corridor

VI. Conclusion

Until now the process of human development has led to transportation problems. Human ability may be enhanced but long term goals may be sacrificed. A transportation system relying on the automobile is a prime example. TOD seems to be a reasonable and attractive solution; however, whether TOD is capable of solving all sorts of traffic problems and can be fully applied remains to be further investigated and executed. In sum, this paper explores the TOD strategies and the corresponding organization of transportation systems by analyzing the large capacity BRT system in Jinan, China.

References:

1. Chen Xueming. BRT and Chinese Transit Priority Development Strategy [J]. City Transportation, 2003, 27(10): 28-33,38.
2. Eric Holeman. Selecting Corridors for Bus Rapid Transit Using a Multi-criteria Method [A].TRB 2004 Annual Meeting CD-ROM.
3. Hong T.,. City Road Planning in Japan [J]. City Planners, 2005, 21(7): 118-122.
4. Lin Yan, Wei D., Liang Ge. The Transit-Oriented Development City Land Development Mode Research [J]. Trans. Engrg & Infor. Journal, 2004, 2(4): 90-94.
5. Luo Sidong. The Spread of American Suburbs: The Analysis of Traffic Congestion and the Loss of Land Resources [J]. City Planning Journal, 2005, (3): 43-46.
6. Mark A. Miller. Bus Rapid Transit: Institutional Issues and Strategies for Resolution[A]. ITS 2001 Annual Meeting CD-ROM.
7. Ren Futian, Xiaoming Liu, Jian Rong. Traffic Engineering [M]. Beijing: People's Communications Press, July, 2003.
8. Zou Zhe, Jinsong Zhou. To Create a New TOD Concept for the 21 Century [J]. City Transportation, 2005 (4): 42-44.

Six-Sigma: Delivering Quality to Mega Transportation Projects

by
Vinay V. Mudholkar
Princeton Junction, New Jersey

Keywords: Six Sigma, Quality delivery, Mega projects, Construction organization, Contracts, Worker Behaviour, Technology, Process.

Quality in Mega Projects:
Quality delivery to Mega Projects is a growing vital concern to all owners, engineers and program managers. As the scale of projects increases, complexity of delivery increases leading to quality control problems. There are many examples out there which are well documented in technical papers as well as public press! This issue certainly leads to degrading public confidence in large complex projects. Rework needed to rectify defects leads to further increase in costs and time consuming litigations. In certain safety critical systems such as High Speed Rail Line, Nuclear Power Plants, Air-craft or Space vehicle construction etc. defect rate must be reduced to zero to avoid component failures which may lead to total system failure causing a lower performance, shutdown or even an accident.

Six-Sigma Methodology for Best Quality Delivery:
Six-Sigma methodology has been successfully used by manufacturing companies for many years and amongst them GE and Motorola have been on forefront of the application for best quality achievement. In recent years, many other product manufacturing companies have followed their lead for quality improvements, cost reduction in rework and consistent consumer satisfaction. Methodology is relatively new to rail and construction industry process applications in construction activities. Recent experiences in UK's West Coast Route Modification upgrade project to increase speed at 125 mph validate by case studies the fact that the methodology can be applied to construction as well as administrative activities to deliver quality, increase project efficiency and reduce cost of rework.

Bechtel/Network Rail team effort in UK:
Bechtel's West Coast Engineering Program Management team and Network Rail (client) have made significant corporate commitments to train project staff to achieve such gains. Contractors and sub-contractors are actively participating in such Process Improvement Programs/ Projects (PIP) and are delivering required results.

Lessons learned from these case studies can benefit US railroads, transportation companies, construction firms, and engineering community at large. Six-Sigma methodology is new to railways and construction industry. Its success for a particular firm or a construction program/ project depends on commitment of management and staff, not only for short term gains but also for long term operation of the firm. It is not a "flavor of month" activity, it requires serious long term training of staff and

resource commitment, which will definitely payoff in short term PIPs as well as long term improvements to Construction Operations. It should become a way of life for examining construction/ project processes on a continuous basis to achieve best results and reduce cost of rework. In production processes Cost of Poor Quality (COPQ) can be significant.
In delivering high quality mega projects, all efforts be made to eliminate COPQ. Construction firms can achieve better profitability by drastically reducing cost of rework, by delivering required quality for first time. Six-Sigma improves reliability, safety and ultimately improves public/ client confidence and relations.

Case study: UK's West Coast Main Line Upgrade Project: Improvements to Rail Welding:
Project Background: West Coast Route (Main Line from London to Manchester to Glasgow, Scotland) was initially constructed in 1840 and today it is the busiest mixed traffic route carrying 2,000 passenger and freight trains per day. 450 mile route has 1660 track miles, over 10,000 structural spans and 13 major junctions. Route is electrified and was upgraded under continued operation on most segments. Due to its heavy transportation demands, project delivery time table has been critical and upgrading is planned by segments, in stages. First major phase of upgrade from Euston, London to Manchester to achieve speed of 125 mph has been successfully completed in Fall of 2004; second phase to Glasgow, Scotland is expected to be complete by Winter 2006. Terminals and Junctions are being modified to meet the needs of proposed traffic flows and work is expected to continue beyond 2006. Total project is expected to reach 10-12 Billion pound investment. It is certainly one of the largest rail investments in Europe.

A mega project such as West Coast, certainly demands delivery on time and best quality to meet the standards of High Speed. Large number of designers, engineers and contractors have been participating in the project delivery. Project construction schedule is complex due to various multi functional activities such as track and structures renewal works, geotechnical improvements of sub-grade, tunnel rehabilitations, signal and communication upgrades, overhead electrical works, renewal of power system,
preparation of design plans and specifications, client approval of submittals, material and equipment acquisition etc. As we closely examine many of these project activities, we find many common processes where Process Improvement Plans (PIP) can be applied to deliver consistent quality to comply with standards established. Examples of some of the repetitive process are:

1. Design plan submittals and client approvals
2. Reporting process of Contractors
3. Material Acquisition and delivery
4. Fabrication
5. Field and shop welding of rail, testing
6. Track surveying and tamping
7. Concrete curing, testing
8. Material distribution by work trains, trucks

9. Many others

A case study presented here confirms merits of Six-Sigma methodology to achieve desirable quality delivery of the technical as well as administrative processes.

Summary of Six-Sigma Process:
Operational excellence can be achieved in complex organizations, mega-projects by adopting Process Improvements in various measurable areas. As concluded by previous successful application:
"Six Sigma focuses on helping organizations produce products and services better, faster, and cheaper by improving the capability of processes to meet customer requirements…" (which are engineering standards defined in construction projects)
" SixSigma is a philosophy of continuous improvements to eliminate defects to near zero level. Six Sigma emphasizes quality improvement, but is much than statistics and tools. When properly deployed on carefully selected business projects, Six Sigma application virtually eliminates defects before they occur, which saves valuable corporate resources…."

Steps involved in process can be summarized briefly as:

1. Identify process/activity that is not meeting standards/ goals and needs improvement. It must be definitive, measurable, production oriented/ repetitive, may or may not be technology based (can be administrative process)
2. Create "Process Improvement Project Team" (PIP Team) that directly affects Quality and/or Production of Infrastructure Works. Assign Champion (Team Leader) for cause who has strong interest in PIP and can unblock any difficulties in the implementation plan; Champion is supported by Six-Sigma trained staff who provides data analysis and required technical support.
3. Create a specific Process Improvement Statement: Goal: which must be measurable and achievable for the team in a reasonable period to improve quality and production. For project purposes, this can not be a very long-term R&D effort. For example: Two case study goal statements for welding PIP were:
   A: Achieve defect rate under 1% for field Flash Butt (FB) Welding process
   B: Achieve defect rate under 3% for field Thermit Welding process
   C: 0% defect rate in laboratory testing of welds (bending, cyclic loading)

   Others can be:
   D: 99% of concrete samples pass in testing
   E: Install new track within 2 mm of gauge variation,10 mm in designed position
   F: Obtain engineering drawing approvals from client in 10 days of submission
   (Example of administrative process)
4. Conduct technical workshops with client staff, project staff, contractors, manufacturers of process/technology (when required), and construction engineers. This team effort is very critical in establishing communication and

gaining credibility to PIP. By only such a team effort, the goal can be achieved.
5. Training/ retraining of staff/ crews for improved methodology is vital not only for project but overall long quality production and attitude in the organization.

Sounds simple and straight forward? Not so! Commercial interests, personalities come into play and they must be dealt with. A sound statistically based data analysis to understand "cause effect" relationships of defects will move the improvement process forward. Sometimes Champion has to actively remove roadblocks as the process change is not always easy to introduce. Commercial rewards and penalties also need to be examined as they drive behavior of firms and managers.

Changing human behavior is the hardest part of introducing any change in existing process or procedure. Many managers, supervisors, technicians are unwilling to even recognize that problems exist due to continued past practices over years. It is important to clearly state process goals, expectations in contracts, daily monitoring and job meetings of what is expected and all efforts be made to avoid cost of rework. To influence behavior, incentives and penalties are essential in contracts and be introduced at lower levels of project control. Examination of tools, skills, training and field procedures and supervision will be necessary as statistical data provides regular process information for understanding contributing factors to defect rate. (Parito charts, flow charts etc.)

Six-Sigma Methodology: Steps summarized: DMAIC:
The following table summarizes steps involved in DMAIC approach:

Define:	• Exact definition of Process Goal to be achieved
Measure:	• Team measures field observations and create a database for statistical study
Analyze:	• Tools: Process Maps, cause and effect matrix, Parito charts, Fishbone diagrams for Primary and Secondary causes
Improve:	• Team work shops, discussions for improvements, solutions
Control:	• Monitor process improvement for continuity/ savings

Welding case studies on High Speed Railway of West Coast, UK:

Best quality welding is highly essential for rail integrity of high speed railway. With the support of the logistic unit of Network Rail of UK, a system-wide weld inventory data base was set up to identify: 1. Line 2. Location 3.Welder 4.Weld ID 5. Contractor
6. Sub-contractor 7. Defect type 8. Date etc.
A sub-set was developed to show defective replacement rework. This inventory/database provided a valuable start for six-sigma analysis. The results of analysis by individual firm, welder and defect types provided a very useful insight on

what was happening in understanding poor performers and defect causes. Removal of defect under traffic is expensive and project management, client staff realized the potential of cash savings that can be gained by improving weld quality urgently. This potential was discussed with senior management of project and railway and it was agreed to hold a joint workshop with all concerned to take PIP forward.
Communication was also kept open with commercial/ contract administration staff and they also participated in contract discussions and modifications for future work.

PIP I: Flash Butt Welding Process:

Chicago based Holland Company participated in these discussions as they were providers of 2 mobile welding machines and operators on the project. In PIP it was agreed that the defect rate below 1% can be achieved technically. Holland Company provided experienced operators from US and training was also arranged for UK welders for better production of quality. Welding heads and operating systems required modifications to achieve higher project standards. Continuous monitoring over different line segments of route did bring down defect rate from initial 2.9% to 0%. Intelligent operating system of welding machines virtually eliminated operator's errors. Welding operation was closely followed by grinding and inspection; reporting defects immediately. Production rate is very much dependent on availability of track time. For very short time availability where access was an issue, thermit welding process was utilized.
Flash butt weld samples were tested satisfactorily by independent testing facility of AEA in UK as per required criteria for railway. Tests included: 1. slow bend test 2. metallographic examination 3. Hardness traverses 4. Fatigue testing to 5 million cycles
Results achieved were more than satisfactory.

PIP II: Thermit Welding Process:
Due to unavailability of longer possession times, thermit welding process is commonly used by railway contractors and sub-contractors. Process provides better access and mobility, but tend to generate higher defect rates. Increased rail replacement programs in UK have put large demands on welding skills and there has been shortage of highly skilled welders and experienced supervision. In certain areas of US we are also facing similar problems of skilled technicians in construction industry.
During PIP study, analysis of data indicated that Principal contractors who have sub-contracted the majority of welding work to meet peak demands for night work and week-end work, showed very high defect rates as compared to contractors who used their own trained staff. As the number of sub-contractors increased, defect rate increased due to lack of supervision and control. Initially defect rate as high as 12% was observed with Principal contractors who largely subcontracted the work. This indication warranted immediate attention of construction and commercial (contract) managers to take appropriate actions to remedy the situation.

It was concluded in the workshops that the technical process can deliver defect rates under 2 to 3 %. Statistical data analysis indicated that the following primary factors were major causes leading to high defect rates:
Weld Defects: Primary Factors (Data elements from "Fish-bone chart")

Commercial Terms	**Technical Details**	**People Skills**	**Materials/Equipment**
No penalties for non-compliance	Substandard training	Retraining needed to understand specifications, goals	Wrong molds
No incentives for good work	Operator errors	Shortage of skilled staff, high demand	Unchecked field equipment / process
No regular feedback	Time pressure due to short work period	No periodic evaluation	Lack of site preparation
Ownership issue	Lack of supervision	Shortage of field supervision	Just in time ordering Low inventory
Small risk to prime	Poor advance plan	Night work issues	Limited access

Action Plan: Weld defect data was carefully analyzed by individual firms and individuals and "high defect generators" were identified and were notified of their poor performance.
Closer supervision, retraining, better equipment etc. were recommended corrective measures for improvements and achieving goals. Certain sub-contractors who were large contributors to defect rates clearly lacked professional skills and were not to be considered for future work, without extensive training.

WCML project as well as rail system as a whole certainly benefited from this Process Improvement exercise along with other Six Sigma PIPs. Six Sigma certainly has earned its place, in my experience, in delivering quality to a Mega Project and I highly recommend it to our construction firms and engineering professionals. A single most important critical success factor was the 'need for perceive change'. A change in focus from 'people' and 'quantity' to 'process' and 'quality' is required. Overcoming 'blame culture' was also found critical and useful for long-term sustainability of the Six Sigma initiatives. Top management commitment is vital for training staff for gaining Six Sigma skills in organization, which in turn will produce short project gains as well as long term quality and financial gains.

Athens 2004 Olympics:
The Importance of a Freeway for an Olympic Size Event

Fanis Papadimitriou[1], Bill Halkias[1], Kostas Papandreou[1], Pantelis Kopelias[1], Panos Prevedouros[2]

[1]Attikes Diadromes S.A., 41.9 Km of Attiki Odos, GR 19002 Peania, Athens, GREECE
[2]University of Hawaii at Manoa, 2540 Dole Street, 383, Honolulu, HI 96822, USA

ABSTRACT

This paper reports on the role of the Attica Tollway, an urban freeway connecting key Olympic venues, the airport, and the city of Athens in Greece, during the Athens 2004 Olympic Games. The paper details the measures taken by the operator of the freeway to ensure that Attica Tollway could fulfil the great challenges presented by this significant global event. It also discusses the lessons learned and performance indicators that justified the successful operation of the Attica Tollway during Games time. In light of the upcoming 2008 Olympic Games in Beijing, the goal of this paper is to present our experience regarding the preparation, management, and operation of a freeway for such a large-scale international event and thus to contribute to the understanding of the issues and challenges faced by this grand endeavour.

INTRODUCTION

Attica Tollway, locally known as Attiki Odos, is a modern 65 km, fully access-controlled freeway, constituting a half ring-road around the city of Athens, Greece. The freeway crosses through the Greek capital's northern suburbs, facilitating the speedy connection between 30 municipalities and all major transportation means and infrastructures in the greater Athens area: the Athens International Airport, metro, suburban and railway stations, ports, the two intercity highways, and major Olympic venues (Figure 1).

The scale of the project is best illustrated in numbers: 38 toll stations totalling 193 toll lanes, 24 grade separated interchanges, 100 overpasses, 4 major Motorist Service Stations, 9 Customer Service Centers, 56 tunnels and Cut & Cover Sections covering about 12% of the freeway's length. The entry to the Attica Tollway is controlled by toll barriers. Drivers pay the toll only once (open system) upon their entry, while the toll is the same for all entry points (flat fee).

The Attica Tollway has been constructed on a concession basis and constitutes the largest co-financed road project in Greece and one of the largest in Europe. The whole project was completed in June 2004, two months prior to the Athens 2004 Olympic Games. By ensuring free flow operating conditions, the Attica Tollway constitutes the "traffic moving" alternative around the extensive metropolitan area of the Greek capital and formed the backbone of the entire road network of Athens during the Olympics, a role that it continues to play to the present time.

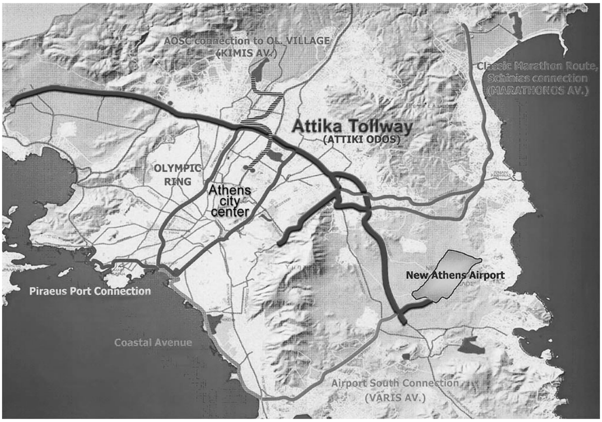


Figure 1. The Olympic Road Network for the Athens 2004 Olympic Games

Attikes Diadromes (AD) is the operating company of Attica Tollway and is in charge of traffic management, road maintenance, and toll collection. AD is operating from a modern Traffic Management Center (TMC), on a 24 hour basis, and coordinates all actions necessary to locate and deal with any kind of incidents. There is quick and efficient intervention at all times and road assistance, a service offered to all in need by trained personnel, free of charge (Halkias, 2005).

In the following sections, the role of and planning procedures of the Attica Tollway for the Athens 2004 Olympics are described. The paper concludes with a discussion of the actions taken by AD to ensure that this once-in-a-lifetime challenge for efficient traffic management during the Games was dealt with successfully.

THE ROLE OF ATTICA TOLLWAY DURING THE OLYMPIC GAMES

The Athens Summer Olympic Games took place in the greater Metropolitan Area of the Greek capital in August 13 to 29, 2004 and were the first Summer Olympic Games in the post “9/11” terrorist attack era. Security was a paramount issue: It was a primary challenge for the organizers and a major headache for the international community. The Athens Games broke many records, since Athens hosted the largest number of athletes ever, from 201 participating countries (more than in any other sport event). Forty one competition venues and the Olympic Village were within the greater Athens area and an additional four competition venues were located in different cities. About four million Olympic vehicles were journeying during the Games while 3.6 million tickets were distributed. About 200,000 people were accredited, of which 75,000 were members of the official Olympic Family (ATHOC, 2005). All of these amounts of activities and extensive transportation needs presented a major concern for the operational units of AD who were preparing not only on the basis of figures and statistics but also on the basis of having the “world” watching, ready to criticize poor traffic management and level of service to athletes, organizers and spectators.

The country’s central responsibility and control of Olympic Transportation was assigned to the Athens Traffic Operations and Control Center. The Olympic

Traffic Management Plan was based on the creation of an Olympic Road Network (ORN), where specially authorized vehicles (fleet, athletes and team officials, technical officials, media, sponsors, members of International and National Olympic Committees and International Federations, etc.), all exclusively marked by boards placed on the windshield, would have priority in traveling to restricted "Olympic Lanes" in these selected city and peripheral roads of the ORN.

The Attica Tollway was the most critical part of the ORN but the only road not implementing the reserved dedicated "Olympic Lane" scheme. Instead, all lanes were to be open to all vehicles. Furthermore, the plan required that Athens Olympic Committee (ATHOC) member vehicles were to travel primarily through the Attica Tollway in order to move faster and safer. Hence, it was AD's role to proactively plan the management of the freeway traffic in order to achieve the required high standard of service and to ensure the quick and safe access to the Olympic sports venues.

PLANNING AND PREPARATION FOR THE OLYMPIC GAMES

Attica Tollway traffic management preparations started well over twelve months prior to the Games. AD set in place a proactive traffic management plan which involved collecting information, data, and solutions concerning the traffic management of the "golden" 20 days challenge; cooperating and negotiating with local public service authorities, ministries, and other entities such as the Ministry of Foreign Affairs for managing the passage of VIPs and Heads of States; working with the Traffic Police for securing the passage of convoys through the toll stations' service lane (with the presence of uniformed officers standing by and managing VIP toll passages); collaborating with the City of Athens on traffic coordination issues and the ATHOC for opening special prepaid Electronic Toll Collection (ETC) accounts to facilitate the quick passage of the Olympic vehicles fleet from the toll stations.

The most critical issue was the application of the "Olympic Lane" concept within the freeway lanes. This became the most sensitive issue in discussions with ministries, the Traffic Police, and ATHOC. Olympic Lane advocates were seeking fast, undisturbed and priority flow for ATHOC and VIP vehicles, thus decided to commit the left lane on the roads of the Olympic Ring to the exclusive usage of VIP and Olympic Family members. On the other hand, AD management was convinced that such an Olympic Lane application along the Attica Tollway, especially occupying the left lane, would cause problems in cases of incidents, accidents, and vehicle breakdowns. Furthermore, the capacity of the Attica Tollway would ensure the smooth flow of all vehicles, including the Olympic Family vehicles, without the need to implement a special Olympic Lane, therefore it was finally decided that no Olympic Lanes would operate on Attica Tollway.

Regarding the passage through the Attica Tollway tolls, AD staff decided to divert all Olympic family vehicles through reserved ETC lanes. ETC lanes normally exclude buses, but especially for the Games, an agreement was reached with ATHOC to allow buses to pass through ETC lanes throughout the period of the Games and this ended up being an important measure. Still, the challenge remained in cases such as before and after the Opening and Closing ceremonies when a large concentration (~500) of Olympic Family vehicles had to pass through a single toll station in order to move between the Olympic Stadium and the Olympic Village. To maintain the required travel time, AD organized a "convoy" formation passage in restricted time schedules allowing vehicles to pass without any disturbances, while at the same time a team was standing by ready to act in cases of need. Overall, ETC prepaid temporary accounts were issued for 6,000 vehicles, of which 2000 were buses.

An Olympic size event requires adequate and well trained staff. In order to comply with traffic level of service requirements, AD recruited and trained well in advance an additional 200 employees, in order to reinforce the readiness of the approximately 850 regular traffic operations employees. Of these, 120 people were recruited to support the toll stations, 30 people were employed as vehicle patrols for incident management and 50 as "ETC lane officers" helping and controlling passage of ATHOC buses and Olympic Family vehicles. The recruitment process for all new staff started five months prior to the Games, a period sufficient for their thorough training. All staff at AD agreed and complied with the "Alert on-Duty Program", which imposed, among other regulations, extra shifts and duties, overtime and no holiday break until the end of the Paralympics Games in mid-September of 2004. The lifting of the ETC lanes height barriers for the ATHOC buses to be able to cross over created extra control and surveillance needs, since staff was needed to make sure that only those buses would pass through the ETC lanes, as well as to be able to respond immediately to any incidents during those passages.

In order to be able to cope with crisis situations, the AD traffic operations team was prepared for facing crisis management scenarios. All Management team members participated in crisis management training courses, learning how to respond to several unexpected occurrences. Examples of such incidents included terrorist assaults to buses or the suburban railway that runs in the middle of the Attica Tollway or incidents that could cause possible freeway closure, especially on crucial days such as the 20th of August when 25 out of 30 Olympic venues were hosting events. This training led to initiatives for tighter cooperation with Traffic Police and Fire Brigade officials, offering AD management the knowledge and mentality needed to be prepared to handle critical situations. An important outcome of the crisis management preparations was the detailed traffic diversion plans that were prepared in case of total closure of the tollway. These were distributed to all cooperating agencies.

The Alert on-Duty Program applied with AD staff was extended also to Traffic Police officers who staffed the Attica Tollway TMC. The same applied to the Athens TMC also. This city-wide 24/7 hour surveillance allowed for better coordination and increased immediate decision readiness. Moreover, under the program, all managers were present on site every day from 6 AM to 2 AM, the Call Center and Help Line were operating on an almost 24 hour basis for both Greek and foreign customers. Finally custom Olympics signage was introduced along the freeway to assist drivers in finding their way to the Olympic venues. Maps of the freeway indicating all Olympic venues were prepared in English and distributed from all toll stations and Customer Service Centers.

EVALUATION OF OPERATIONS DURING THE OLYMPIC GAMES

Judging by the results, the advanced preparation described above proved to be crucial. Everyone involved worked hard and followed the recommended plans. As a result, no serious incidents or crises occurred during the 2004 Olympic Games. To keep matters under control, a Management Team meeting was taking place daily, focusing on the status of real vs. anticipated traffic conditions and setting up priorities for the next day. The Alert on-Duty Program that was in place for the month of August established the permanent presence of a Traffic Police Officer in the TMC and thus a permanent link of communication between AD and the Traffic Police. This cooperation resulted in an average reduction of 4 minutes to the Traffic Police response time during Games time (Figure 2).

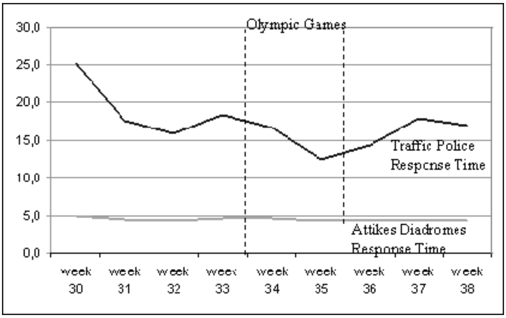


Figure 2. Reduction of Attica Tollway Traffic Police response times during the Olympics.

The management of traffic, the coordination between all involved agencies, the coordinated implementation of Olympic action plans and the management of non-scheduled incidents affecting traffic, demonstrated excellence, as there were no serious consequences from traffic incidents. Moreover, a certain reduction of accidents was recorded during the period of the Olympic Games. The number of fatal accidents in August was zero and the number of accidents with light injuries was four, i.e., the minimum value of every other month of the year 2004 (Papadimitriou, 2006).

The passage of buses from ETC lanes relieved the manual lanes from queuing, thus accelerating and securing accredited fleet movement from the tolls. Specifically, the ETC passage of the ATHOC vehicles, in line with the decision not to implement the height restriction in ETC lanes and the effective management of “convoy” traffic for the ceremonies, allowed for about 15,000 Olympic vehicles to pass quickly through the ETC lanes during the Games, increasing the average percentage of ETC transactions for that period to 45% compared to the 25% level before the Games (Papadimitriou, 2006).

Finally, persuading spectators, volunteers and workforce to use public transport and abstain from using private vehicles was of major importance and measures were taken to improve and to strengthen all available public transportation means. The overall plan was proven to be successful in reducing traffic demand, as the 320,000 average daily spectators and the 80,000 average daily workers and volunteers mostly used public transport (Papadimitriou, 2006).

CONCLUSION

The challenge of hosting a large and significant event such as the 2004 Summer Olympic Games was immense for the entire country and for the operation company of the Attica Tollway in particular. The sheer size of the event placed a huge load on the planning and preparation stages: there would be no room for mistakes after the Opening Ceremony because millions of people are set into motion and very little, if anything, could be changed or delayed to allow for covering of any omissions.

The internal evaluation of the entire Olympic Games Traffic Program was vital for AD and for the city of Athens at large. Negotiations, cooperation and daily meetings with other involved agencies and authorities resulted in managing certain incidents faster and more efficiently, and the AD personnel felt more confident sharing responsibility. The added value gained by this planning process manifests itself in the form of procedures that have remained and continue to be followed long after the Olympics, e.g., the procedures implemented for the passage of the VIPs and the presence of the Traffic Police at the Attica Tollway TMC is now permanently in place.

In conclusion, despite the many challenges presented, both on a technical and an organizational level, the Attica Tollway operator, with the aid and support of external and special entities, succeeded in overcoming the difficulties of hosting such a major global event. Most important, however, is the learning process that took place for both managers and employees of AD, who won a “gold metal” for handling such an intensive and prolonged situation successfully. To live the experience of such an event and to gain this unique knowledge becomes a powerful asset for any organization; moreover it is valuable know-how for the transport engineering community and a significant gain for the country at large. In light of the upcoming 2008 Olympic Games in Beijing, this experience can serve as a basis for the understanding of the issues regarding the preparation, management, and operation of Olympic Games urban transportation.

REFERENCES

Halkias, B., Papandreou K., Kopelias P., Prevedouros P. and Skabardonis, A. (2005). "Attica Tollway Management in the Athens 2004 Olympic Games." *Proceedings 12th World Congress on Intelligent Transportation Systems,* San Francisco, CD-ROM.

ATHOC – Athens Olympic Organizing Committee (2005) “Official report of the Games of the XXVIII Olympiad”, *Games Services – Transport*.

Papadimitriou, F., Thibaut, F., (2006). “Managing Operations for Large Scale Events: The Case of the Attica Tollway during the 2004 Olympic Games.” *Proceedings 1st International Symposium on Freeway & Tollway Operations,* Athens Greece, CD-ROM.

Model Development for Comprehensive Evaluation of Freeway Maintenance Quality

Guoqiang Zhang[1], Jian Lu[2], Qiaojun Xiang[3] and David Peshkin[4]

[1]Department of Traffic Engineering, Transportation College, Southeast University, 2 Si-Pai Lou, Nanjing City, Jiangsu Province, 210096, China, PH(86)25-83790776; FAX(86)25-83795352; email: flyingfeather_zhang@yahoo.com.cn
[2]Department of Traffic Engineering, Transportation College, Southeast University, 2 Si-Pai Lou, Nanjing City, Jiangsu Province, 210096, China, PH(86)25-83790776; FAX(86)25-83795352; email: jianjohnlu@sina.com.cn
[3]Department of Traffic Engineering, Transportation College, Southeast University, 2 Si-Pai Lou, Nanjing City, Jiangsu Province, 210096, China, PH(86)25-83790776; FAX(86)25-83795352; email: xqj@seu.edu.cn
[4]Applied Pavement Technology, Inc. Executive Towers West III, 1400 Opus Place, Suite 905, Downers Grove , IL 60515; email: dpeshkin@pavementsolutions.com

Abstract

Freeways are the backbone of a highway system and they are designed to provide users with high standards of service. In order to maintain a high level of freeway service, adequate and appropriate maintenance should be routinely performed. Determining the need for, and the appropriate type of, routine maintenance may be accomplished with systematically performed, comprehensive evaluations of freeway maintenance quality. This paper summarizes a research project sponsored by the Ministry of Communications of China and the Highway Bureau of Jiangsu Province, which was to develop comprehensive evaluation procedures and models for freeway maintenance. The freeway in this system consists of six primary subsystems: pavement, bridge, subgrade, on-road facility, culvert, and roadside vegetation. Based upon survey and research, comprehensive evaluation models for the integrated estimation of maintenance quality of the six subsystems are set up by way of linear regression. Analysis of the correlation between the subjective evaluation of survey crews and objective evaluation of the models is carried out, and the result shows that the models are valid and reasonable.

Introduction

The Department of Transportation has published several important specifications concerning highway and freeway maintenance, such as *Technical Specifications of Maintenance for Highway* (Department of Transportation, 1996) and *Expressway Maintenance Quality Evaluation Standards* (Department of Transportation, 2002).

In 2003, to further develop these specifications and to promote a more rational approach to freeway maintenance, a research project sponsored by Jiangsu Province Department of Transportation was initiated to develop comprehensive evaluation models for freeway maintenance quality. This paper mainly summarizes the model development for comprehensive evaluation of freeway maintenance quality. Details about the model development for the separate subsystems will be published in other papers.

Evaluation Framework

For the purposes of maintenance, freeways can be divided into distinct subsystems; six such subsystems are used in this project: pavement, bridge, subgrade, on-road facility, culvert, and roadside vegetation, and each of these subsystems is evaluated using the following methodologies. Pavement evaluation is based upon the following indexes: Pavement Condition Index, Riding Quality Index, Pavement Rut Depth Index, and Pavement Skid Resistance Index. Bridge evaluation is carried out by rating three components: the deck, superstructure, and substructure. The subgrade is evaluated by integrating the ratings for the shoulders, side slope, and drainage. On-road facilities include signs, markings, guardrails, separation fences, anti-glare installations, and emergency telephones: their evaluation is the integration of the ratings of each of those components. Culverts are rated by integrating a damage rating and drainage rating. Roadside vegetation is evaluated by integrating the ratings for median vegetation, slope embankment vegetation, and roadside vegetation.

The comprehensive evaluation of freeway maintenance quality is based upon the separate evaluation of maintenance quality of the six subsystems. The detailed discussion of the evaluation of the six freeway subsystems will be published in other papers; this paper mainly focuses on the unification of separate evaluations of the six subsystems into a comprehensive evaluation. Figure 1 shows the basic evaluation framework (concept).

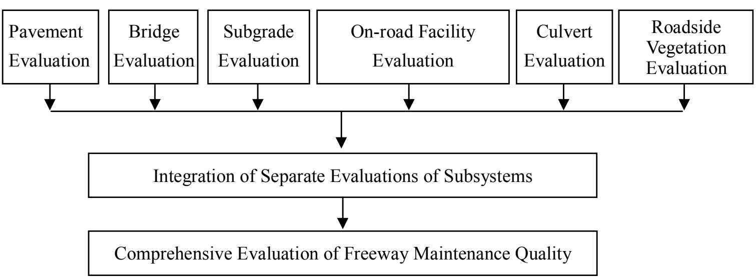


FIGURE1. Evaluation framework of freeway maintenance quality.

Evaluation Modeling Procedure

For practical reasons, freeways are divided into individual sections (in China, the length of a section is usually 1 kilometer.) and a comprehensive evaluation of freeway maintenance quality is carried out to evaluate the individual sections. It should be pointed out that although the freeway is divided into six subsystems, many of the sections do not contain all of the six subsystems. According to our field survey, there are four possible situations in all: freeway sections with all six subsystems; freeway sections with all subsystems except culvert; freeway sections with all subsystems except bridge; and freeway sections with only four subsystems: pavement, subgrade, on-road facility, and roadside vegetation. Therefore, four comprehensive evaluation models should be established separately to represent the four situations. A linearly weighted combination of separate estimations of subsystems is applied to formulate the comprehensive evaluation models, which are expressed as follows:

$$TQI = W_{PQI} \cdot PQI + W_{BQI} \cdot BQI + W_{SQI} \cdot SQI + W_{FQI} \cdot FQI + W_{CQI} \cdot CQI + W_{GQI} \cdot GQI \tag{1}$$

$$TQI = W_{PQI} \cdot PQI + W_{BQI} \cdot BQI + W_{SQI} \cdot SQI + W_{FQI} \cdot FQI + W_{GQI} \cdot GQI \tag{2}$$

$$TQI = W_{PQI} \cdot PQI + W_{SQI} \cdot SQI + W_{FQI} \cdot FQI + W_{CQI} \cdot CQI + W_{GQI} \cdot GQI \tag{3}$$

$$TQI = W_{PQI} \cdot PQI + W_{SQI} \cdot SQI + W_{FQI} \cdot FQI + W_{GQI} \cdot GQI \tag{4}$$

where *TQI* is the comprehensive evaluation of freeway maintenance quality; *PQI*, *BQI*, *SQI*, *FQI*, *CQI* and *GQI* separately stand for pavement maintenance quality index, bridge maintenance quality index, subgrade maintenance quality index, on-road facility maintenance quality index, culver maintenance quality index and roadside vegetation maintenance quality index, which are individual evaluations of maintenance quality of a freeway's six subsystems obtained through their maintenance quality evaluation models; W_{PQI} , W_{BQI} , W_{SQI}, W_{FQI} , W_{CQI} and W_{GQI} are corresponding weights to be estimated through the modeling procedure.

To calibrate the models shown in Equations 1 through 4, several approaches could be used. In the research, a multi-variable regression analysis method was used. Conceptually, the entire comprehensive evaluation procedure is to convert the indices to be obtained by those individual subsystem evaluation models into the overall rating *TQI,* through the application of the equations. The data to be obtained by the field panel and subsystem evaluation models are used to calibrate the models shown in Equations 1 through 4.

In order to calibrate the models shown in Equations 1 through 4, ratings which represent the subjective evaluation of the whole freeway system are needed. The subjective rating scores are used as dependent variables in the calibration to obtain the weights shown in Equations 1 through 4. To obtain the subjective ratings, a review panel consisting of a group of roadway maintenance engineers and users will be organized. Members of the panel will subjectively rate the freeway maintenance quality, considering maintenance quality of all the subsystems as a whole. The results from the panel's subjective rating will be used to calibrate the models as *TQI*. The indices *PQI*, *BQI*, *SQI*, *FQI*, *CQI* and *GQI* are obtained by calculation of corresponding maintenance quality evaluation models of the freeway's subsystems, which depend upon data collected from field survey engineers (called surveyors). Details of those evaluation models of the six subsystems will be discussed in other papers.

Field Data Collection

In this research project, the models specified in Equations 1 through 4 were developed with data collected in the field. Throughout 2004, 2005, and 2006, the research team collected field data for the development of maintenance quality evaluation models of freeway's subsystems and the comprehensive evaluation models of freeway as a whole. Members of the survey group were well trained with sufficient knowledge to collect all kinds of data, both subjective and objective, for the research project. Most of the field data were collected from eastern China's Jiangsu Province.

In order to calibrate the comprehensive evaluation models, subjective ratings for overall freeway maintenance quality were provided by a field panel (as discussed previously). The rating results from the panel were only based on the overall subjective perception for maintenance quality of freeway. On the other hand, the separate evaluation models of the six subsystems gave the indexes of maintenance quality of the six subsystems. The data related to subjective ratings were used as references (dependent variables) and the data obtained by the evaluation models of subsystems were used as independent variables for calibration purposes. Data from more than one hundred freeway sections were used, covering a wide range of different freeway maintenance qualities between *Poor* and *Excellent*.

Modeling Results

The main objective of the modeling procedure is to develop the models specified in Equations 1 to 4, or to obtain the parameters of the equations. The calibration procedure was really simple because it can be completed with the use of linear regression methods. The following equations represent the regression analysis results:

$$TQI = 0.37 \cdot PQI + 0.18 \cdot BQI + 0.14 \cdot SQI + 0.12 \cdot FQI + 0.06 \cdot CQI + 0.13 \cdot GQI \quad (5)$$

$$TQI = 0.4 \cdot PQI + 0.19 \cdot BQI + 0.14 \cdot SQI + 0.13 \cdot FQI + 0.14 \cdot GQI \quad (6)$$

$$TQI = 0.45 \cdot PQI + 0.17 \cdot SQI + 0.15 \cdot FQI + 0.07 \cdot CQI + 0.16 \cdot GQI \quad (7)$$

$$TQI = 0.49 \cdot PQI + 0.18 \cdot SQI + 0.16 \cdot FQI + 0.17 \cdot GQI \quad (8)$$

The above equations are the models to comprehensively estimate freeway maintenance quality based upon the estimate of the separate subsystems of freeway. The correlation coefficient values (R^2 values) for Equations 5 to 8 are 0.86, 0.90, 0.92 and 0.91 respectively. Since all the correlation coefficient values are high, it can be concluded that all these models correlate well with the subjective ratings.

Model Validation

Model validation is necessary for this research project. The validation procedure was based on the comparison between the evaluation from the models shown in Equations 5 to 8 and the evaluation performed by freeway survey crews. In fact, the evaluation by freeway survey crews was a subjective evaluation intended to give a subjective rating of the maintenance quality of a freeway section, and the subjective rating was then used as a reference to assess whether the model results were good enough or not. In this validation, 20 test sections of freeway were used. Two teams were organized, with the first team focusing on the subjective evaluation of freeway maintenance quality and the second team working on collecting freeway maintenance quality data to be input to the maintenance quality evaluation models of the six subsystems, whose evaluation results are to be used by Equations 5 to 8 for comprehensive evaluation of freeway maintenance quality.

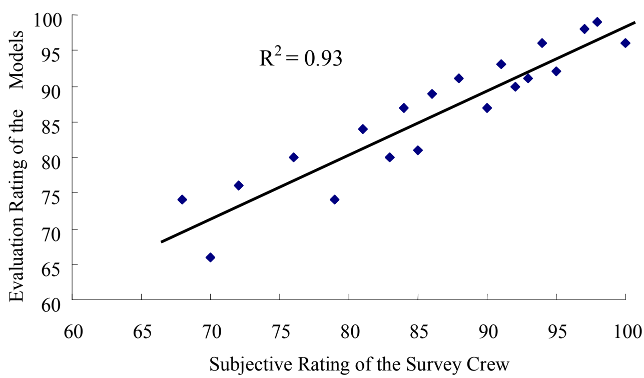


FIGURE 2. Comparison between subjective rating of the survey crew and evaluation rating of the models.

It can be seen from the validation results shown in Figure 2 that the correlation between the evaluation rating of the models and the subjective rating performed by freeway survey crews is very good. The coefficient of correlation (R^2 value) is larger than 0.9, indicating the models shown in Equations 5 to 8 are valid and reasonable (Frank E. Harrell, Jr. 2001).

Model Application

The models developed above can be applied to direct the maintenance activities of the freeway. Based on field survey results, freeway maintenance quality can be classified as *Excellent* (score = 90 to 100), *Good* (score = 80 to 89), *Fair* (score = 70 to 79), and *Poor* (score < 70). For freeways with a maintenance quality rating of *Excellent* and *Good*, there is no need to propose any special maintenance activities; regular or routine maintenance will be sufficient. For freeways with a rating of *Fair*, certain adjustments or minor additions in maintenance are needed. For freeways with a rating of *Poor*, which means some of the subsystems (such as pavement, bridge and roadside vegetation are not acceptable), major adjustments in maintenance are certainly needed.

Conclusions

Freeway systems are an important component of the transportation network, and users expect the highest quality of service on these systems. As such, the expenditures associated with the construction and operation of freeways usually are extraordinarily high. Because of their importance and the investment they represent, it is necessary that freeways be carefully and routinely surveyed to monitor that their overall quality is above a certain level. Comprehensive evaluation models of freeway maintenance quality can be developed to facilitate the maintenance activities of a freeway system.

With comprehensive evaluation models developed in the research project, freeway maintenance engineers are able to better or more objectively quantify the overall maintenance quality of freeways so that optimized maintenance options could be planned. However, in order to reasonably use the evaluation models, the models should be upgraded from time to time with the new field data collected.

References

Department of Transportation. (1996). *Technical Specifications of Maintenance for Highway* (JTJ073-96), People's Jiaotong Press. Beijing China.

Department of Transportation. (2002). *Expressway Maintenance Quality Evaluation Standards* (Test Version), People's Jiaotong Press. Beijing China.

Frank E. Harrell, Jr. (2001). *Regression Modeling Strategies: with applications to linear models, logistic regression, and survival analysis*, New York : Springer.

A Self-adjusting Urban Traffic Control

Chi Qi[1], Zhong-Sheng Hou[2], Aleksandar Stevanovic[3]

[1]Advanced Control Systems Lab, Beijing Jiaotong University,
ST 100044; PH (0086) 010-51684679; FAX (0086) 010-51682487;
e-mail: qichi1982@126.com
[2]Advanced Control Systems Lab, Beijing Jiaotong University,
ST 100044; PH (0086) 010-51682487; FAX (0086) 010-51682487;
e-mail: houzhongsheng@china.com
[3]Utah Traffic Lab, University of Utah, Salt Lake City, UT
ST 84112-0561; PH (801) 581-4151; FAX (801) 585-5860;
e-mail: aleks@trafficlab.utah.edu

Abstract: Controlling traffic in cities is a very complex task, since efficient traffic flows require coordination of service requests from many participants. Traditional methods optimize signal timing plans, to maximize efficiency of traffic, based on predefined traffic flows e.g. fixed-time signal timing plans. Disadvantages of these methods are that they cannot cope with constantly changing traffic flows. Therefore, fixed-time signal timing plans are not robust enough to provide near-to-optimal service to fluctuating traffic flows. We propose a simple and feasible method which adjusts traffic lights to better utilize phase green splits. This method is based on detection of traffic flows and represents a form of actuated traffic control with a flexible phase sequence. The initial results show that self-adjusting traffic control significantly reduces average vehicle delays when compared with conventional traffic control. The method also shows some promising potential for coordinated traffic operations on urban arterials.

Keywords: self-adjusting; traffic control; actuated control;

1. Introduction

With ever-increasing number of vehicles on urban networks, the conflict between transportation supply and demand becomes one of the vital issues for human mobility. Applications of Intelligent Transportation Systems (ITS), such as Advanced Traffic Management Systems (ATMS), are widely used to help reducing the gap between traffic supply and demand (Zhang and Peng, 2003).

In traditional area traffic control, signal timings such as phase splits, offsets, and cycle times are calculated for representative periods of diurnal traffic demands. Alternatively, on-line traffic control systems are used to adjust these signal timings to fluctuations in traffic flow (Liu, 2003). Both methods have disadvantages. The traditional fixed-time signal timing plans do not respond to traffic fluctuations. On the other hand, on-line traffic control is often too slow in detecting current changes in traffic flow. Hence, on-line traffic control may sometimes be few steps behind current traffic conditions (Duan, 2004).

In this paper, we present a method for the self-adjusting traffic control, which is based on local traffic detectors at each intersection. Our method utilizes both upstream and stop line detectors that report both vehicle arrivals and queue lengths at the respective intersection approaches. The method is based on a simple concept: the fewer number of vehicles arriving at an intersection approach the longer waiting time at the stop line. So, phase green durations are based on number of vehicles detected at each intersection approach. This method, when compared to the traditional methods, generates less average delay per vehicle and reduces computational needs.

2. Self-adjusting traffic control

When evaluating quality of traffic control regimes we use a set of performance measures, which are either system-based (travel times, stops along the route, etc.) or intersection-based (average vehicle delays, queue lengths, etc.). There is a strong relationship between queue length and an average vehicle delay at an intersection approach. As queue length grows the average vehicle delay is increased (Tan and Zheng, 2006).

Traditionally, most of the current traffic control methods use a fixed phase sequence, which does not always match actual traffic flows at an intersection. For example, if traffic demand for a phase is low (only few vehicles or a single vehicle) certain amount of green is dedicated for that phase in fixed phase sequence. This approach requires that some minimum green and some inter-green (amber + all red time) times are used, sometimes only to serve a single vehicle. Primary purpose of this approach is to provide service to each phase on the expense of efficiency of utilizing available green time. For example, it may be more beneficial if the green time used to serve only one vehicle was used for other phases with higher traffic demands.

The method of self-adjusting traffic control is based on flexible phase sequence and variable cycle length. In other words, detected traffic flows govern length of phase splits and phase sequence. This simple method self-adjusts green splits in the following way: if there is a single vehicle (or few vehicles) waiting for a phase green, these vehicles will wait longer. At the same time green will be available to a phase with the highest traffic demand. Traffic demand accumulates at the approaches which wait for green light. As more vehicles join the queue, the chances of getting green increase. Upstream detectors are used to detect vehicles before they reach the queues. With sufficient number of vehicles the green light will be provided even before this platoon reaches the intersection, thus generating “green waves” (Gershenson, 2005). Potentially, this method can be used to provide good coordination between all intersections within a region.

3. The single intersection self-adjustment control method

In a traffic signal cycle, an interval during which a right-of-way is given to traffic flow from certain approach is called a phase. Minimum number of phases to operate a traffic signal is two, while most of the traffic controllers currently use four or more phases (Xu et al., 2005).

For any type of traffic control length of green should be shorter for the movements with lighter traffic flows. For fixed-time control these green splits are determined based on average hourly traffic flows for each traffic movement. Alternatively, actuated traffic control detects number of vehicles arriving during green interval to extend the green interval between its predefined minimum and maximum values. Green extension is done in increments of Δ_t, where actuation of the relevant detector(s) is checked at the end of each interval Δ_t. There are two ways to end the green: Gap Out and Max Out. Gap Out means that all vehicles from the queue cleared the intersection and green interval ends a second or two after the last detector's check shows no vehicle actuation. Max Out means that the queue is so long that it cannot clear the intersection before green reaches its maximum time. The approach presented here is quite similar to the one used by actuated traffic control. However, the major difference between the actuated traffic control and the self-adjusting method is in the way they respond on detection calls for next phases. While actuated traffic control always serves a phase which has a placed detection call, the self-adjusting control reacts differently. The fact that a vehicle is waiting for a phase's green does not mean that the green light will be given in the self-adjusting control. Decision to provide the green light depends on how badly each of currently non-served phases needs the green. In other words, certain phase may be skipped several times before it accumulates enough traffic to get the highest priority when traffic control allocates the green lights.

Based on these principles, we suggest the following implementation for green splits within a single traffic signal cycle:

Step 1: Define maximum and minimum green times (G_{max} and G_{min}) as well as passage time (Δ_t) for each phase in sequence. Passage times should be based on the judgment of an experienced person who regulates traffic at intersections and geometric features of the intersection.

Step 2: Run each phase green at list until minimum green time (G_{min}).

Step 3: When minimum green time is reached check the upstream detector to see if there is a vehicle arriving. The upstream detectors should be installed approximately around 80 to 100 meters from the stop line.

Step 4: If there is a vehicle arriving, extend the phase green time for the length of passage time (Δ_t)

Step 5: At the end of current (extended) green time repeat Steps 3 and 4 until detector's check reports no vehicles arriving, then go to Step 6.

Step 6: Sum number of vehicles queued at each of the intersection approaches which did not get green light in this cycle.

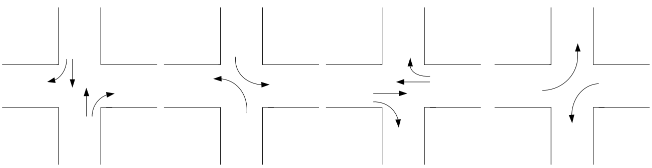
Step 7: Start the phase with highest number of vehicles waiting.

Step 8: Repeat Steps 2 to 7 until all phases in the sequence are run.

Step 9: Go to the next cycle and repeat the entire procedure.

4. Simulation Tests

The self-adjusting traffic control for a single intersection is compared in simulation with a traditional fixed-time timing plan. The essential part in the fixed-time plan for single intersection is to obtain the optimal timing plan according to the traffic flows and its distribution. By optimizing cycle length we provided adequate capacity for the intersection whereas by optimizing green splits we minimized average delay per vehicle. The fixed-time signal timings used in the simulation were: cycle length - 120s; green splits for the four phases shown in Figure 1- 30s, 34s, 26s, 30s, respectively. Figure 1 shows a traditional four-phase traffic signal design used in China, where arrows show protected traffic movements for each phase.



(a) Phase 1 (b) Phase 2 (c) Phase 3 (d) Phase 4

Figure 1 Four-phase signal control

Programs in Matlab were written to simulate both the self-adjusting method and conventional fixed-time signal control. As a test-case authors used a single four-leg intersection, with following traffic conditions:

- Each intersection approach has 3 lanes
- Maximum red - 300 s
- Minimum green - 15 s,
- Maximum green - 60 s
- Amber time - 3
- All red time - 1 s
- Vehicle arrivals are defined by Poisson distribution
- 20% of approaching vehicles turn left or right
- Saturation flow rate - 1800 vehicles/hour

- Upstream detectors are located 80 m from the intersection

To obtain statistically valid results we repeated 20-minute simulations 25 times. The results of our comparison are shown in Table 1. The results show that the self-adjustment control responds quickly by adjusting phase greens to fluctuations in traffic flows. During heavier and saturated traffic conditions the method reduces average vehicle delay and improves intersection capacity. Results also show that benefits of using the self-adjusting traffic control for light traffic conditions are similar to those of the traditional fixed-time traffic control.

Table 1 Average Vehicle Delays from Two Traffic Control Methods

East-West Average Traffic Flows (veh/h)	North-South Average Traffic Flows (veh/h)	Average vehicle delay (sec/vehicle)		
		Self-adjusting	Fixed-time	Difference
400	400	54.43	58.19	6.4 %
800	800	69.87	78.25	10.7 %
2000	2000	171.74	194.68	11.7 %

Conclusions

The self-adjusting traffic control presented in this paper uses flexible phase sequence. It extends green lights of the current phases as long as there is traffic demand. The self-adjusting traffic control decides which phase should be served next based on the traffic accumulated at the respective approach during the previous phases. The results show that the self-adjusting control brings significant benefits in average vehicle delay when compared with traditional fixed-time signal control. These benefits are higher for heavier traffic conditions which increases potential of using the self-adjusting traffic control for oversaturated traffic conditions.

5. References

Gershenson, C. (2005). “Self-organizing traffic lights.” *Complex Systems,* <http://uk.arxiv.org/abs/nlin.AO/0411066 > (Feb. 1, 2005).

Xu, L.H., Xi, L.A., and Zhong, L.S. (2005). “Adaptive multi—phase fuzzy control of single intersection based on neural network.” *China Journal of Highway And Transport.,* 18(3) 90-93. Chang'an University.

Duan, Q. (2004). “The Urban Traffic Lamp Control System with Self-Organization.” *Journal of South-Central University for Nationalities* (Nat. Sci. Edition)., 23(3) 64-67 South-Central University.

Tan, W.X., and Zheng, Y.W. (2006). “Fuzzy control for urban isolated intersection traffic lights.” *Journal of Fuzhou University.,* 34(1) 71-75. Fuzhou University.

Zhang, G.W., and Peng, H.Q. (2003). *“The system engineering of the intelligence* *transportation introduction."* The Electronics Industrial Press, Beijing.

Liu, Z.Y. (2003). *"The intelligence transportation control theories and application."* The Science Press, Beijing.

Downtown APM Circulator: A Potential Stimulator for Economic Development in Newark, New Jersey

RONGFANG (RACHEL) LIU, PH.D. AICP, PE
NEW JERSEY INSTITUTE OF TECHNOLOGY

CHOIKWAN (SHIRLEY) LAU
ADVANTEC CONSULTING ENGINEERS, INC.

ABSTRACT:
After several decades of decay and decline, the central cities in America have finally caught the attention of governments, the private sector, and ordinary citizens. There are numerous programs that entice businesses to relocate or stay in these areas. The programs also aim to improve the occupational and academic skills of local residents, encourage the creation and retention of new jobs, and encourage entrepreneurship and the formation of new businesses. However, congesti
on and parking problems continue to impede the progress of such downtown revitalization and economic improvement efforts.

This manuscript documents a feasibility study of an Automatic People Mover (APM) application for the City of Newark in New Jersey. The objective of this feasibility study is to evaluate the applicability of the APM technology for increasing mobility, efficiency, and accessibility, and ultimately for enhancing transit service quality. The potential benefit of replacing the current LOOP, a small bus circulation system in downtown Newark, with APM technology, will not only reduce passenger walking, waiting, and overall travel time, but will also provide an anchor for downtown business development, which in turn will have great potential to stimulate economic development in the city of Newark and surrounding areas.

KEY WORDS:
Economic Development Stimulator, Automatic People Mover, Downtown Circulation, Driverless Transit, downtown revitalization, activity center

INTRODUCTION

Living in the 21st century, Automated People Movers (APM) are no longer limited to airport use, but also stretched its potential to recreational parks and central business districts. In the U.S, APM has been adapted by many major international airports and amusement parks, as well as central business districts (CBD) of major cities (Panayotova, 2003).

Defined by its critical location, rich traditions of all American ethnicity, and natural, cultural and human resources, Newark is clearly emerging at the top of the wave of downtown revitalization processes. Economic stimulation is one of the main themes to be explored by many downtown redevelopment processes; this particular study explores the possibility of building an Automated People Mover in downtown Newark, which will not only replace the existing bus LOOP services, but will also work as an economic stimulator to the CBD and the city.

APM TECHNOLOGY

As defined by the Transportation Research Board Committee APO40: Major Activity Circulation Systems (TRB, 2006), the APM is a driverless, electric-powered, automated vehicle operating in a single or multi-car train on steel or concrete guideways, monorail, or dual track systems. Over the last two decades, the APM has become one of the most significant developments in transit industry. The APM system has become the standard solution in large, complex airports due to its capability to move passengers fast while minimizing waiting and walking time and to provide high service quality (Shen, 1992).

Historical Development

In late 1970, the Urban Mass Transportation Administration (UMTA) has signed a contract with West Virginia University to construct the first automated people mover in the United States *(Schneider, 1999)*. In 1976, the UMTA has selected proposals for APM on-site inspections from Los Angeles, St. Paul, Minnesota, Cleveland, Houston, Miami, Detroit and Baltimore, of which only Detroit and Miami actually built.

Typical APM Features

Used in many airports, theme parks, and downtown circulations, APMs may have as short as 60 second headways and are capable of moving 2,000 to 25,000 passengers/hour per direction. The construction of tracks can be elevated above the ground. The duration of the ride is usually short; a typical trip length is between 3-6 miles, and the travel speed is typically between 35-60 mph, with smooth acceleration & deceleration. The system may be configured in an on-demand mode during off-peak hours to minimize energy consumption.

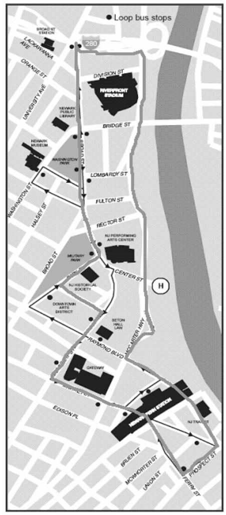
APM Applications

There have been over fifty airport APM applications worldwide due to its advantages, such as high ride quality, short headways, flexibility in operation, excellent reliability and safety records (Shen, no date). APM technology has also been used for trunk link transit service, such as the SkyTrain in Vancouver, Canada, and the VAL system in Lille, France, both of which are significantly successful. Today, one sixth of all transit passengers in the Vancouver region use the SkyTrain for all or part of their daily trip. The VAL system in Lille, France is also an example of a successful line haul APM application in Europe.

POTENTIAL APM APPLICATION IN NEWARK, NJ

Located in northern New Jersey, Newark is the largest city in the state with a population of 273,546 (US Bureau of Census, 2003). The city of Newark is not only a gateway and a major financial and commercial center to the State of New Jersey, but it also serves as a transportation hub due to its accessibility in the transportation network. The city is surrounded by major interstate highways (such as I-78, the New Jersey Turnpike, and I-280), local routes (Route 21, US 1 & 9)/roadways, and it is also near one of the busiest international airports in the U.S. Newark Liberty International Airport. In addition, Newark has a high population density with a high usage of public transit, such as trains (Pennsylvania Station), buses, cabs and subways. Therefore, the city generates heavy traffic volume, especially during peak hours as well as high pedestrian flow during the day. Due to its CBD area type, there is limited available land for roadway expansion when it comes to problem solving of traffic congestion.

Besides the commuter trains from New Jersey suburbs to New York City, there are a large number of buses that make up the public transportation network in North Jersey, especially in Newark. According to New Jersey Transit (NJ TRANSIT), the public transportation agency in New Jersey, about 50 buses per hour (with 6 different routes) operate along Broad Street & Market Street, the downtown or central business district of Newark.



There is currently a downtown circulation service in Newark called the “LOOP”, which is a minibus shuttle operated by “Red and Tan Tours,” a Coach USA Company under contract with NJ Transit. It provides a connection between Newark Penn Station and Newark Broad Street Station during morning and afternoon peak hour periods. The service offers circulation throughout downtown Newark with stops at Broad Street Station, Newark Museum, NJPAC, Seton Hall Law School, Gateway Center, Penn Station and the Ironbound, as demonstrated in Figure 1.

Limitations of the LOOP

The LOOP service has provided connections to people who work in downtown Newark or commute to and from downtown locations. However, due to the small capacity of the minibuses, limited operating schedule, and low visibility of bus services, the current LOOP service faced major challenges in meeting the true demand of downtown development in Newark.

The shuttle service often experiences delays since it runs with mixed traffic in downtown Newark, where congestion occurs frequently. Several stops are located in middle of a roadway segment with standing and parked cars. When passengers enter or leave the shuttle, the traffic flow could be interrupted by blocking other vehicles in the lane. This would increase the accident rate when drivers tried to shift lanes. Some other stops are located at corners or intersections before or after the traffic signal. This could also interrupt the traffic flow if the signal was green while the shuttle stopped to drop off passengers.

A high pedestrian flow is expected in a CBD area during peak hour periods when people get to and off from work. It can be seen not only on major roads such as Broad Street, Market Street, Raymond Boulevard and Mulberry Street. However, it is more apparent in a pedestrian-oriented community – the Ironbound. The roadway merges with Raymond Boulevard and Market Street after Penn Station in Ironbound, and has one lane in each direction with parked cars on each side. There are running and stopping buses and pedestrians crossing. Traffic conditions can be intense and affect overall driver and pedestrian safety, especially when weather becomes severe.

Encouraging Environment for APM Applications

According to the City of Newark (2004), one of the objectives of the Master Plan is to strengthen and improve downtown Newark as a regional office center, with private market retail, cultural and entertainment uses, with hotels and dense residential developments. Vertical mixed use of pedestrian-oriented retail and ground floor services, as well as open spaces and parks, is encouraged.

The city has implemented its Master Plan with an additional element in land use planning and it is expected to improve the way that the city looks with new residential, commercial, entertainment, sports and recreation developments. According to the Master Plan, the goal and recommendation of the plan is to "reinforce and capitalize on Newark's existing physical assets, including its historical areas, waterfront, major new developments and Newark Liberty International Airport. Finally, the plan specifically recommends that future development and redevelopment respect and repair the city's street grid, which provides the best means of accommodating both development and transportation needs."

Various entertaining, cultural, and service establishments in downtown Newark will have significant impact on improving the downtown image. They can bring additional retail revenue to the city by offering night -time and weekend entertainment activities. As shown in Figure 2, the New Jersey Performing Arts Center, the Newark Bears and Eagles Riverfront minor league baseball stadium, an arena for the NJ Nets National Basketball Association franchise, and an additional symphony concert hall are all within walking distance of the existing LOOP route.

Recent development and proposed downtown improvement projects will bring a significant amount of employment, entertainment, and retail activities into downtown Newark. Given the existing pressure on the LOOP system and anticipated growth in downtown Newark, it is reasonable to plan some dramatic increase in circulation capacities. What better application can you find than the APM system in Newark?

FEASIBILITY OF AN APM IN DOWNTOWN NEWARK

The objectives of installing an APM system in downtown Newark are three-fold: First, an APM circulation system may improve mobility in downtown Newark via a dedicated guideway of APM systems. Second, APM circulation will provide better connections to various sites by offering more frequent services, extended operating hours, and a more comfortable ride. Third, the permanent investment of APM circulation and its high visibility will create a better environment to stimulate economic growth.

The challenges of implementing an APM circulation system in downtown Newark go hand in hand with the benefits it brings. However, the most significant challenges may include the negative experiences or perceptions experienced by various central cities and difficulties to raise the funds for the capital cost of APM systems.

The encouraging news is that a number of international cities in Asia, South America, and Europe have successfully implemented an APM circulation application in dense downtown areas. For example, Indonesia has successfully gathered private equity to fund an entire APN circulation project in Jakarta, and the circulator has continuously operated for more than 15 years without any major incidents. Daily operational expenses can be covered solely from revenues, with no public financing required. This financing applies to direct operations and maintenance expenses only. Although revenues could cover capital costs, this was not an objective of the Jakarta project

Benefits of the APM Applications in Newark

Replacing the LOOP in Downtown Newark with an APM service, as shown in green color of Figure 1, can provide a quiet, fast, efficient and environmental-friendly travel mode choice for the public. It will also attract more transit ridership for a city that has a diverse population group by imposing a minimal physical environmental impact on existing infrastructure. The system can also implement the connectivity for downtown district and nearby recreation areas, such as the riverfront stadium, arena, concentration for shoppers and workers, and direct street access, in order to offer a sense of community and improve quality of life for the surrounding neighborhood. In addition, an APM can bring a more balanced mode split to the city by accommodating walking and bicycling in a friendlier pedestrian environment, driving with less congestion while lowering the automobile use (minimizing pollution), promoting transit accessibility by sharing with other modes, and providing adequate access for handicapped users.

The APM proposal for downtown Newark stems from the need to understand and properly address the unique requirements of different socioeconomic groups *(Philips, 2002)*. The decision makers should make better transportation decisions that meet the needs of all people. The planning and engineering profession should design transportation facilities that fit more harmoniously into communities and enhance the public involvement process, strengthen community-based partnerships providing minority and low-income populations with opportunities to learn about and improve the quality and usefulness of transportation in their lives.

Installation of the APM system in downtown Newark will increase the service capacity to satisfy the travel demand of downtown population. In addition to assisting the people who work or travel en route to downtown Newark, the newly installed APM system will also serve the travel needs of downtown dwellers, who are mostly low income or minority groups.

Theater goers and sports fans can go to the New Jersey Performing Arts Center (NJPAC) or to the stadiums to participate in entertaining or sports events via APM. Travelers may stay in downtown Newark without renting vehicles since they would be able to get around among various business, entertainment, and service venues via the APM circulation.

Investment Returns

As an investment, the use of an APM application can as well correlate with the city of Newark's future land development plan. According to the city's Master

Plan: future airport-related hotel and other retail services will have to relocate elsewhere in the city in the long run due to the limited land availability in Newark Airport's northern and western border area (current hotel district), and the demand for downtown office space rentals will increase.. Newark downtown has been seen as a "developing area."

In accordance with the city's future development, existing land and transportation availability, an alternate long-term solution should be considered in order to relieve the growing traffic volume and population density without traditional roadway expansion. An APM system can be used to improve land use configuration and parking adjustment by creating stronger pedestrian focus in the center core area. That allows the city to stay more focused for higher density centers by reducing the use of cars, and that makes the city area more accessible by improving pedestrian environments for comfortable shopping trips by foot without the need to increase on-street parking capacity.

When compared to the light rail or other rail transit, an APM requires much smaller guideway and station dimensions that can provide sharper turns and steeper grades. APM stations (with 1500 feet or 500 meters comfortable distance) are more closely connected to each other than light rails (3300 feet or 1 kilometer) or rapid rail transit (1-2miles or 3 kilometers). Physical roadway geometric change would be minimal when the system is designed as "elevated."

Overall, the installation of APM services may enhance and strengthen the region's center for commercial services, such as hotel, entertainment, cultural, educational and institutional activities. It will promote interconnected community development. It will also maintain and enhance its traditional role as a transportation hub in the metro region for people and goods.

REFERENCES

City of Newark, New Jersey. "Newark Master Plan." 2006.
New Jersey Devils. "City Unveils Model of Newark/Devils Arena."
http://newjerseydevils.com/2005/html/theteam/teamnews/newarkarena040705.ph

Panayotova, T. "People Movers: Systems and Case Studies." Facilities Planning and Construction UFL, May 28th, 2003.

Schneider, J. B. "A Brief History of UMTA's Downtown People Mover Program."
http://faculty.washington.edu/~jbs/

Shen, L. D. "Implications of Automated People Movers for Airport Terminal Configurations." ITE Journal, February 1992.

Transportation Research Board, 2006."APO40 Major Activity center circulation committee" "http://gulliver.trb.org/directory/comm_detail.asp?id=1168.

The ASCE Automated People Mover Standard: Codifying Best Practices

P.V. Didrikson[1]

[1]Director, Marketing & Product Planning, Bombardier Total Transit Systems, P.O. Box 220, Station A, Kingston, Ontario, Canada K7M 6R2; PH (613) 384-3103 x 4212; FAX (613) 634-5321; email: paul.didrikson@ca.transport.bombardier.com

Abstract

Since their inception in the U.S. in the 1970s, automated people mover systems have become numerous, with installations around the world. Because APMs are inherently complex systems that involve multiple interacting subsystems, new technology and public safety, it is essential to establish minimum standards for their design, construction, operation and maintenance. The paper describes the benefits of standardization to organizations that specify and procure APM systems, regulatory authorities, system suppliers, system operators, system users and the general public. The consensus approach to standard writing is described and evaluated in terms of fairness, objectivity and effectiveness. Balance in APM Standards Committee membership and openness and transparency of the standards creation process are evaluated. The scope of the APM standard and potential overlap with other related standards produced within other standards-developing organizations is discussed. The degree to which a standard can gain worldwide acceptance, having been developed by an American organization with global membership, is also considered.

Introduction

The first automated people mover (APM) system in the world was installed at Tampa International Airport in 1971. Since that time there have been multiple installations in North America, Europe, the Middle East and Asia. Today there are (by our count) 67 APM systems in operation world-wide, with another 30 under construction. (These numbers include all driverless people movers and transit systems including driverless steel-wheeled systems). Clearly the APM is here to stay.

The Need for Standardization

Automated people movers are inherently complex systems that involve new technology, system automation and public safety; therefore there is a need to ensure that such systems are designed, manufactured, installed, tested, operated and maintained in accordance with strict standards.

A large portion of the cost of a modern APM system arises from systems engineering, documentation, safety, test and quality control requirements that are not necessarily obvious or evident in the finished product. As one example of such costs, the validation and verification process for the software that performs the automatic train protection (ATP) function in an automatic train control (ATC) system requires exhaustive analysis and testing of all ATP functions to ensure safety under all foreseeable conditions. This verification process alone can consume 50% of the resources deployed during development of safety-critical software (Bhansali 1997).

Clearly it would not be in the public interest to allow negligent or unscrupulous suppliers to produce substandard APM systems that did not include this type of safety and performance analysis and testing, and hence were not safe or did not deliver the required performance. There may be other approaches that can avoid such undesirable situations, but the creation of consensus industry standards (such as the ASCE APM Standard) and their enforcement with appropriate contractual and quality assurance measures is a method that has proven to be effective.

Scope of the APM Standard

The APM Standard ASCE 21 is intended to provide a minimum set of requirements for the design, system verification, demonstration, operation and maintenance of an automated people mover (APM) with an acceptable level of safety and performance. Consultants, transit authorities and other organizations who write specifications for the procurement of APM systems are free to specify design or performance characteristics above and beyond those required by the APM Standard, but should be aware that such requirements will probably add cost and may or may not provide corresponding benefits. Any organization that would procure an APM system without reference to the APM Standard, specifying instead requirements less demanding than those in the Standard, should be aware that the safety and/or performance of the system may be compromised.

Benefits of Standardization

Standardization Minimizes Duplication of Effort

In the absence of an accepted standard, organizations wishing to procure an APM system would be required to develop their own specifications and/or hire specialist consultants to produce such specifications. If multiple organizations followed this approach, inevitably there would be a considerable amount of duplicate effort and cost involved in producing these redundant specifications.

Standardization results in Efficiencies and Reduces Cost

When suppliers are able to standardize their products and minimize design changes from one project to the next, the cost of redesign is avoided and the savings can be passed on to buyers. In the absence of such standardization, suppliers would be faced with divergent requirements from multiple buyers, requiring a large engineering effort for each project, with the result that costs would be driven upwards.

Standardization Facilitates New Product Development

When a standard such as the ASCE APM Standard becomes widely adopted throughout the industry, it provides system suppliers with a stable set of minimum performance requirements that they can use as a benchmark or baseline. This allows them to proceed with confidence to develop new products that can meet or exceed the baseline requirements, relatively free from concerns that unexpected new specification requirements will make their new design non-competitive or non-marketable.

Standardization Ensures Safety and (at Least) a Minimum Level of Performance

An organization that references the ASCE APM Standard in its procurement documents for an APM system receives the benefit of the expertise of a diverse group of professionals who have been intimately involved in the design, manufacture, commissioning, operation and maintenance of such systems for (in some cases) more than thirty years. The combined relevant experience of the committee is easily in the hundreds of years.

The APM Standard has been developed using a rigorous and transparent process, ensuring that APM systems designed in accordance with the Standard will be safe and provide a level of performance that will be acceptable to the vast majority of users. The APM Standard is intended to ensure that the public interest is preserved.

If, on the other hand, an organization develops its own set of specifications for an APM system in isolation, the results achieved will be completely dependent upon the capabilities of the organization's specification writers, consultants, and quality assurance personnel without the benefit of the experience and expertise collected in the APM Standard.

Standards Facilitate the use of Performance Specifications

In producing a specification for a complex system such as an APM, the organization intending to procure such a system has the choice of two general approaches:

(a) creation of a detailed, prescriptive specification of the design and construction of all system components, including their installation, interconnection, test and commissioning; or

(b) creation of a performance specification that defines the required system performance but leaves the details of the system design and implementation to a turnkey supplier.

In case (a), production of the detailed, prescriptive specification will require a very large engineering effort, which will be costly whether this effort is performed by in-house engineers or by consultants. The resulting detailed specifications are likely to be voluminous, but may not result in an optimum system solution, for the following reasons:

- ***Detailed specifications may exclude competent suppliers:*** By defining the details of component or subsystem design and construction, the specifications may seriously limit the number of APM system suppliers who can respond with competitive bids. Faced with detailed requirements that may require redesign, a supplier may decline to submit a bid. Thus APM System designs that are equally capable as the solution specified, but which do not conform to the detailed, prescriptive solution may be excluded.
- ***Detailed specifications may increase costs:*** Detailed specifications may also increase costs. If suppliers are required to redesign their vehicles, their ATC systems, or any component of their APM systems, such redesign will incur cost but will not necessarily provide a corresponding benefit.
- ***There may be unintended consequences:*** If the supplier has designed his APM system using a rigorous process of systems engineering, changes to the details of the design in an effort to conform to a detailed specification may have negative unintended consequences.
- ***Detailed specifications may deter systems engineering:*** In a purchaser's effort to produce detailed specifications, the systems engineering effort required to ensure that the system will meet system performance requirements in an optimum and cost-effective manner may be lost. The principles of systems engineering require a top-down approach to system design that may tend to be overlooked when the primary objective is seen as the production of detailed specifications. When numerous specialists are assigned to write detailed specifications, they may concentrate on the detailed requirements of their assigned subsystem to the detriment of the overall system design and integration. Interfaces may be incompatible, and solutions that appear to minimize cost or maximize performance at a subsystem level may not produce an optimum overall system design.

Option (b), the performance specification, has the great advantage of clearly stating the desired system performance, but leaving the details of how such performance is achieved to a turnkey system supplier. The supplier is able to use his best existing designs or new developments to meet the performance requirements in the most cost-effective manner. Thus the process of system optimization takes place at the APM system level, which is where it is most effective.

The ASCE APM Standard is written with system performance optimization and systems engineering principles clearly in mind. It specifies methods of calculating system-level performance criteria such as System Availability, and includes requirements such as System Hazard Analyses, which must by definition be performed at the APM system level. The APM Standard facilitates the production of performance specifications by defining a set of minimum requirements that are generally applicable to APM systems and accepted in the industry, with the result that an author of a performance specification need only specify the characteristics that are unique to the specific application at hand.

By including references to the ASCE APM Standard and other similar internationally accepted standards, it is quite possible to create a twenty page performance specification for an APM system that is clear, concise, comprehensive, and which defines a state-of-the-art system. By avoiding the specification of details that are unnecessary and possibly counterproductive, such a specification will also tend to result in minimum cost.

To the extent possible, the requirements of the ASCE APM Standard are technology-independent. The intent is to allow diverse technical solutions while prescribing the essential safety and performance characteristics. Thus it is possible to design and supply an APM system that meets the requirements of the Standard using several different technologies, whether they use conventional rubber-tired vehicles, cable-pulled vehicles, or magnetic levitation or steel-wheeled vehicles powered by linear induction motors.

Fairness and Objectivity in the Consensus Approach to Standards Writing

The ASCE Codes and Standards Committee manages the consensus process by which standards such as the APM Standard are produced. The consensus process includes safeguards such as balloting by a balanced Standards Committee comprised of both members and non-members of ASCE, balloting by membership of ASCE membership as a whole, and balloting by the public. The process is designed to produce useful, competent standards and prevent undue influence from any special interest group.

It is of interest to note that since 1998 the U.S. has been a signatory to the World Trade Organization (WTO) *Agreement on Technical Barriers to Trade*, which obliges WTO members to ensure that regulations, voluntary standards, and conformity assessment procedures do not create unnecessary obstacles to trade. The WTO is responsible for ensuring that central government standardizing bodies comply with a *Code of Good Practice for the Preparation, Adoption and Application of Standards* and take measures to ensure that voluntary standardizing bodies do the same (McGean 2005).

Balance in APM Standards Committee Membership

The ASCE APM Standards Committee is a volunteer organization whose members are not paid by ASCE. Some members are paid by their employers to participate in the Committee’s standards-producing process (and may therefore be expected to reflect the views of their employers) but Committee membership is balanced to ensure that no single member or group can sway decisions in an inappropriate manner to the detriment of the public interest.

To this end, ASCE has four membership categories for its committees: Consumers, Producers, Regulatory, and General Interest. Consultants are placed into one of these groupings based upon the group for whom they generally spend most of their time working (usually the Consumer group).

The Committee presently has 67 members and is in balance as defined by ASCE rules, with 38.8% of members in the Consumer group, 37.3% in the Producer group, 6.0%

in the Regulatory group and 17.9% in the General Interest group. When the Committee membership reaches a maximum limit in one of the categories, new applicants for membership who fall in that category are put on a wait list.

Membership in the ASCE APM Standards committee includes representatives from all major APM system suppliers worldwide, and members in other categories from North America, Europe and Asia. Approximately 15% of the Committee membership is from EU nations. The Committee holds three meetings per year, with one meeting each year held outside of the U.S. (e.g. 2006, Singapore; 2007, Vienna).

Potential Overlap with other Standards

The ASCE APM Standard is presently the only generally accepted international standard that covers complete APM systems, although there are existing standards that address parts of such systems, such as communications-based train control (CBTC). The APM Standard avoids overlap in this case by making reference to IEEE P1474.1-2004 and stating that the IEEE standard shall apply for the CBTC portion of APM systems that use CBTC, in lieu of the corresponding sections of the APM Standard.

The IEC formed Technical Committee TC9 Working Group 39 in 2001 to develop an international safety standard for automated urban guided transport (AUGT) systems. Working Group 39 was reorganized as an ad hoc group in the spring of 2006. When the ad hoc group completes its work and the IEC publishes this standard, there will be some overlap between the IEC and ASCE standards, although an effort is underway to coordinate between the two standards groups to avoid inconsistencies.

International Acceptance of the APM Standard

The ASCE APM Standards Committee is certified as a Standards Development Organization by the American National Standards Institute (ANSI), however, given that the APM Standard is produced by an American organization and contains references to documents produced by other U.S.-based organizations such as IEEE, NFPA and UL (as well as references to U.S. military standards), it is legitimate to ask whether it is, in fact, an international standard accepted outside of the U.S. Indications of such acceptance include the fact that compliance with the APM Standard has been required in several international procurements for APM systems in Europe, Asia and the Middle East. Also, ASCE membership is global. The APM Standard is thus a “de facto” international standard, although it is not an official international standard under the ISO/IEC process (McGean 2005).

References

Bhansali, P.V. (1997). “Perspectives on Safety-Critical Software” Proceedings of the Australian Software Engineering Conference (ASWEC ’97), Publication 0-8186-8081-4/97, IEEE

McGean, Thomas J. (2005). “The International Standards Process – Implications for the Transit Industry” International Best Practices Workshop, Orlando, Florida Jan 27-28, 2005. Study funded by TCRP J6 (48)

Evaluation of the Railroad Crossing Wayside Horn

R. Jafari[1], J. H. Hummer[2]

Abstract

One potential solution for the problem of annoying locomotive horn sound is a stationary horn mounted at the crossing. This "wayside horn" is sounded in place of the locomotive horn when a train approaches and is positioned to direct the sound precisely down the intersecting roadways rather than along the track. A wayside horn can therefore operate at a lower sound level than a locomotive horn and produce less area sound exposure.

The objective of this project was to evaluate a wayside horn produced by Railroad Controls Limited. We conducted the evaluation through observation of a test installation in Rocky Mount before and after wayside horn installation. Before wayside horn installation, the site had a typical array of safety devices (gates, lights, signs, and marking). The site was a nearly ideal crossing of a road with one through lane in each direction of a single track with low train volumes and speeds in a moderate density suburban area. We examined the reliability of the system and also measured sound in the area, motorist behavior, motorist opinion, area resident opinion, and train engineer opinion.

Based on the results from previous studies and the results from our test, the study team concluded that the wayside horn offers significant sound relief to residents and others in the area around a crossing. The team also concluded that the wayside horn has led to slight, if any, shifts in driver behavior and opinion. Finally, the study team concluded that the wayside horn appears to be reliable and acceptable to train engineers. The team strongly recommends that the NCDOT, other relevant agencies, and railroads continue to allow wayside horns and begin preparing policies and procedures to decide where they should be installed and who should pay for them.

Introduction

Federal Railroad Administration (FRA) rules require that locomotive horns be sounded at most public highway-rail intersections (grade crossings) in the United States. This is due to the incontrovertible evidence showing that horns are an effective safety device preventing collisions at crossings. However, the rule also provides an opportunity for any community to establish a "quiet zone" in which the locomotive horn would not be sounded at crossings where supplementary safety measures (SSMs) fully compensate for the absence of the warning provided by the locomotive horn (1). Numerous communities may seek to implement such quiet zones, including several in North Carolina (2). In fact, New Bern and Rocky Mount in NC have had horn bans in their downtowns for years (2). However, the cost of the SSMs the FRA rules require in quiet zones may be prohibitive for some communities. For example, the installation costs for crossing gates could reach $250,000 with

[1] Chas. H. Sells, Inc, 15401 Weston Parkway, Cary, NC 27513; PH (919) 604-0808; FAX (919) 678 0202; e-mail: rjafari@chashsells.com

[2] North Carolina State University, Department of Civil, Construction, and Environmental Engineering, 423 Mann Hall; PH (919) 515-7733; FAX (919) 515-7908; e-mail: hummer@eos.ncsu.edu

yearly maintenance costs of $5,000. In view of these costs it may not be practical to establish or maintain quiet zones in most communities that may desire them.

The major reason for establishing a quiet zone is to eliminate the disruptive and intrusive affect of the locomotive horns on the community that has developed around the tracks. Federal regulations require that the horn volume be sufficient to reach motorists on roadways perpendicular to the train, far enough in advance for the motorist to be able to stop before reaching the crossing. The horn must reach down intersecting roadways, penetrate the automobile, and compete with other internal and external audio sources in order to alert the motorist of the train's approach. In order to do this, the horn must be loud and the horn must be used well prior to the train reaching the crossing, allowing the sound to spread into the surrounding community and impacting residents outside their vehicles.

Objective

The objective of this research project was to evaluate the Automated Horn System (AHS) or wayside horn; a stationary horn mounted at the crossing produced by Railroad Controls Limited. We conducted the evaluation through observation of a test installation in Rocky Mount, NC before and after AHS installation. Before AHS installation, the site had a typical array of safety devices (gates, lights, signs, and markings) and employed the locomotive horn in typical fashion. After AHS installation the array of devices remained mostly unchanged except for the installation of a median barrier, but the wayside horn replaced the locomotive horn. During the evaluation, we examined the mechanical reliability of the system and also measured: Sound level, Motorist behavior, Motorist opinion, Area resident opinion, and Train engineer opinion.

The results from the evaluation will show whether the wayside horn should be considered as an alternative to a quiet zone. While not cheap, the AHS is less expensive than some of the supplemental safety devices needed for a quiet zone. If the wayside horn proves reliable, proves to be as safe as a locomotive horn, reduces sound as anticipated, and makes nearby residents happier, it could be an attractive alternative for the NCDOT, other DOTs, and local communities.

Literature Review

Prior to this project, there have been six major efforts to evaluate the effectiveness of wayside horns. Table 1 summarizes the work done in each of those studies. Note that the methodology we followed in this test was more comprehensive than the previous studies, in that we looked at all five dimensions listed. However, everything we analyzed had been examined in at least one of these previous studies.

Table 1 -Previous evaluations of wayside horn

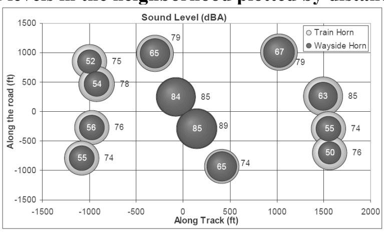
				Did the authors analyze data on…				
Ref. no.	Year of study	City	No. of crossings	Noise	Motorist behavior	Motorist opinion	Resident opinion	engineer opinion
3	1995	Gering, NE	1	√	√		√	
4	2000	Gering, NE	1		√			
5	1998	Ames, IA	3	√		√	√	√
6	2001	Richardson, TX	1	√				
7	2002	Mundelein, IL	3	√	√		√	√
8	2003	Roseville, CA	2	√	√		√	√

It is easy to summarize the findings from the previous six studies, because they are so uniformly positive. In particular: a) The AHS-focused sound radiation minimizes community intrusion. b) Community responses to the AHS and changing sound pattern have been favorable. c) The AHS is viewed favorably by both motorists and train engineers. d) There is no evidence that the AHS is less safe than the current practice of using train-mounted horns.

Sound Levels in the Neighborhood

The sound levels were recorded in dbA. The values recorded were the peak values stored by the meter or seen by the collector on the digital readout during the sound event of interest, such as a blowing of the wayside horn. Points of interest in the area 2000 ft by 3000 ft around the crossing were considered to find the effect of the horn. Figure 1 shows that there was typically a 10 to 25 dbA difference between the wayside and train horns. Notice that the farther from the crossing, higher the difference between the wayside and train horn volumes.

Figure 1 - Sound levels in the neighborhood plotted by distance from crossing



Resident Opinion

We gathered these data before and after installation of the wayside through a postcard-type written survey. The questions were: 1) What is the impact of railroad crossing horn in your life? 2) How is the sound of the horn at the crossing? In the after period we also added two questions to the form: 3) Have you noticed a change in the horn system at Winstead Avenue in the past few months?" 4) If so, do you prefer new or old one? We were able to complete 57 surveys in the before period. We received 29 responses from these postcards, providing a lower response rate. A spatial analysis showed that the responses from the before and after periods covered the same areas. Table 2 summarizes the responses from the resident opinion surveys.

Table 2 – Summary of resident's responses

Question	Response	Percentage	
		Before	After
1. Impact	Very good	20	21
	Good	23	18
	No effect	50	50
	Bad	4	7
	Very bad	4	4
2. Sound	Too soft	5	0
	Soft	16	32
	No idea	35	36
	Loud	39	29
	Too loud	5	4
3. Noticed	Yes		38
	No		62
4. Prefer	New		21
	No preference		67
	Old		13

Driver Opinion

Opinions from a sample of drivers using the test crossing were one of the five types of data we analyzed in this research. We distributed 100 survey forms before and after the wayside horn installation. In the before period we received 49 responses for a very respectable response rate, and in the after period we received 23 responses with the decreased response rate.

The questions were: 1) How often do you drive through the crossing at Winstead Avenue? 2) What device first alerts you of the train at that crossing? 3) How is the volume of the train horn at that crossing? 4) Do you usually slow down while crossing the railroad? In the after period, we added two more questions: 5) Have you noticed a change in the horn system at Winstead Avenue in the past few months? 6) If so, which one do you prefer? Table 3 shows a summary of the results from the driver survey.

Table 3 - Summary of driver's responses

Question	Response	Percentage	
		Before	After
1. How often?	Day	68	22
	2-day	20	13
	Week	6	22
	Month	0	43
	Just today	6	0
2. Device?	Horn	20	26
	Lights	49	43
	Gate	29	9
	Other cars	0	4
	Don't know	2	17
3. Volume?	Very good	35	22
	Good	45	22
	Don't know	18	48
	Bad	2	4
	Very bad	0	4
4. Slow?	Yes	82	91
	No	18	9
5. Noticed?	Yes		4
	No		96

Driver Behavior

Driver behavior was one of the five types of data we analyzed during the study. The study team judged driver behavior before and after installation of the wayside horn by viewing the video images recorded with a camera system. Our measure of effectiveness were 1) time after gates activated until the last vehicle crossed stop bar and 2) time after gates activated until the first vehicle stopped. These were the best measures we had available of driver reaction to a train event that was related to safety. Table 4 shows the data.

A qualitative look at the motorist behavior data shows that there may have been a slight shift to more vehicles choosing to stop rather than going through the crossing after device activation with the wayside horn in use. There were only 13 such cases in the "after" period as opposed to 37 cases in the "before" period. Otherwise, the distributions of events look quite similar.

Table 4 - Driver behavior data

	Number of vehicles	
Seconds after gates activated that vehicle crossed stop bar	Before wayside horn	After wayside horn
0.0-0.5	9	4
0.5-1.0	9	1
1.0-1.5	4	3
1.5-2.0	9	1
2.0-2.5	3	1
>2.5	3	3
Total	37	13
Seconds after gates activated that vehicle stopped		
0-2	5	6
2-4	34	14
4-6	23	18
6-8	28	15
8-10	20	8
10-12	13	9
12-14	6	3
14-16	6	2
16-18	1	0
>18	4	0
Total	140	75
Grand total	177	88

Train Crew Opinion

The fifth and final method of evaluating the test crossing in this research was the opinion of the crews who operated the train through the crossing. The study team provided a short written questionnaire to the Nash County Railroad, and two of the crew members who regularly operate trains through the test crossing responded. The first question was, “How is the crossing after installation of the wayside horn?” One crew member responded that it was “safer”, one responded that was “the same” and neither responded that it was “less safe”. The second question was, “Have you had to blow the train horn at lease once when approaching Winstead Avenue since the wayside horn has been operating?” Both crew members responded “no”. In sum, the response of the train crews using the crossing was generally positive. They appreciated the purpose of the horn, thought that, at least, it did not make the crossing less safe, and had no problem responding properly to the signal.

Conclusions

Based on the results from previous studies and the results from our test described above, the study team arrived at several conclusions regarding the use of a wayside horn at railroad grade crossings. First, the team concludes that the wayside horn offers significant sound relief to residents and others in the area around a crossing. This conclusion is based on our most robust data. Along the track and throughout the neighborhood the wayside horn reduced sound levels by 10 to 25 decibels compared to the locomotive horn. Along the road, the wayside horn produced about the same sound levels as the locomotive horn at a couple of points, but was generally 5 to 10 decibels quieter. Our resident survey picked up some indication that this lower sound level made some residents happier, particularly those near the crossing and track. This finding has been noted in other studies as well. Second, the team concludes that the wayside horn has led to slight, if any, shifts in driver behavior and opinion. Driver opinion is difficult to judge based on our small, changing sample, but seemed to indicate overall that most drivers do not notice a change from locomotive to wayside horn. The driver behavior data we collected are more important and robust than the driver opinions. Those data showed only small shifts in the number of drivers crossing the stop bar after device activation, with some of the small changes better for the wayside horn and some better for the locomotive horn. Overall, this research team agrees with conclusions of previous wayside horn researchers that the wayside horn is likely to provide crossings that are as safe as crossings with locomotive horns. However, there does not appear to be any significant increase in the level of safety over locomotive horns.

Finally, the study team concludes that the wayside horn appears to be reliable and acceptable to train crews. It is likely that the early research experiences with less reliable wayside horns have been corrected. The fail-safe design of the wayside horn is also comforting in this regard.

Recommendations

Based on the study results and the conclusions provided above, the study team recommends that the FRA, the NCDOT and other state DOTs, local agencies, and railroads continue to allow the use of wayside horns when evaluated as part of an engineering diagnostic. The evidence from this study and the previous literature show that wayside horns provide significant sound relief and do not compromise crossing safety.

The study team recommends a couple of avenues for follow-up future research. First, at the Rocky Mount test site the NCDOT should consider collecting more “after period” data on motorist opinion and behavior. Second, someone at the state or Federal levels should fund a before and after collision study of wayside horn installations after a number of them have been working for a few years. Such a study would be fairly easy to conduct methodologically, since regression to the mean would not be a threat to validity. The success of that study would just hinge on the size of the sample.

References

1. Redden, J.W.P., “Is Train Horn Noise a Problem in Your Town?” APWA Reporter, Volume 72, Number 9, September, 2005.
2. Lim, P., “Some Cities Consider Hushing the Horns,” News and Observer, Raleigh, NC, February 13, 2006.
3. Multer, J. and A. Rapoza, “Field Evaluation of a Wayside Horn at Highway-Railroad Grade Crossing,” DOT-VNTSC-FRA-97-1, USDOT, Washington, DC, 1998.
4. Roop, S., “A Safety Evaluation of the RCL Automated Horn System,” Texas Transportation Institute, College Station, May, 2000.
5. Gent, S., S. Logan, and D. Evans, “Evaluation of an Automated Horn Warning System at Three Highway-Railroad Grade Crossings in Ames, Iowa,” Iowa Department of Transportation, Ames, 1998.
6. Poe, C.M., “Automated Wayside Train Horn Warning System Evaluation,” PB Farradyne, Inc., Dallas, TX, 2001.
7. Northwestern University Center for Public Safety, “Evaluation of the Automated Wayside Horn System in Mundelein, Illinois—Final Report,” Evanston, IL, January, 2003.
8. City of Roseville, California, “Evaluation of the Automated Horn System in the City of Roseville, CA,” January, 2004.

A New Economic Appraisal System for Railroad Construction Projects in China

Peihong Chen [1], Jiaqiong Wang, Carolyn Dougherty
1. Research Center for Beijing Transportation Development, School of Economics and Management, Beijing Jiaotong University, Beijing 100044, China; PH (86) 010-81730252; email: chen_peihong@163.com

Abstract: This paper outlines a new set of criteria for the economic appraisal of railroad construction projects, based on overall comprehensive objectives including socioeconomic benefits. Multi-Criteria Analysis (MCA), in conjunction with Cost-Benefit Analysis (CBA), was deemed appropriate for analyzing and comparing these objectives. The proposed criteria are applied to the South Xinjiang Railroad project.
Keywords: Railroad Construction Projects, Economic Appraisal, Multi-Criteria Analysis (MCA), Cost-Benefit Analysis (CBA)

1. Introduction

Because large railroad construction projects significantly affect the national economy, careful economic appraisal should be done as a part of feasibility research and strategic planning before the government invests in these projects. Appraisal is also necessary to determine the effectiveness of a completed project. The Chinese Railway Ministry's "Economic Appraisal Measures Used for Railroad Construction Projects" (2nd edition, 1997), currently used to evaluate railroad projects, has several shortfalls and weaknesses; a set of more reliable criteria adapted to recent economic issues should be considered.

2. The Appraisal Objectives of Railroad Construction Projects

An appraisal study should begin with a clear understanding of the objectives that the project seeks to achieve. EU and ECMT have made great progress in the field of transportation assessment (ECMT, 2001). The EU research program TENASSESS outlines ten general transport objectives (Steen, 2001); the list below has been adapted to suit China's situation:

1) Maximize transport efficiency (improved performance, comfort and level of service, etc.)
2) Improve transport safety
3) Contribute to environmental improvement (reduced local air pollution, noise, use of energy resources, ecological damage, etc.)
4) Contribute to strategic economic development (progress toward government goals in regional economics, spatial planning, etc.)
5) Improve strategic mobility (increased accessibility and railway networks,

provision of missing links, etc.)
6) Contribute to the social dimension (increased equity, stabilization, improved working conditions, etc.)
7) Contribute to the external dimension (development and integration of multimodal networks)

3. Appraisal Methods for Railroad Construction Projects

3.1 Cost-Benefit Analysis (CBA)

In order to adequately evaluate a project, relevant circumstances with and without the project and among alternatives are compared. CBA can be used to determine which option is likely to provide the best economic performance. Although "Economic Appraisal Measures Used for Railroad Construction Projects" uses CBA, it has shortfalls and weaknesses; what follows is a more adequate appraisal framework.

1. Benefits Appraisal. Most railroad construction projects involve improving service and reducing costs. Direct benefits include user effect, provider effect and environmental effect (Japanese Transport Ministry Railway Bureau, 1999). Besides these, benefits to the network as a whole should be considered. The new service will remove some trips from parallel corridors, which decreases the journey time on these lines; however, lines that connect to the new service may experience increased volume and journey time. Net benefits should consider these negative external effects as well as the positive effects (Pedro, 2001).

$$B = DB + PB - CB$$

(1)

B: Net benefit of the project
DB: Direct benefits
PB: Benefits to parallel corridors
CB: Costs to connecting lines

$$DB = UE + PE + EE$$

(2)

UE: User effect *PE:* Provider effect *EE:* Environment effect

$$UE = Q_1(C_1 - C_2) + \tfrac{1}{2}(C_1 - C_2)(Q_1 - Q_2) = \tfrac{1}{2}(C_1 - C_2)(Q_1 + Q_2)$$

(3)

C: Overall user costs including vehicle operating cost, value of journey time, value of comfort, cost of accidents, and user payments
Q: Traffic volume in unit time (1 year)
C1: Total costs without the project
C2: Total costs with the project
Q1: Traffic volume without the project
Q2: Traffic volume with the project

$$PE = MC_2 - MC_1 \tag{4}$$

MC1: Maintenance costs without the project
MC2: Maintenance costs with the project

$$EE = E_2 - E_1$$

(5)
E: Monetized environmental costs and benefits including air pollution, noise, energy consumption
E1: Monetized environmental costs and benefits without the project
E2: Monetized environmental costs and benefits with the project

2. Cost-Benefit Analysis. The sum of all costs (including construction costs) (PV_c) and benefits (PV_b) across a project's lifetime, converted into present day values, is calculated across the appraisal period. The project can then be analyzed with three indices:
a) Cost Benefit Ratio (*CBR*)

$$CBR = PV_b / PV_c$$

(6)
If *CBR>1* the project is feasible.
b) Economic Net Present Value (*ENPV*)

$$ENPV = PV_b - PV_c$$

(7)
If *ENPV>0* the project is feasible.
c) Economic Internal Rate of Return (*EIRR*)

$$EIRR = I \quad if \quad ENPV = 0$$

(8)
If *I>i* the project is feasible.

3.2 Multi-Criteria Analysis (MCA)
CBA assumes that the project should maximize economic benefits; a project with many functions, however, may have multiple appraisal targets. Appraisal of such projects should use multi-criteria analysis (MCA). MCA methods, particularly AHP, have been used to appraise railroad construction projects in China since 1990. MCA methods can be divided into four types:

- Qualitative Appraisal Methods (e.g. Analytic Hierarchy Process (AHP), Expected Value, Frequency, Geometric Scaling)
- Methods between Quantitative and Qualitative Appraisal (e.g. Elimination Set Choice Translating Reality (ELECTRE), Value Matrix, Simple Multi-Attribute Rating Technique (SMART))
- Quantitative Appraisal Methods (e.g. Technique for Order Preference by

Similarity to Ideal Solution (TOPSIS), Goal Achieve, Pair Wise Comp., Multi-Attribute Utility, Satisfying, Weight Sum)
- Methods with Quantitative and Qualitative Appraisal (e.g. Multi-Criteria Evaluation with Qualitative and Quantitative Data (MEQQD) (FENG, 2000))

3.3 Combined approach

MCA is often seen as competing with CBA, although there is no reason why the two approaches may not be used together, for example by adding quantitative and qualitative aspects to the CBA core analysis. The combined approach criteria set consists of core and non-core strategic impacts (ECMT, 2001):

Core impacts		
Construction costs	Materials, construction labor, land and property acquisition (including compensation)	CBA
Operation and maintenance costs	Vehicle and road maintenance, fuel and oil consumption, depreciation, operation and maintenance labor	CBA
Travel cost benefits	Tickets and shipping charges	CBA
Journey time benefits	Travel time, stopover time, commute time	CBA
Safety	Fatalities, injuries, property damage	CBA
Convenience	Comfort, crowding	CBA
Local environment	Noise, air pollution	CBA
Non-core strategic impacts		
Economic development	Land use, economic development, employment	MCA
Mobility	Accessibility and networks	MCA
Environment	Greenhouse effect, loss and damage of ecological, historical, archaeological and scientific sites, energy consumption, natural re-sources , visual environment	MCA
Other strategic policy and planning impacts	social dimension, integration of transport system	MCA

Source: EUNET 2001(E.U.) （Revised by author considering the circumstances of China）

Core impacts (valuated impacts) are well defined and can be monetized. These criteria measure the first two objects and part of the third object mentioned in section 2. Core impacts are generally the most important, and are measured by the CBA method, which determines the feasibility of the project. Non-core strategic impacts are not easily monetized, but can in several cases be modeled. These criteria measure part of the third object and the last four objects mentioned in section 2, and are measured by the MCA method. This analysis will set the priorities for the project (H. Morisugi 2000).

4. An Empirical Study--South Xinjiang Railroad

The 970 km South Xinjiang Railroad, built between 1996 and 2000, traverses South Xinjiang from Kashi to Kuerle near the western border of China. There are five commercial ports in South Xinjiang. About 7,660,000 people from diverse racial and ethnic groups live in the region. The project's evaluation period is 30 years, between 1996 and 2025. The Railway Ministry's First Prospecting and Designing Institute estimated its Financial Internal Rate of Return at 4.95%, less than the railway industry benchmark yield ratio of 6%; thus the profit margin is not optimal. Because the freight volume is not adequate, the South Xinjiang railroad line has been running at a deficit since 2000 (Huang, 2005).

However, from the social and economic point of view, according to the CBA method described in section 3.1, the project is feasible. User effects include savings in travel time and cost, enhancement of passenger comfort, and reduction of freight damage. Provider effects include savings in railroad maintenance and vehicle operating cost. Environmental effects include reduction of traffic accidents and reduction in air pollution (NO, CO_2). In addition, the EIRR of the project is 16.67%, more than the Social Discount rate (10%), the ECBR is 1.85, more than 1, and the ENPV is 3.8 bn RMB, more than 0 (HUANG, 2005).

MCA can be used to measure non-core, strategic impacts of the project. A study by LIN, RONG and CHEN (2005), evaluated all effects of the project using AHP. The study defined 17 indices grouped under four criteria, and classified appraised scores into five categories: best, better, ordinary, worse, worst. The study compared the values of each index with and without the project, and assigned appraisal scores to each.

Criterion (% weighting)	Index	Weighting (%)	Scores (%)				
			best	better	ordinary	worse	worst
Economic (25)	Operating revenue	4	0	0	100	0	0
	Maintenance and operating cost	4	0	0	100	0	0
	Regional economic development	66	20	80	0	0	0
	Regional economic structure	26	0	40	60	0	0
Social (25)	Travel time	4	100	0	0	0	0
	Comfort	4	100	0	0	0	0
	Safety	4	100	0	0	0	0
	Regional employment	23	60	40	0	0	0
	Distribution equity	21	0	0	0	60	40
	Communication	44	0	50	50	0	0
Environmental (25)	Pollution	45	0	0	100	0	0
	Landscape	13	0	90	10	0	0
	Ecological	31	0	0	100	0	0
	Historical	11	0	0	100	0	0
Political (25)	Society stabilization	22	0	40	60	0	0
	Society Order	8	0	0	40	60	0
	State safety	70	100	0	0	0	0

(LIN, RONG and CHEN, 2005)

The results of the appraisal for this project were as follows:

	Total % Score				
Criterion	**Best**	**Better**	**Ordinary**	**Worse**	**Worst**
Political	70	9	16	5	0
Social	34	17	19	21	8
Economic	12	54	34	0	0
Environmental	0	12	88	0	0
Total	29	23	40	6	2

Based on the AHP analysis, the study drew the following conclusions:

- The political effect of the project is most notable, with the highest “best” score of 70%
- The social effect is the second most important, with the highest “best” score of 34%
- The economic effect is the third most important, with the highest “better” score of 54%
- The environmental effect is the least important, with the highest “ordinary” score of 88%

The South Xinjiang Railroad brings great benefits to the community. Of the economic, social, environmental and political effects, the political effect of the project is most notable.

5. Conclusion

As the economic appraisal method currently used for railroad construction projects in China contains serious drawbacks, more accountable railroad construction economic appraisal criteria should be established. The appropriate appraisal method, combining CBA and MCA approaches, is suggested. This approach could provide comprehensive evaluation for a railroad construction project, assessing its feasibility and assisting in identifying the most suitable alternative. New criteria were applied to a representative project, the South Xinjiang Railroad; the CBA analysis showed that the project is feasible, while the MCA showed that its political effect was most significant.

References

ECMT Group on Transport and the Environment (2001). *Assessing the Benefits of Transport*. ECMT, 27 pp. (available at http:// www.oecd.org).

FENG Zhengmin (2000) *Methods of Urban and Regional Analysis*. Taiwan Jiandu Culture Business Co. LTD, Taiwan, pp. 190-200.

HUANG Min (2005) "Theoretic and Empirical Study of Commonweal Characteristics of the Railway Construction Projects". Ph.D. Diss., Univ. of Beijing Jiaotong. pp. 126-144.

Japanese Transport Ministry Railway Bureau (1999). *Railway projects cost-benefit analysis handbook.* M. p. 9.

LIN Xiaoyan, RONG Chaohe, and CHEN Youxiao (2005) “Study on Comprehensive Fuzzy Hierarchy Analysis of Investment Effects of the Land Development Railway”. *Journal of the China Railway Society*, 27 (1) pp. 106-113.

Pedro (2001). *Economic analysis of investment operations: analytical tools and practical applications*. The World Bank, pp. 87-95.

Steen Leleur (2001) *New Appraisal Methods in Transport Planning with an Emphasis on Large Transport Infrastructure Projects*, p.2. (available at http://www.ctt.dtu.dk).

Risk Analysis and Management of Ship Collisions with Bridges

Wang Zeng-zhong[1], Wang Jun-jie[2], Fan Li-chu[3]
Co-author: Cristina L. Brush
[1] Architectural Engineering College, Shanghai Normal University, 20118, Shanghai, P.R. China; Post Doctor of the Department of Bridge Engineering, Tongji University; PH&FAX(021) 571252281; email: wangzz@shnu.edu.cn
[2] Department of Bridge Engineering, Tongji University, 200092, Shanghai, P.R. China; PH&FAX (021) 6598219; email: jjwang@mail.tongji.edu.cn
[3] Department of Bridge Engineering, Tongji University, 200092, Shanghai, P.R. China; PH&FAX (021) 6598219; email: lcfan@mail.tongji.edu.cn

Abstract

On June 15, 2007 a great tragedy occurred due to a ship collision with a bridge pier which resulted in nine fatalities in Guangdong Providence, China. This event greatly increased concern over the safety of bridges crossing navigable waterways. Due to the rapid expansion of highway networks in China many bridges which span rivers, lakes and crossing bays have been constructed or are under construction. At the same time, waterborne traffic has also been increasing which has caused an increased risk of ship collision with bridges (SCB). The risk analysis and management of SCB are vital to the safety of both bridges and ships. In this paper the risks of SCB are analyzed, evaluated, classified and rated according to their impact distribution, probability and loss of the accident. The risk acceptance criteria are analyzed and several kinds of ship collision protection methods are discussed. The framework of the "life-cycle cost oriented" risk reduction methodology is proposed to guide the bridge design and/or indicate the appropriate level of collision protection and management.
Keywords: Ship Collision, Bridge, Risk Analysis, Risk Management.

1. Introduction

On the morning of June 15th, 2007 about 200 meters of the 1,600 meter long bridge which connects Jiujiang in Foshan to Heshan City in Guangdong Province, China, collapsed causing four cars to plunge into the river and seven passengers in the cars along with two bridge workers to perish ("Tragedy of 6/15"). (Shenzhen Daily 2007) It is ascertained that the accident was caused by a boat carrying sand that was sailing out of the main navigation channel colliding with a bridge support, causing a chain reaction to be triggered. It is ironic that no agency or unit of government was held responsible for this tragedy and that the cable-stayed bridge was said to be in stable condition since it came to use in 1988. The bridge was built with its pillars designed to withstand an impact of 1,200 tons and its arches of 40 tons, both exceeding mandatory safety standards. The boat seems to be almost the only cause of the catastrophic accident. The "Tragedy of 6/15" has sparked discussions about whether the bridge was strong enough and highlights the importance of waterway transport safety in the vicinity of large-scale bridges.

Highway bridges over navigable waters are of great importance to the transportation system as they connect key highways at crucial nodes. This connectivity has made significant contributions to the nation's economy. It is unfortunate that there is not a comprehensive national specification for ship collision design of highway bridges in China at present. The risk analysis and management of SCB is a key issue not only in bridge design, but also in lifetime management of bridges. The tasks of ship collision risk analysis include the consequence assessment for accidental ship collisions, the classifications of the risks of SCB, the development of risk acceptance criteria, and the optimization of risk reduction measures. These can be intentionally used to guide the bridge design, collision protection and risk management.

2. Classifications of the Risks of SCB and Risk Acceptance Criteria

Highway bridges in the vicinity of extreme ship traffic inevitably pose potential collision hazards. Both the probability of SCB and the magnitude of its consequences are involved in risk analysis. Risk management is the process used for risk assessment, controlling such risks, and arriving at an acceptable level of risk.

2.1 Classification of the Risks of SCB

The risks of SCB affect bridge owners, facility users, ship owners, and other stakeholders. SCB may cause the risks of loss of lives, economic loss, environmental damages, and other unwanted effects. (Xiang 2002)

All of the risks of SCB could be expressed in costs that include pier and span replacement costs, motorist inconvenience costs, port interruption costs, and other additional costs. The pier and span replacement costs include the costs associated with debris removal from the waterway due to the collapsed bridge sections along with engineering and construction inspection costs. The motorist inconvenience costs include the additional vehicle operating costs incurred by motorists who must take a longer, more congested route, along with the toll revenues lost by the bridge owners. The port interruption costs include the costs associated with the temporary closure of port facilities. Additional costs include environmental, business, social, and loss of life costs. Although in some cases subjective and probabilistic in nature, the costs associated with injury and the loss of human life are very important in the assessment of the SCB risk and must be considered. Several researchers have provided some excellent methods to help quantify these costs. (Chang 1996)

2.2 Risk Acceptance Criteria

One of the many performance goals in the design of highway bridges is to ensure that the risks of major accidents and service disruptions are low enough to be acceptable to users, the public, and those responsible for public safety. Serious ship collisions with bridges are extreme events associated with a great amount of uncertainty, especially with respect to the impact loads involved. Since bridges designed for the worst case would be overly conservative and economically unreasonable, a certain amount of risk should be considered as acceptable. The risk acceptance criteria could be expressed in two main groups: target safety acceptance criteria and target economic acceptance criteria. Target safety acceptance criteria deals with the collapse probability of the bridge and must be less than an acceptable level. The AASHTO Specifications outline that the acceptable annual frequency of collapse (AF) of critical bridges shall be less than or equal to 0.01 in 100 years and 0.001 for non-critical bridges. (AASHTO 1991) The annual frequency of bridge collapse due to ship collision (AF) is the product of the annual number of ships which can strike the bridge, the probability of ship aberrance, the geometric probability of a collision, and the probability of bridge collapse due to collision of an aberrant ship. Target economic acceptance criteria deals with cost of the bridge and must be feasible and rational. A balance between safety and economic cost must be sought.

The risk acceptance criteria for significant bridge structures should be established and approved by the owner. Once established, a rational risk assessment procedure can be developed and the risks can be calculated and compared to the acceptance criteria. Cost effective analysis is used to assess economic feasibility and desirability. This can be done with a conventional benefit/cost (B/C) ratio calculation. The benefit, B, is expressed in the total present value of the avoidable risks, and the cost, C, is expressed in the total present value of the cost to build, maintain, and operate the system required to provide those benefits.

3. The Reflection of the "Tragedy of 6/15" and Risk Reduction Methods

3.1 The Reflection of the "Tragedy of 6/15"

Although the exact reasons for the "Tragedy of 6/15" have not been thoroughly revealed, it is believed that the root cause of the bridge collapse accident involved several factors. The disobeying of traffic regulations seems to be the main cause of the disaster. The captain of the vessel that collided with the bridge was traveling at 20.8 kilometers an hour when the visibility was only 70 meters. He was also said to be ignorant of the use of radar and had only 1.5 hours of sleep in the last 24 hours before the collision. The area had also been experiencing heavy rainfall and the higher, faster water may have contributed to the accident. The vessel sailed out of the main navigation channel and collided with one of the piers causing all four connecting piers to collapse. The impact force from the collision is estimated to be 1,000 tons.

This accident has sparked many discussions about whether the bridge was strong enough, raising the age old question that has puzzled engineers all over the world; "How safe is safe enough?"

The issue of waterway transport safety was also brought into the limelight after the tragic collision accident. China has nearly 200,000 vessels with combined capacities of more than 100 million tons. Harbors in the country handled 5.6 billion tons of goods in 2006. Inland waterways in China run approximately 120,000 kilometers in length. An evaluation of accident statistics indicates that human errors and adverse environmental conditions are the primary reasons for accidents rather than mechanical failures. As vessels are getting bigger and traffic flows are getting denser, the comprehensive all-weather and quick-response water transport safety supervision and rescue system on major waterways are in deadly need to intensify monitoring and management to ensure water transport safety and order.

3.2 Risk Reduction Methods

The objectives of the risk reduction methods are to reduce the probability and consequences of ship collisions. The methods used are: 1) Physical protection systems for bridge piers, 2) Motorist warning systems, and 3) Aids to navigation alternatives.

Bridges may be designed to resist ship impact loads or be protected by a bridge protection system. The first option involves designing the bridge to withstand the ship impact loads with little to no damage. The second option allows the designer to provide a protection system of fenders, pile supported structures, dolphins, or artificial islands to either reduce the magnitude of the impact loads to within the allowable strength of the bridge or to independently protect the bridge elements.

The greatest loss of life in catastrophic ship/bridge collisions has resulted from the continuation of highway traffic after the span has been severed, as was the case in the "Tragedy of 6/15". The motorist warning system, which includes hazard detection systems, verification devices, and traffic control and information devices, is essential to risk reduction, especially during peak traffic hours or when a large ship approaches the bridge.

Improvements in navigation within the navigable channel at a bridge location would significantly reduce the vulnerability of bridge to ship collisions. The aids to navigation involve operational alternatives, standard navigation alternatives, and electronic navigation systems. Since sixty to eighty-five percent of all ship collision accidents are attributed to pilot error, it is important that all aspects of bridge design, location, and aids to navigation with respect to the navigation channel be carefully evaluated with the purpose of improving or maintaining safe navigation in waterways in the vicinity of bridges. (AASHTO 1991) The implementation of radio-telephone communication between the ship and the bridge operators or toll personnel is one of the most effective and least expensive alternatives to improve the safety of bridges.

Extensive analytical and experimental investigations have to be carried out to verify the functions of the risk reduction methods and risk analysis tools that can be applied for evaluation of their effects. However, the basic design philosophy is that it is possible to design a bridge in a cost-effective manner which minimizes the risk of catastrophic superstructure collapse from ship collisions.

4. Life-Cycle Cost Oriented Risk Reduction Methodology

The design strength of the bridge or type of protection to be provided shall be selected based on cost-effectiveness acceptance criteria where the cost of the bridge strengthening or bridge protection system is compared against the benefits of risk reduction. Until recently, design optimization research focused mainly upon reducing the initial costs of design and construction. However, research has shown that these initial costs are often dwarfed by the costs of repairing, inspecting, and maintaining a bridge over its useful life. (Frangopol 2007) Indeed, a life-cycle cost analysis consists of the calculation of not only the initial costs, but also includes the costs due to operation, inspection, maintenance, repair, and damage/collapse consequences during a specified lifetime. Future costs are typically discounted to the present value for meaningful comparison. The goal is to produce a cost-effective design solution that balances the initial cost and lifetime cost in the preferred manner.

The minimum expected life-cycle cost with respect to lifetime performance is the most widely used criterion for design optimization of risk reduction measures. All costs, benefits and other values shall be expressed in constant dollars. The general form of the expected life-cycle cost can be calculated as:

$$C_{ET} = C_T + C_{PM} + C_{INS} + C_R \tag{1}$$

Where C_{ET} = expected total cost, C_T = initial design/construction cost of the bridge, C_{PM} = expected cost of routine maintenance, C_{INS} = expected cost of performing inspections, and C_R = expected cost of the risk due to SCB.

Inclusion of risk reduction methods into this general form results in:

$$C^0_{ET} = C^0_T + C^0_{PM} + C^0_{INS} + C^0_R + C_M \tag{2}$$

Where C_M = expected cost of the risk reduction system which is best treated with respect to a life-cycle cost as:

$$C_M = M_T + M_{OP} + M_{INS} + M_{REP} \tag{3}$$

Where M_T = expected initial design/construction cost of the risk reduction system, M_{OP} = expected operational cost of the risk reduction system, M_{INS} = expected inspection cost of the risk reduction system, and M_{REP} = expected repair cost of the risk reduction system.

The benefit of the risk reduction system, B_M, is then computed through a comparison of the expected life-cycle total cost with and without the risk reduction system by subtracting Equation [1] from Equation [2]:

$$B_M = C_{ET} - C^0_{ET} \tag{4}$$

The risk reduction system would only be justified if $B_M > 0$, which would mean that the risk reduction system is cost effective, unless it was code-driven. The expected cost of the risk due to SCB, C_R, is arguably the most difficult and controversial term in the equation, as discussed in section 2.1 of this paper. The most important cost to be considered for this term is the cost associated with injuries and loss of human life.

In some cases the bridge costs of initial design/construction, routine maintenance, and performing inspections are the same with or without the risk reduction system. The Equation [4] could be expressed as:

$$B_M = C_R - (C^0_R + C_M) = (C_R - C^0_R) - C_M \tag{5}$$

Therefore if the risk reduction cost is more than the cost of implementing the risk reduction system and the safety level is in accordance with the mandatory code, the alternative of risk reduction methodologies is cost-effective.

5. Conclusions

The risk analysis and management of ship collisions with bridges are of great significance not only in bridge design but also in the operation of bridges. It is very important to consider the risk of ship-bridge collisions as early as possible. In the planning phase for a new bridge the conceptual design of the risk reduction systems are the most essential. The objectives of the risk reduction methods are to reduce the probability and the consequences of ship collisions. Risk analysis tools should be evaluated for their effects and risk management should be involved.

The "Tragedy of 6/15" taught us a profound lesson. Under increasing pressure as ships get bigger and traffic flows get denser in China, it is an important task for researchers in the field of bridge engineering to develop design codes or guidelines for ship collisions with bridges. Meanwhile the cooperation between the relative agencies or units of government and bridge owners is needed to intensify monitoring and management to ensure water transport safety and order.

Acknowledgements

The support of the Shanghai Municipal Education Commission through grant No-06DZ023 is greatly acknowledged. The opinions and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the sponsoring organizations.

References

AASHTO (1991). "Guide Specification and Commentary for Vessel Collision Design of Highway Bridges." Washington, D.C..

Chang, S.E., and Shinozuka, M. (1996). "Life-Cycle Cost Analysis with Natural Hazard Risk." Journal of Infrastructural Systems, ASCE, 2, 118-126.

Frangopol, D.M., and Messervey, T.B. (2007). "Integrated Life-Cycle Health Monitoring, Maintenance, Management, and Cost of Civil Infrastructure." Proceedings of International Symposium in Integrated Life-Cycle Design and Management of Infrastructure, 1-20.

Shenzen Daily (2007). "No Flaws Before Bridge Collapse." http://news.xinhuanet.com/english/2007-06/content_6263295.htm. June 21, 2007.

Xiang, H.F., Fan, L.C., and Wang, J.J. (2002). "State of Art Collision Design for Bridges and Future Research." Journal of Tongji University, Vol. 30(4), 386-392.

Traffic Incident Duration Prediction Based on the Bayesian Decision Tree Method

Bei-bei Ji Yang[1], Xiaoning Zhang[2], LiJun Sun [3]

[1] Candidate for PH.D.Tongji University.Shanghai, P.R. China, 200092; PH (+86-21) 65980561; email:jiyang.1006@gmail.com
[2] Associate Professor. Institute of Traffic Engineering Tongji University, 239-0826, China; Email: cexzhang@mail.tongji.edu.cn Phone (Fax): 86-216-598-2443
[3] Prof., Department of Transportation Engineering, TongJi University, Shanghai, 200-0092, China; PH (+86-21) 65980561; email: ljsun@mail.tongji.edu.cn:

Abstract

Effective incident management requires a full understanding of various characteristics of incidents to accurately estimate incident durations and to help make more efficient decision to reduce the impact of non-recurring congestion due to these accident. This paper presents a new prediction model base on the Bayesian decision model to estimate the traffic incident duration. The basic theory of Bayesian decision model is presented and the prediction model is created based on this theory using incident data collected in Rijkswaterstaat Verkeerscentrum Nederland from various sources. Compared to most existing methods, the proposed model is unique in two aspects: firstly, the model is adaptive in the presence of real incident for which data might only be partially available or in the presence of incomplete information. Secondly, this model is shown to perform better theoretical prediction accuracy compared to the decision model or Bayesian model.

Introduction

Traffic congestion is significant problem in most large urban areas. Incidents are major contributors to delay and have far-reaching consequences for safety, congestion, air pollution, and the costs of travel. These events include vehicle breakdowns, traffic accidents, debris on the road, and other unplanned events significantly affecting roadway operations. Consequently, effective traffic management techniques have the potential to substantially reduce congestion. The actions taken by traffic managers require a full understanding of the nature and tendencies of freeway incident. Thus the specific field of Incident Management has become an important component of traffic management, which involves the steps of clearing traffic incidents quickly and then minimizing the congestion effects on the traffic flow.

Incident duration is made up of four components, namely detection time, response time, clearance time and recovery time. We are concerned with the time between the detection of a traffic incident and the clearance of the road in this research. In the past, quite a few researchers investigate the traffic incident duration problem. Roughly, the traffic incident prediction models can be classified into the following categories: probabilistic distributions (Sullivan, E. C,1997), linear regression models(Garib 1997), decision tree models(Sethi 1994), time sequential models(Khattak 1995), Fuzzy Logic models(Wang, 2002), and so on. The implementation of most of these models is often difficult because the knowledge of all traffic incident characteristics is often incomplete and obtained sequentially.

In this paper, we develop a decision tree model based on Bayesian method which can deal with the incomplete data or inconsistent data(Fan Jian-cong 2005). Firstly, the paper describes the decision tree algorithm based Bayesian method and Secondly, it describes the data used to calibrate and validate the model, and thirdly it describes the development of the model using this data. Finally, the prediction accuracy of this model is presented in the section5.

Methodology for estimating duration

The Bayesian node is inserted into the decision tree model, which is defined Bayesian decision tree model, as showed in Figure1. The Bayesian node contains two values that are “0” and “f”. If the object characteristics information is completed then the value of Bayesian node is “0”, there is no any calculations. However if the object characteristics is missing then the value of Bayesian node is “f”, we have to calculate the value of the Bayesian node(Fan Jian-cong 2005).

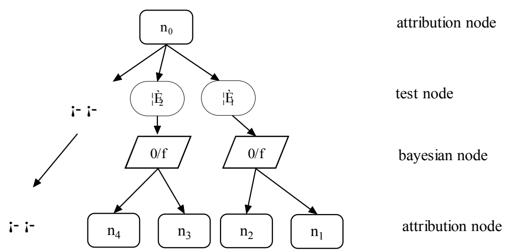


Figure 1. Bayesian decision tree model

The ID3 algorithm is used to construct a decision tree in this paper. The attribute that gives the best prediction at each internal node is determined using information theory. The attribute that reduces the entropy of the category information the most is selected as the best attribute. Assume that there are m categories $C_1, C_2, \ldots, C_m$, and a set S of objects whose categories are known. Let s_i be the set of objects with category C_i. Then the information needed to classify an object in S is:

$$I(s_1, s_2, \cdots, s_m) = \sum_{i=1}^{m} \left(-\frac{|s_i|}{|s|} \cdot lb \frac{|s_i|}{|s|}\right) \tag{1}$$

Let us assume that the objects have n attributes $A_1, A_2, \ldots, A_n$, In order to assess the prediction quality of an attribute A, we need to calculate the information needed to classify an object in S given its value for attribute A. Assume that A can have v values $a_1, a_2, \ldots, a_v$. Then the information of S given A is:

$$E(A) = \sum_{j=1}^{v} \frac{s_{1j} + s_{sj} + \cdots + s_{mj}}{s} I\left(s_{1j}, s_{sj}, \cdots, s_{mj}\right) \tag{2}$$

The information gain for each attribute A is:

$$Gain(A) = I(s_1, s_2, \cdots, s_m) - E(A) \tag{3}$$

The best attribute A is then the one that has the maximum gain, i.e., minimum E(A), among all considered attributes.

F is known as Bayesian posterior probability, which is written as

$$P(C_j|X) = \frac{P\left(X|C_j\right)P\left(C_j\right)}{P(X)} \tag{4}$$

It makes the strong assumption that X_i belongs to a particular class given the value of some attribute is independent of the values of all other attributes. Under this mutual conditional independence assumption, this reduces to

$$P(C_j|X_i) = \frac{P(C_j)}{P(X_i)} \prod_{j=1}^{m} P(X_{ij}|C_j) \tag{5}$$

For each category C_j. Since the denominator will be the same for all categories, we need only calculate the numerator for each category j, choosing

$$P(C_i|X_i) > P(C_j|X_i), \quad 1 \le j \le m, j \ne i \tag{6}$$

And assigning I to category C_i.

In the context of incident duration prediction, the attributes X correspond to observable incident characteristics, the number of blocked lanes, the location of the incident, weather conditions, and so on. The categories C correspond to the incident duration that is the interval of the time.

Data description

The data used in this research was collected by the Rijkswaterstaat Verkeerscentrum Nederland from various sources. These data are normally used specifically for monitoring incident management, and the amount of detail and completeness was sufficient for our research. (Willem 2006).

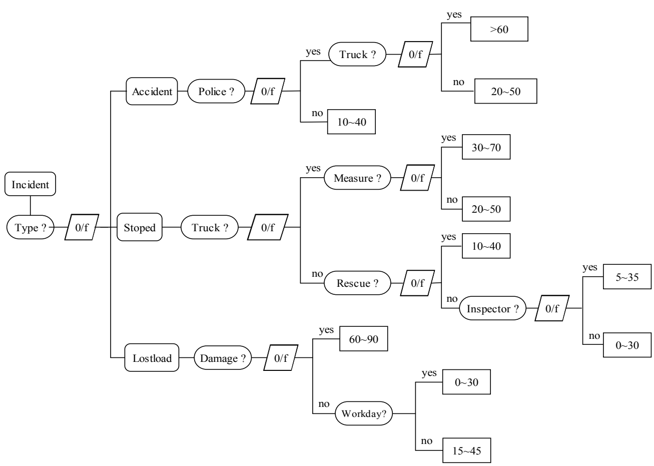
We selected the region of Utrecht (a central province of The Netherlands) and the period from 1st May through 13th September 2005 and found in total 1853 registration of incidents. Table1 contains the descriptive statistics for the incident duration used in the decision model base on Bayesian method.

Table1 1. Descriptive Statistics of Incident duration

	accident	stopped	Lost load
N	903	571	379
average	42.74	43.73	26.64
S.E.mean	1.377	1.586	1.006
Median	33	31	23
Mode	25	26	18
Std.deviation	41.383	37.901	19.58
Range	435	280	138
Minimum	0	2	1
Maximum	435	282	139

Duration prediction models

Based on the principles described in the previous section, we developed a prediction model for traffic incident duration. All 1853 incident registrations were used to build this model corresponding to all different attribution value combinations. Figure2 illustrate the detailed model of traffic incident duration. In this figure 'Police' means whether the Police is required in this incident, and 'Truck' means whether the kind of the vehicles involved is truck or not. 'Measure' represents whether the roadway is closed, and 'Rescue' represents whether the tow truck is required, and 'Inspector' represents whether there is road manager patrol nearby the location of the incident. 'Damage' represents whether there is road equipment damaged, and 'workday' represents whether the day incident happened is workday or not.



Figrue 2. Decision tree model for traffic incident in Nederland

As an example figure 3 shows the distribution of a kind of traffic incidents in intervals of 30 minutes. The characteristic of this kind of incident is accident without the help of police. Each horizontal bar corresponds to an interval, and its length is determined by the number of elements having duration within interval. The intervals are shifted 5 minutes with respect to their neighboring intervals. The third interval with an interval of 10 to 40 contains the 415 incidents that is the largest frequency in the figure4, so the leaf node of this kind of incident is 10 to 40 minutes.

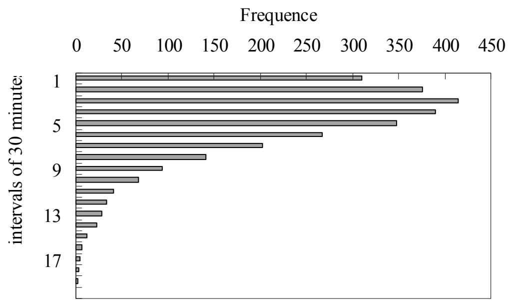


Figure 3. Number of incidents (horizontal) for different intervals of incident duration (vertical)

In order to test the adaptive ability and robust of this model in the presence of missing information, we create 20 percent missing or incomplete data in the validation data set randomly. So when the data has the enough information which could be categorized according to the test node of decision model, else the value f of Bayesian node should be calculated based on the Bayesian theory. On the basis of the value of the Bayesian node the incident could be classified to the C_i.

The theoretical accuracy is shown in table 2. For this table, we used 11 parameter value combinations and different kind incident has different parameters. For each combination, the table shows the number of registered incidents and the percentage of those incidents for which the estimated incident duration interval contained the actual incident duration. Averaging the results of Table2, we find a theoretical reliability of 74%. So, the duration of roughly 3 out of 4 incidents can be estimated correctly. Contrastively the classification tree method correctly categorized 46% using the same incident data set(Willem 2006).

Table 2. Theoretical accuracy of prediction model

	type	#reg.	range	Accuracy
accident	police=1,truck=1	49	>60	0.7029
	police=1,truck=0	322	20-50	0.6384
	police=0	532	10-40	0.8751
stopped	truck=1,measure=1	117	30-70	0.5033
	truck=1,measure=0	72	20-50	0.5247
	truck=0,rescue=1	152	10-40	0.8480
	truck=0,rescue=0,inspector=1	188	5-35	0.7092
	truck=0,rescue=0,inspector=0	42	0-30	1.0847

Lost load	damage=1	5	60-90	0.6667
	damage=0,workday=0	68	15-45	0.6209
	damage=0,workday=1	306	0-30	0.6790
sum		1853		0.7357

Conclusions

In addition to the description of the theory of Bayesian decision tree and its applicability to our problem, we use real world data set collected in Rijkswaterstaat Verkeerscentrum Nederland to develop the prediction model. To our knowledge, this is the first application of this type of model to incident duration prediction, which can deal with dirty traffic incident data such as incomplete or inconsistent data. Furthermore the theoretical accuracy of this model is more accurate than the classification tree method(Stephen Boyles 2007)(Willem 2006). Such a prediction model can help the operators to determine the interval of the incident duration accurately.

References

Fan Jian-cong and Zhang Wen-jin. Decision tree classification algorithm based on Bayesian method. Computer Applications. Vol.25 No. 12 Dec.2005. pp 2882-2885.

Garib, A., A. E. Radwan and H. Al-Deek. "Estimating Magnitude and Duration of Incident Delays." Journal of Transportation Engineering. Vol. 123, No. 6, November/December 1997. pp. 459-466.

Khattak, Asad J., Joseph L. Schofer and Mu-Ham Wang. "A Simple Time Sequential Procedure for Predicting Freeway Incident Duration." IVHS Journal. Vol. 2, No.2, 1995. pp 113-138.

Sethi, V., Koppelman, F. S., Flannery, C, P., Bhomderi, N., Schofer, L., 1994. Duration and travel time impacts of incidents- ADVANCE Project Technical Report, Northwestern University, Evanston.

Stephen Boyles and David Fajardo. A Naïve Bayesian Classifier for Incident Duration Prediction. TRB 2007 Annual Meeting.

Sullivan, E. C. "New Model for Predicting Incidents and Incident Delay." ASCE Journal of Transportation Engineering. Vol. 123, July/August 1997. pp 267-275.

Wang, W., H. Chen, MC Bell, 2002. A Study of Characteristics of Motorway Vehicle Breakdown Duration. 11th International Conference on Road Transport Information and Control, London, UK.

Willem Jan J. Knibbe and Tom P. Alkim. Automated Estimation of Incident Duration on Dutch Highways. IEEE 2006.

Effective Freeway Incident Response: A Bayesian Network Based Algorithm

Shunxin YANG[1], Wei WANG[2] and Feng GUO[3]

[1]Transportation College, Southeast University, No.2, Sipailou, 210096, Nanjing, China, Phone: 86-025-83794814, E-mail: shunxin.yang@gmail.com
[2]Transportation College, Southeast University, No.2, Sipailou, 210096, Nanjing, China, Phone: 86-025-83794814, E-mail: ww.movealong@gmail.com
[3]Department of Statistics, Virginia Tech, 415B Hutcheson Hall, 24061-0439, Blacksburg, VA, U.S.A. Phone: 540-231-7933, E-mail: feng.guo@vt.edu

Abstract: Prompt and accurate incident response plan is an essential part of an effective freeway incident management system. However, uncertainties associated with the incident often lead to delayed or inaccurate response decision and processing plan. The core of the proposed algorithm is a Bayesian network based decision-making system that quantifies the causal dependencies between response decisions and incidents information. With input from real time incident data, the Bayesian networks update the probability of uncertainties associated with highway incidents through a two-way inference mechanism. We apply the proposed method to the incidents data recorded in Nanjing-Lianyungang freeway from 2001 to 2004 combined with more complete incidents data recorded in 2005 obtained from RK-IMS (a rule-based freeway incident management system). The model was validated using incidents data of the year 2006 from RK-IMS. The comparison results indicate that the proposed algorithm is accurate and reliable.

1 Introduction

Monitoring system, telecommunications, and emergency telephones equipments have been installed on most freeways in China. However, the ability of real-time incident response, including incident detection, incident verification, response plan, and on-the-spot management, is generally at a low level and there is an imperative need for improvement.

Since the 1979's, many studies have been focused on freeway incident responses. The methods used in these studies can be generally classified into the following categories: Case-based Reasoning (Chowdhury et al., 2006), Rule-based Reasoning(Arti Gupta 1992; Ritchie 1990; Hu 2003; Nan Zou 2004; Zografos 2002), Probability Density Function(Phillip Rust 2003), Risk-based Reasoning (Ghada Mahmoud Hamouda 2004), Bayesian Networks Reasoning (Ozbay 2006; Shiliang Sun 2006), Bayesian hierarchical analysis (Huang Helai 2007).

One challenge in real-time incident management is that when making decisions, the available information is often incomplete or inaccurate, while incident management system based on rule-based reasoning could generate clearly formulated decisions with accurate input. Although 42 tables including more than 500 rules have been used in the RK-IMS (a rule-based freeway incident management system developed by our research group), which has been applied in the Nanjing-Lianyungang Freeway, there are still some uncertainties of incidents which affect the operators to make accurate decisions during practical use.

The objective of this paper is to develop an algorithm for incident response under non-optimal input information. The remainder of this paper is organized as follows: Section 2 describes the model and methodology used in this paper, including Bayesian Networks (BNs), which is the most critical part of the algorithm; the parameter-learning method; sequential updating and the algorithm architecture. Section 3 presents a case study using incident data in the northern section of Nanjing-Lianyungang freeway from the year 2001 to 2004, and this case is illustrated in three steps, i.e., data analysis, Bayesian Network model and comparative analysis. In Section 4, we summarize the contribution of this paper and discuss future research directions.

2 Model and Methodology

Bayesian Network

BNs consist of two components. The first is a directed acyclic graph (DAG) G, whose nodes correspond to the variables $X_1, \cdots, X_n$. The DAG defines the structure of a BN. The second component is the probabilistic part Θ, which is a set of θs and describes a conditional probability distribution for each variable, given its parents in the graph G. The graph G encodes conditional independence among the variables. Each variable X_i is independent of its non-descendants, given its parents in G. Thus, the joint probability distribution of the variables can be decomposed in the product form of the Equation (1), where $pa(X_i)$ is the set of parents of X_i in G. The conditional probabilities that appear in the product form are stored in conditional probability tables as the parameters θ that we want to estimate. Estimation includes assigning a conjugate before the parameters θ, computing the likelihood function and multiplying the prior density and the likelihood function to obtain an algebraic form proportional to the posterior density.

$$P(X_1, \cdots, X_n) = \prod_1^n P(X_i \mid pa(X_i)) \tag{1}$$

Parameter Estimation

In many case, we have only compute sample statistics for each node and its parents. For example, we estimate $P(InvolvedHAZMAT = yes \mid isTruck = no)$ by the fraction n_c/n where n is the total number of training examples for which $isTruck = no$, and n_c is the number of these for which $InvolvedHAZMAT = yes$. However, when only a few incidents cases are observed and recorded, the estimation results would be less accurate. To overcome this obstacle we adopt a Bayesian approach to estimate the probability using the M-estimate defined in Equation (2),

$$P = \frac{n_c + mp}{n + m} , \tag{2}$$

where, p is our prior estimate of the probability we wish to determine, and m is a constant called the equivalent sample size, which determine how heavily to weight p relative to the observed data.

In case of missing date, we applied the standard Expectation Maximization (EM) method, a widely used approach to estimate parameters in the presence of missing variables.

Sequential Updating

There is an obvious need for improving the performance and accuracy of a Bayesian network as new data is observed. Because of errors in model construction and changes in the dynamics of the domains, we cannot afford to ignore the information in new data. Sequential updating or adaptation (Spiegelhalter and Lauritzen. 1990) is the task of sequentially updating the conditional probability distributions of a Bayesian network when the structure and an initial specification of the conditional probability distributions are given in advance.

Algorithm Architecture

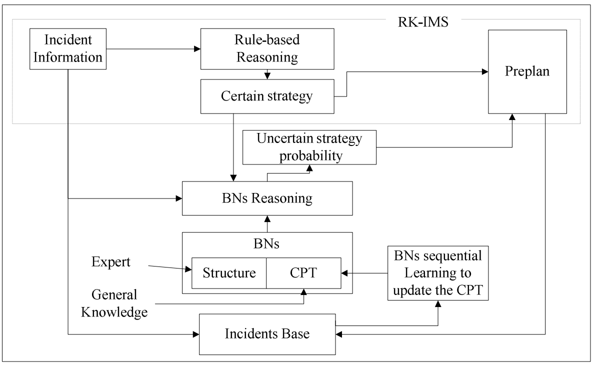


Figure 1. Freeway Incident Response Algorithm.

Freeway incident response is a decision making process. We use Bayesian networks to store prior traffic knowledge and to perform inference in our proposed algorithm architecture. As shown in figure 1, both rule-base reasoning and Bayesian reasoning is carried out on the real-time incident information. With the combination of rule-based reasoning results and actual incident data as input for BNs, the probabilities can be calculated. Moreover, when the incident is disposed finally, the BNs parameters can be relearned through sequential updating model by the incident database consist of incident information and incident response preplan.

3 Case Study

Data Analysis

We use incident data in the northern section of Nanjing-Lianyungang freeway from the year 2001 to 2004. By analyzing the distribution characteristics of traffic incidents occurred along this freeway, including time, weather pattern, incident types etc., we find that many incidents occurred during the time interval from 4 to

8 a.m. and from 14 to 18 as well as from 22 to midnight, with the amounts of incidents occurred in these three periods accounted 23.84%, 18.90% and 9.86% of the total amounts of incidents occurred in that year. In terms of the severity of traffic incidents, majority of serious accidents occurred during 20 to midnight and 6 to 8 a.m., with more fatalities occurred from 20 to midnight. On the other side, the total number of traffic incidents occurred in the Nanjing-Lianyungang freeway during daytime took up 60.86% of the overall amounts, which was more than the 39.14% incidents occurred during night. In terms of seasonal fluctuation, we find that incidents occurred in July, August and September accounted more than 31.47% of all accidents occurred in that year. With regard to the incident type, most incidents occurred along the Nanjing-Lianyungang freeway related to rear-end collision with vehicles ahead, hit fixed facilities on the freeway and turnover. More specifically, truck-involved incidents are mainly hitting the vehicles ahead while car-involved incidents are mainly hitting the guardrail.

Based on the above analysis, we determined the most significant uncertainties during the process of freeway incident response, including the severity of driver injuries and vehicle damages, incident duration, lane blockage, and estimated the probability of secondary incidents. Variables affecting these uncertainties include the time when incidents occurred, the number of vehicles involved in the incident, incident type, weather pattern, pavement conditions, the age and gender of drivers, whether there were turnovers, presence of vehicle fires, and corresponding traffic conditions. Through analysis on large quantities of incident data, the prior knowledge of these uncertainties is roughly estimated. As demonstrated in table 1, the severity statistics are summarized, with two categorical severity indicators represent the injury severity of drivers. Each data in the table indicates the casualties of every incident according to different occurring time and incident type.

Table 1. Summary of crash severity

Month	DI(A)	DI(B)	Time(h)	DI(A)	DI(B)	Incident Type	DI(A)	DI(B)
Jan	0.063	0.266	0~2	0.118	0.706	Crash head-on	3.000	2.500
Feb	0.143	0.270	2~4	0.167	0.333	Side crash	0.100	0.800
Mar	0.107	0.393	4~6	0.125	0.500	Rear-end	0.264	0.708
Apr	0.284	0.473	6~8	0.205	0.500	Opposing scratch	0.000	0.000
May	0.209	0.970	8~10	0.100	1.300	Parallel scratch	0.000	0.143
Jun	0.159	0.381	10~12	0.160	0.920	Hit fixture	0.000	0.246
Jul	0.345	0.476	12~14	0.071	0.357	Pedestrian	1.000	0.200
Aug	0.136	0.693	14~16	0.313	0.500	Turnover	0.182	1.394
Sep	0.152	0.565	16~18	0.000	0.400	Drop	0.000	1.000
Oct	0.137	0.562	18~20	0.100	0.700	Vehicle	0.000	0.500
Nov	0.121	0.345	20~22	0.500	0.600	Fogy	0.080	1.000
Dec	0.135	0.404	22~24	0.211	0.421	All	0.019	0.156

Bayesian Network Model

There are two methods for constructing BNs. One is by consulting experts, and the other is through data analysis. We applied the first method by consulting two experts in the field of freeway incident management. The aim of constructing BNs is to develop a graphical representation representing all the key factors that affect the successful operation of an incident management. This model was built with GeNie software as shown in figure 2.

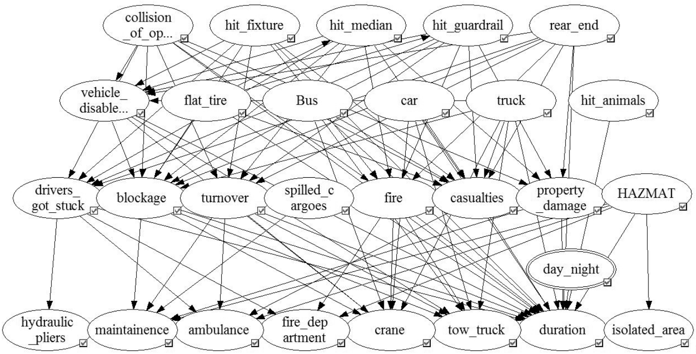


Figure 2. Freeway Incident Response Algorithm.

Comparative Analysis

Incidents data in the northern section of Nanjing-Lianyungang freeway of the year 2006 is used for a comparison between the calculating results and the actual data. Among the 48 sets of data of incidents with time information recorded, 29 sets of calculated data obtained from our model accorded to the actual situation while the left 19 sets of data had a comparatively larger difference from the actual situation. Table 2 demonstrates the comparison results of a set of data. Moreover, by establishing different Bayesian network structures (e.g., native Bayesian network), different calculating results may be obtained. The primary reason leading to some differences is that the quantities and qualities of data we obtained and applied are somewhat below the requirements. However, the general results are acceptable.

Table 2. Comparison between the calculating results and the actual data

Time slice	Duration		Severity			
	Fact(min)	Model			Fact	Model
30~60		0%	Driver	No injury		46.2%
60~90		10%		Slight	4	32.3%
90~120		45%		Serious		21%
120~180	145	43%		Fatal		0.5%
180~240		20%	Vehicle	Slight		3%
240~300		12%		Serious	Serious	62.2%
300~more		1%		Extensive		34.8%

4 Conclusions and Future Work

Focusing on the uncertainties in freeway incident response, an algorithm architecture for the uncertainties reasoning with BNs as the most critical part is proposed in this paper. Comparison between the calculated results from our model and the actual data implies that 60% of the simulation results are reliable.

References

Chowdhury Mashrur, Sadek Adel, Ma Yongchang, Kanhere Neeraj and Bhavsar Parth. (2006). “Applications of artificial intelligence paradigms to decision support in real-time traffic management". *Transportation Research Record,No. 1968*, TRB, National Research Council, Washington, D.C. pp. 92-98.

D. Spiegelhalter and S.L. Lauritzen. (1990). "Sequential updating of conditional probabilities on directed graphical structures". Networks, Vol.20. pp.579–605.

Stephen G. Ritchie. (1990). "A knowledge-based decision support architecture for advanced traffic management". *Transportation Research Part A: General*, Vol. 24. pp. 27-37.

Ghada Mahmoud Hamouda. (2004). "Risk-Based Decision Support Model for Planning Emergency Response for Hazardous Materials Road Accidents." *A thesis presented to University of Waterloo in fulfillment of the thesis requirement for the degree of Doctor of Philosophy in Civil Engineering.* Waterloo, Ontario, Canada, 2004.

A. Gupta, I.S.Mumick, and V.S.Subrahmanian. Maintaining views incrementally. In *SIGMOD* Conference, pages 157-166.1993.

Mingwei Hu. (2003). "Development of the real-time evaluation and decision support system for incident management". *Intelligent Transportation Systems, 2003*. Proceedings. IEEE, Vol. 1. pp.426 – 431.

Kun Zhang, Michael A.P. Taylor. (2006). "Effective arterial road incident detection: A Bayesian network based algorithm". *Transportation Research Part C*, Vol. 14. pp.403–417.

Kaan Ozbay, Nebahat Noyan. (2006). "Estimation of incident clearance times using Bayesian Networks approach". *Accident Analysis and Prevention*, Vol. 38. pp.542–555.

Shiliang Sun, Changshui Zhang and Guoqiang Yu. (2006). "A Bayesian Network Approach to Traffic Flow Forecasting." *IEEE Transactions on Intelligent Transportation Systems*. March 2006. Volume 7, issue 1. pp. 124-132.

Zhang Hongjun, Ritchie Stephen G. (1994). "Real-time decision-support system for freeway management and control". *Journal of Computing in Civil Engineering*, Vol. 8. pp. 35-51.

Nan Zou, Gang-Len Chang, and Saed Rahwanji. (2004). "Simulator: An integrated knowledge-base/simulation system for real-time incident management". *TRB* CD-ROM, Transportation Research Board, Washington, D.C.

Konstantinos G.Zografos, Konstantinos N. Androutsopoulos,George M. Vasilakis. (2002). "A real-time decision support system for roadway network incident response logistics". *Transportation Research Part C,*Vol. 10. pp. 1-18.

Research on Traffic Management For On-Street Intersection Connecting to Urban Expressway Off-Ramps

LONG Kejun[1], YANG Xiaoguang[2], ZHENG Jianlong[1]
1 School of Traffic & Transportation, Changsha University of Science & Technology, Changsha Hunan, China
2 Department of Traffic Engineering, Tongji University, Shanghai China

Abstract: The paper analyzes elevated expressway off-ramp connections to arterial street intersection. Direct left-turns and U-turn left turns were analyzed in terms of travel time and delay. VISSIM 4.1 simulation models for direct left-turn and indirect left-turns were constructed. The simulation tests were executed under two traffic conditions. Delay and travel time of straight movements in the U-turn scenario were reduced by 40-90 percent, while delay and travel time of the direct left-turn increased. Research also showed traffic volume and composition can affect the results of this comparison.

Key words: *access management, ramp terminal, U-turn, simulation, complex urban intersections*.

1 Introduction

Many urban expressways were built in China in recent years to alleviate congestion. Urban expressways provide benefit of high speed, safety and high capacity. However, the off-ramp terminals connecting expressways to local street intersections suffer from congestion, high accident rates and reduced capacity. Figure 1 illustrates this situation. The main problem here is insufficient separation distance between a ramp terminus and a downstream signalized intersection. The short separation could contribute to traffic conflicts and excessive congestion. This paper will focus on this type intersection, research the optimum traffic management method and control plan. The research deals with spacing of access points along the arterial/collector street system, access management.

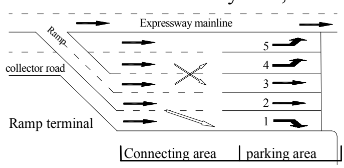


Figure 1. Connection between expressway ramp terminal and intersection

In urban areas, access management has proven an efficient method to improve signalized intersections in terms of signal spacing, lane marking, U-turn, provision of auxiliary lanes and median treatments. To simplify conflicts many American cities have introduced indirect left turns or U-turns. Indirect left turns are accomplished using median U-turns in advance or beyond the main intersection. The Michigan Department of Transportation [1] has provided median openings for U-turns on highways with wide-medians and prohibited all left-turns at major signalized intersections for many decades. The result indicated that capacity of the intersection was increased by 15 to 18 percent compared with dual left-turn lanes, and additional delay due to the increased travel distance of the indirect left turn was less than the conventional left-turn treatment. Hummer and Reid [2] studied a high-volume intersection, using CORSIM to simulate traffic performance and using SYNCHRO to develop optimized signal timings. Four time periods were analyzed, including morning, noon, midday (2:00 PM to 3:00 PM), and evening peak periods. They found that the median U-turn was superior to the two-way left-turn lane in capacity and corridor speed. Xiaokuan Yang [3] used CORSIM to compare the direct left turn with U-turn in terms of delay and travel time. Yang concluded that the U-turn was superior to the direct left turn in the high volume situations. Kentucky Transportation Center [4] studied New Circle Road using CORSIM, concluded that U-turn left turns enhance the operation of the corridor due to the more efficient processing of left turn vehicles at the downstream intersection. Also, they recommended that U-turns be considered for corridors with peak volumes greater than 1,500 vph or for cases where the expected total turn volume is greater than 20 percent of the total approach volume. Other researches [5,6,7] also proved superiority of U-turn left turn treatment for signalized intersections various perspectives.

Traffic mix at ramp terminal intersections is more complex than normal signalized intersections. Ramp terminal locations include ramp traffic merging with traffic from collector roads, within a short distance making left-turns, proceeding through and turning right at the main collector/arterial intersection.

This paper will apply basic access management principles to improve a complex intersection connecting a ramp terminal of the 2nd Ring Expressway and the Wanbao Arterial in Hunan Changsha city. This paper will evaluate the operation of two alternatives: direct left turn and indirect left turn (median U-turn) using VISSIM.

The following sections will discuss existing conditions of the intersection, improvement alternatives analyzed, simulation results using VISSIM for the two alternatives for the AM & PM peak periods and conclusions of the analysis.

2 Existing Conditions

2.1 Geometric condition of intersection

The study intersection is located in the northeast district of Changsha in Hunan province of China, intersecting roads include collector road of 2nd Ring urban expressway and Wanbao arterial road (Figure 2). Here, the 2nd Ring urban expressway and collector road is north-south bound, and Wanbao aterial road is east-west bound. The expressway is an elevated road lying over its parallel collector road, and connects to the collector road by on and off ramps.

Both north and south approaches have 5 lanes, 2 exclusive left-turn lanes, 2 straight lanes and one right-turn lane. Both east/west approaches have 3 lanes, one for left-turn, one for straight and one for right-turn movements dividedly. Two lane ramps are located 100 meters from stop line upstream of main intersection.

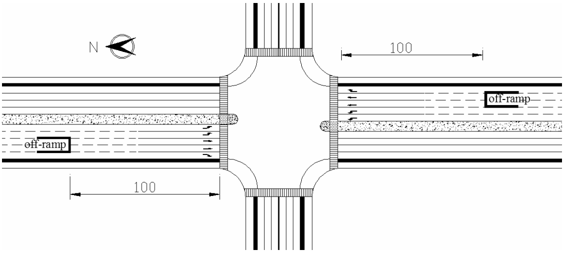


Figure 2. Geometric condition of the intersection

2.2 Traffic demand

This major expressway/arterial intersection carries heavy traffic. During the AM peak hour from 7:20 to 8:20, the total intersection volume was up to 6428 pcu/h, in which ramp in the north approach shared 557pcu/h and ramp in the south approach shared 689 pcu/h. During the PM peak hour from 17:55 to 18:55, the whole traffic volume was up to 6562 pcu/h, in which ramp in the north approach shared 573 pcu/h, and ramp in the south approach 742 pcu/h. Traffic volume data is included in Table 1.

The composition of traffic was: large vehicle 22.5%, car 39% and others 38.5%.

Table 1. Traffic volume of the intersection (pcu/h)

Approach \ Direction		AM peak hour (7:20 - 8:20)			PM peak hour (17:55 - 18:55)		
		Left-turn	Straight	Right-turn	Left-turn	Straight	Right-turn
East		238	580	253	333	382	127
West		448	608	341	298	356	549
North	Collector	521	576	306	70	969	269
	Ramp	207	229	121	30	425	118
South	Collector	462	552	297	326	1318	250
	Ramp	243	290	156	128	516	98

3 Evaluation of Direct Left-turn vs. Indirect Left-turn

3.1 Direct left-turn

Advantages of the direct left turn alternative include[6]: (1) For low traffic volumes, delay and travel time may be less than the indirect left-turn alternative; and (2) Vehicles making direct left turns travel less distance than the indirect left-turn. Disadvantages of the direct left turn alternative include: (1) Traffic delay and travel time may greatly increase under high traffic volume conditions; (2) Four-phase timing plans will result in more signal time loss; (3) Left-turn vehicles from the ramp and straight vehicles from the collector road must weave to enter the their destination lane; (4) Increased volumes and queues may extend to expressway mainline.

Based on above, the total travel time can be defined by the following equation:

$$T_L = t_{L1} + t_{L2} \tag{1}$$

Where: T_L – average total travel time of direct left-turn alternative,
t_{L1} – average delay of vehicles due to signal,
t_{L2} – average delay of vehicles due to interweave.

3.2 Indirect Left Turn Alternative

Advantages of the indirect left turn include: (1) Travel time is shorter, delay is reduced and capacity is increased. For distances of less than 0.5 mile the indirect left turn will be more effective because the travel time will be comparable with the travel time of the direct left turn. (2) High left turn volumes requiring left turn phases with long green times may reduce intersection capacity and increase the delay of the though movements, U-turns will improve the traffic flow conditions by reducing the vehicle travel time. (3) Interweave between left-turn movements from ramp and straight movements from collector road can be eliminated due to left turn being relocated to the U-turn beyond the main intersection.

Disadvantages of the indirect left turn include: (1) Higher delay compared to the direct left turn alternative if major road traffic volume is low; (2) Longer travel distance may increase fuel consumption when compared to a direct left turn.

Based on above, the total travel time can be defined by the following equation:

$$T_S = t_{S1} + t_{S2} \tag{2}$$

Where: T_S – average total travel time of indirect left turn alternative,
t_{S1} – average delay of vehicles due to the signal,
t_{S2} – average delay of vehicles at the U-turn median opening.

4 Simulation Models

VISSIM 4.1 was used to simulate the traffic operation of the two alternatives for AM peak and PM peak traffic volume scenarios.

4.1 Simulation model of direct left-turn

To model real traffic operations, the following assumptions were made: (1) The five approach lanes for north/south are numbered in sequence increasing from curb to median, the right turn lane is lane 1, as shown in Figure 1. (2) At the north/south approach, left-turn movements from the collector road will enter the approach lane 5 without disruption; straight movements from the ramp can enter the approach lane 2 without disruption; left-turns from ramp and straight from collector road must interweave using lanes 3 and 4; the area between ramp terminal and stop-line of intersection is a classic A-type weaving section. (3) Overflow was not evaluated and will not extend to upstream intersections. (4) The four-phase timing plan provides protected left-turn phases for both N/S and E/W.

4.2 Simulation model of indirect left-turn

In the indirect left-turn alternative, the following assumptions were made: (1) The three-phase signal provides a protected phase for east/west left turns. (2) Left-turn movements of north/south approach are indirect, and left-turns of east/west approach pass through the intersection directly. (3) The indirect left turn (median U-turn) is designed on the parallel collector of the 2nd Ring expressway. Hummer and Reid [2] pointed out that the optimum distance from the opening of U-turn to intersection is about 180m. The distance for this analysis was designed to be 250m. The median U-turn model is illustrated in Figure 3.

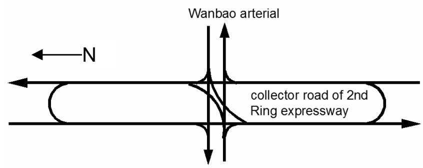


Figure 3. Scenario of indirect left turn (median U-turn)

5 Results and Discussion

The operational effects of the two alternatives were compared using travel time and delay. Tables 2-5 shows the simulation results of the two alternatives for AM and PM peak. Alternative 1 analyzed direct left-turns with a four-phase signal, and Alternative 2 analyzed indirect left turns with a three-phase signal. Data was collected for left-turn and straight movements (Table 2). North/south approaches were separated: expressway collector road and expressway off-ramp.

Table 2. Comparison of Travel Time (seconds)

Approach \ Direction		AM peak hour (7:20 - 8:20)				PM peak hour (17:55 - 18:55)			
		Left-turn		Straight		Left-turn		Straight	
		Alt. 1	Alt. 2	Alt. 1	Alt. 2	Alt. 1	Alt. 2	Alt. 1	Alt. 2
East		13.9	6.1	14.7	6.5	15.8	6.8	16.9	6.5
West		11.9	5.8	15.2	6.8	14	10.2	23.5	8
North	Collector	87.8	141	91.5	30.7	97.5	129.1	83.7	21.2
	Ramp	95.6	150.8	45.5	25	83.5	112.3	54.5	18.1
South	Collector	68	130.2	95.5	34.8	79.8	111.1	89.4	29.9
	Ramp	75.6	139.1	89.1	28.5	79.8	89	61.4	18.8

Note: "collector" refers to the 2nd Ring expressway collector road, "ramp" refers to off-ramp of expressway.

From Table 2, we can find that: (1) Compared to Alternative 1, travel time of straight movements at any approach was reduced by 45-75 percent during AM and PM peak hour. Travel time was less for Alternative 2 due to fewer signal phases. (2) Travel time of left-turns on east/west approaches was reduced by 27-57 percent. (3) Travel time of left-turn on north/south approaches increased by 11-91 percent. The increased travel time is caused mainly by the additional distance of required by the indirect left-turn.

Table 3. Comparison of Delay (seconds/vehicle)

Approach \ Direction		AM peak hour (7:20 - 8:20)				PM peak hour (17:55 - 18:55)			
		Left-turn		Straight		Left-turn		Straight	
		Alt. 1	Alt. 2	Alt. 1	Alt. 2	Alt. 1	Alt. 2	Alt. 1	Alt. 2
East		23.3	9.1	26	8.1	11.1	6.3	11.4	5.9
West		88.9	20	27	9.2	11.5	6.4	18.3	6.2
North	Collector	72.1	124.9	75.1	10.1	63.1	77.7	66.5	8.8
	Ramp	77	125.9	31.8	10.1	69.8	80.9	42	6.5
South	Collector	42.4	115.9	83.1	10.1	56.1	48.1	72.9	16.5
	Ramp	64	110.8	72.4	10.1	77.4	35.4	48.7	7.8

From Table 3, we can discover: (1) Compared to Alternative 1, delays of straight flow in Alternative 2 were reduced by 48-87 percent during AM and PM peak hour. Delay was less for Alternative 2 due to fewer signal phases. (2) Delays of left-turn in Alternative 2 were reduced by 43-77 percent compared to Alternative 1 for the east/west approaches. (3) During AM peak hour, delays of left-turns at the north/south approaches increased by 64-170 percent.

Overall, we find: (1) Straight movements were improved by the indirect left turn. One reason is the reduction in the number of signal phases, the other is that weaving movements between flow on the expressway collector road and flow from expressway off-ramp are reduced due to the indirect left turn. (2) Indirect left-turns caused additional travel distance to complete the left-turn, and may add travel time. However, when the overall effect is evaluated this additional time can be cancelled out by its efficiency. (3) Traffic conditions can affect the operational efficiency of the indirect left-turn. From the results of AM and PM peak hour, we can conclude that traffic volume and traffic composition are key factors.

6 Conclusions

This paper applied the method of access management to a complicated signalized intersection. Indirect left turns were used to replace the direct left-turn. Using VISSIM 4.1, the direct left-turn and indirect left-turn were compared using travel time and delay. The results indicated that indirect left-turns can improve the operation quality to some extent especially on the straight movements. Traffic volume and mix significantly affect the indirect left-turn efficiency.

7 Acknowledgements

This work is supported by the China Natural Science Foundation (50608010).

References

[1] Herbert S. Levinson. Indirect left turns－the Michigan experience [C]. In：*4th Annual Access Management Conference*，Portland, OR 2000.

[2] Hummer, J.E. and Reid, J.D. 2000 “Unconventional left-turn alternatives for urban and suburban arterials”, *Transportation Research E-Circular*[R], Transportation Research Board, PP. E-3/1-E-3/17.

[3] Yang Xi.. CORSIM-based simulation approach to evaluation of direct Left turn v.s. right turn plus U-turn from driveways [D]. Univ. of South Florida，2001.

[4] Kentucky Transportation Center. U-turns at signalized intersections[R]. KTC-04-12/SPR258-03-3F, 2004.

[5] Levinson, H.; Koepke, F.; et al, Indirect Left Turns—The Michigan Experience[C], In：*4th Access Management Conference*, Portland, OR 2000.

[6] Huaguo Zhou, Jian John Lu, Nelson Castillo, Operational Effects of a Right Turn Plus U-turn Treatment as an Alternative to a Direct Left Turn Movement from a Driveway[C], In：*4th Access Management Conference*, Portland, OR 2000.

[7] Bared, J.; Kaisar,, Median U-Turn Design as an Alternative Treatment for Left Turns at Signalized Intersections, *ITE Journal*, Vol. 72-2, Feb 2002, pp. 50–54.

[8] Topp, A; Hummer, J E, Comparison of Two Median U-Turn Design

Alternatives Using Microscopic Simulation[C], *3rd International Symposium on Highway Geometric Design*, 2005.

Experimental Research on the Effects of Oxygen-Enriched Air on Combustion in a Small Spark-Ignition Engine

Shengqin Li, Qiang Guan，Wenhui Zhang

Shengqin Li, Traffic College, Northeast Forest University. Harbin, Heilongjiang Province, China（150040） Tel:+86-451-82191004. E-mail: lishengqin@126.com

Abstract - An investigation on the effects of oxygen-enriched air on engine performance was proposed. This investigation focuses on combustion performance. As the oxygen concentration of the air for combustion increases, the heat of combustion for an engine of the same size would increase. The size (displacement) of an engine can be effectively reduced to achieve the same power. Oxygen- enriched air will affect the combustion process and the overall engine thermodynamics. For example, as oxygen concentration is increased, under the same operating conditions, combustion temperature and cylinder pressure will be increased. The purpose of this study was to quantify these results for a range of operating conditions.

Results include detailed thermodynamic output of pressures and other properties as related to the oxygen concentration of the air used for combustion. Results also include engine performance parameters such as power, torque, fuel consumption, and thermal efficiency. In one case, the combustion performance of the engine was recorded for an equivalent load with different oxygen concentrations of air. For air enriched with 24% oxygen, the cylinder pressure was increased by about 15%, the heat of combustion increased, and both the maximum pressure and the increased heat of combustion were achieved more quickly. Other results were recorded.

I. Introduction

A car is driven by the power that is released in the form of heat energy after the combustion of fuel in the engine. However, this same process of combustion releases pollutants to the atmosphere. The rising costs of petroleum, combined with the desire to reduce the pollutant effects of combustion have led to a finding ways to increase the efficiency of this combustion process. Thus, the latest research by experts and institutions around the world is being driven towards ways to improve power, decrease fuel consumption and reduce pollution from the combustion engine.

In the combustion process of an engine, full heat energy can only be released when all of the fuel has burned. Supplying sufficient fuel to the combustion chamber is easy enough; it is difficult to provide sufficient oxygen to burn all of the fuel in the combustion process. Therefore, it can be said that the main factor in determining the power an engine can produce is the amount of air (oxygen) available for combustion in the chamber, not the amount of fuel. Thus, it is very important to research a way to apply a technique, with regards to engine combustion, to increase the oxygen content in the intake air that will satisfy the minimum requirement of the amount of oxygen necessary to achieve a complete combustion of fuel without changing the quantity or quality of engine exhaust, and will also improve the power and economy of the engine as well. Theoretical analysis shows that when air is introduced into the combustion chamber, and when the air contains the optimum ratio of oxygen, the heat of combustion is increased and the rate of heat release is improved. The efficiency of the engine is improved while the hazardous gases resulting from the combustion process are reduced compared with the combustion resulting from the use of common air of the same mass.

The power output of a gasoline engine is lower than that produced by a diesel engine. There are more hydrocarbons (*HC*) and carbon dioxide (CO_2) present in the exhaust produced in a gasoline engine, causing more pollution to the environment, especially since a catalytic converter does not function when an engine is first started and the hazardous gases in the exhaust can not be effectively degraded effectively.

With a focus on the operating characteristics of a gasoline engine, an experiment was performed on a single-cylinder carbureted gasoline engine, with oxygen-enriched air. The experimental results were analyzed and are discussed in this paper.

II. Experimental System and Procedures

The engine used in the experiment is the Honda Model G200, a single-cylinder carburetor engine. The technical specifications of the engine are shown in Table 1. A schematic of the experimental test system is shown in Figure 1.

A pressure sensor was installed on the engine cylinder head and a combustion analyzer was used for combustion analysis. Meanwhile, additional data collection and analysis tools were applied to collect and analyze the parameters of oxygen concentration in the entering air, engine speed, engine power, exhaust temperature and fuel consumption. Finally, the performance of combustion of gasoline combined with oxygen-enriched air is analyzed.

During the experiment, an oxygen cylinder (oxygen bomb) was used to supply oxygen to the engine. The oxygen is introduced the combustion chamber of the engine directly through the intake manifold by taking advantage of the negative pressure produced when the engine intakes air. With an oxygen analyzer placed in the engine intake manifold, analysis is performed to measure the oxygen concentration of the air that has been introduced into the engine cylinder to support combustion. At the same time, the amount of oxygen flow is controlled by a flow control valve. The combustion process of the engine is then analyzed at oxygen concentrations of 21% and 24%, respectively.

The purpose of this experiment is primarily to monitor internal combustion, power and torque, as well as tail gas emission of the engine when oxygen concentrations are increased. The analysis and research provided in this paper is focused on the combustion process of the engine cylinder only.

III. Results and Discussion

A DEWE-800 combustion analyzer was used to measure the output of the engine with combustion supported by air in different oxygen concentrations. Through mathematical analysis of the results, maximum combustion pressure and the maximum rate of change of pressure, and location of the crank angle at each occurrence are calculated. Additionally, the parameters of cyclic variations of combustion for different air conditions are analyzed. Heat release law for various intake air conditions is determined through thermodynamic analysis of the results. Further, analysis is performed on the influence of engine combustion characteristics by the oxygen concentration in the entering air. The combustion process of the engine is studied primarily through the following four parameters.

A. Analysis of Combustion Pressure

Analysis is performed regarding the maximum combustion pressure, P_{max}, and the position of the crank angle at P_{max}. A corresponding relationship exists between the maximum combustion pressure and an engine's thermal efficiency and power. The higher the maximum combustion pressure, the greater is the indicated power and the better the combustion status. However, limited by engine strength and mechanical efficiency， the maximum combustion pressure should not be too high. The maximum combustion pressure and the crank angle are used to show whether the combustion of engine is in time or not.

Figure 2 shows the result of the combustion process when oxygen concentration is 21% and 24% respectively, from which it may be found that, as the oxygen concentration in the intake air is increased, the maximum value of cylinder pressure is increased, and the time that the maximum value of pressure is reached is advanced. When oxygen concentration is 21%, the maximum cylinder pressure measured was 21.239 bars. At that moment, the crank angle was 117°. When oxygen concentration was increased to 24%, the maximum cylinder pressure measured was 24.493 bars， and at that instant the crank angle was 63°. The relative increase in cylinder pressure is not very much, but the maximum pressure value was reached earlier with the increase in the ratio of oxygen. This occurred because the fuel was able to combust more completely due to the increase in oxygen concentration. The amount of heat produced was also increased, which led to an increase in cylinder pressure. Additionally, the fuel was able to combust faster and flame velocity was reached more quickly, therefore the time of maximum pressure was reached earlier.

B. Study of the Rate of Change of Cylinder Pressure

The maximum rate of increase of pressure and its corresponding crank angle were recorded. The maximum rate of pressure increase is related to the rate of combustion. If the rate of combustion is too fast, the engine will run very roughly, while a combustion rate that is too slow will result in poor power and poor fuel economy. Oxygen in the air at intake has a large influence on the rate of combustion and the flame velocity. Therefore, the research shows there is justification to use changes in the maximum rate pressure increases to reflect the influence of oxygen concentrations in the intake air on the properties of combustion, and the characteristics of an engine.

Figure 3 shows the rate of change of the cylinder pressure during the combustion process when the oxygen concentration reaches 21% and 24%, respectively. From this, it can be seen that the maximum value of the rate of change of cylinder pressure is 0.789 bar/°CA and the corresponding crank angle is 107°, when the oxygen concentration in the intake air is 21%. The maximum value of rate of change of cylinder pressure rise rate is 1.287 bar/°CA and the corresponding crank angle is 58°, when the oxygen concentration in the intake air is 24%. As the same, due to the increase of oxygen entering the cylinder, the fuel combusts more quickly and the flame velocity is increased, both of which have lead to an increase in rate of change of pressure and a quicker time to reach the moment of maximum value the rate of change of pressure. Accordingly, in order to reduce the deflagration trend of engines, the ignition advance angle should be delayed or postponed appropriately when oxygen concentration in the intake air is increased.

C. Study of Heat Release

Heat Release Law is achieved by calculation from the measured indicator according to the thermodynamic *Law of Conservation Of Energy* and the *Law of Conversion of Energy* under a given hypothesis. It may reflect more accurately the property/characteristics of a combustion process. Therefore, it is an effective way to evaluate the combustion process of an engine under different concentrations of oxygen in the intake air, and to research and analyze the combustion process in an engine by using the heat release law. In this paper, the combustion performance of an engine is influenced by oxygen concentration in the intake air, and two aspects of the combustion process are analyzed: heat release and the rate of heat release.

Figure 4 shows a comparison of heat release during the fuel combustion process at oxygen concentrations of 21% and 24% respectively. From this, we find that the maximum value of heat release of fuel during the process of combustion is 1.858%/°CA with a corresponding crank angle of 89.5° when the oxygen concentration is 21%. And the maximum value of heat release of fuel is 2.759%/°CA with a corresponding crank angle of 58° when oxygen concentration is 24%. Meanwhile, the initial time/moment of heat release is advanced and the maximum value of heat release due to combustion is increased. Due to an increase in oxygen concentration, same amounts of fuel will combust more completely， or the fuel sprayed into the cylinder within the same amount of time is increased and such leads to an increase of heat released during the combustion process. As a result, due to the increase of oxygen concentration, the fuel can be combusted more rapidly, the flame velocity becomes faster and initial time of heat release is advanced.

D. Study of the Rate of Cycle Alteration During the Combustion Process

The cyclic variations of combustion are constantly changing during the combustion process between the first circulation and the next when an engine is operating under given stable working conditions. Its result is that the pressure curve, flame propagation and engine's power output are different. Cyclic variations in combustion have an unfavorable influence on the engine and lead to overall deterioration of engine's performance in power, fuel economy, emissions and maneuverability.

There are many factors that influence cyclic variations during the combustion process, of which the major factors include the cyclic variations of air movement in cylinder, the location of the flame center formed after spark ignition, the variations in the flame's initial growth rate due to changes in speed and direction of airflow velocity, and the content of gas mixture (especially the content adjacent to the spark plug at the moment of ignition).

The cyclic variation rate of maximum combustion pressure is used as the major parameter for evaluating cyclic variation in combustion. The equation for calculating the cyclic variation rate is:

$$\delta = \frac{P_{\max STD}}{\overline{P_{\max}}} \times 100\% = \frac{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(P_{i\max} - \overline{P_{\max}})^2}}{\frac{1}{n}\sum_{i=1}^{1} P_{i\max}} \tag{1}$$

where P_{maxSTD} is the standard deviation of maximum combustion pressure, $\overline{P_{\max}}$ is the averaged value of the maximum combustion pressure of n continuous cycles. P_{imax} is the maximum combustion pressure of cycle i, n is the number of samples. The sample length during the experiment for this paper is 120 continuous cycles.

During the experiment, cylinder pressure was recorded for different oxygen concentrations of intake air by making use of a spark plug type pressure sensor made by the Kistler Company. Maximum combustion pressure was recorded for each cycle using a DEWE-800 Analyzer. Using these results and equation (1), the cyclic variation rate of combustion under different oxygen concentrations of intake air may be calculated.

Using equation (1), we may also calculate the cyclic variation rate of combustion under different concentration of oxygen in intake air for idle operating conditions, as shown in Figure 5.

It can be seen from the Diagram that the amount of cyclic variation of combustion, due to variations in engine cylinder pressure, is increased. More specifically, the stability of the combustion process caused by cyclic variations is increased, although parameters such as the gas pressure in the cylinder, rate of pressure increase, average measured pressure, and the amount of heat released due to combustion are all increased, and the engine exhibits a deflagration trend, when oxygen concentration in the intake air is increased to 24%. This shows that the cylinder pressure in an engine is increased when the air for combustion is oxygen-enriched and the engine is stable during the combustion process. Therefore, by enriching the oxygen-content of air for combustion, combined with an advance of the ignition angle delay may combust the fuel more fully, may improve the rate of heat release, and may reduce the deflagration trend of an engine, all while improving the stability of the combustion process.

IV. CONCLUSIONS

1. The experiment was performed as described. The cylinder pressure increased and maximum pressure was reached earlier. The rate of increase in cylinder pressure increased and also was reached earlier. All of this occurred when the oxygen concentration in the intake air was 24%.

2. The heat of combustion of the fuel was increased and was reached earlier, when the oxygen concentration in the intake air was 24%.

3. The cyclic variation of combustion within the engine was reduced and the operation of the engine was more stable when the oxygen concentration in the intake air was 24%.

REFERENCES

Jerald A. Caton. “Use of a Cycle Simulation Incorporating the Second Law of Thermodynamics: Results for Spark-Ignition Engines Using Oxygen Enriched Combustion Air.” SAE：2005-01-1130

John B. Heywood, “Internal Combustion Engine Foundamentals,”*J.* McGraw-Hill, 1988, pp. 158-176.

Shengqin Li, “The Study of the Oxygen-Enriched Combustion to Diesel Engines”. Journal of Northeast Forest University. 2006.2: 45-46

Appendices:

Table 1 Test Engine Specifications

Item	Value Used
Engine Displacement:	193 cc

Bore× Stroke:	68 ×45 mm
Compression Ratio:	8.5 : 1
Maximum speed	3600rpm
Maximum Power:	3.7 kW
Maximum Torque:	10.4 Nm

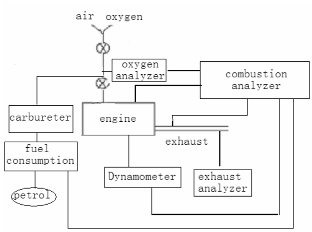


Fig. 1. Experimental Setup

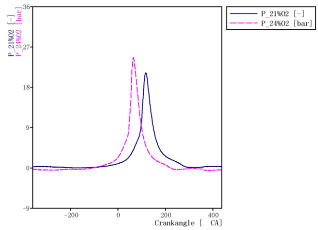


Fig. 2. Contrast of Pressure Variation

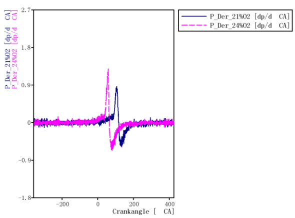


Fig. 3. The Rate of Change of Cylinder Pressure

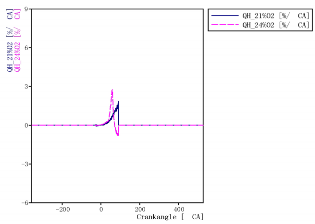


Fig. 4. Combustion as a Function of Crank Angle

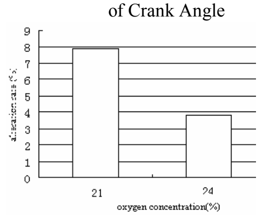


Fig. 5. Change in Rate of Combustion

Evaluation of Effectiveness of Dynamic Traffic Guidance: A Case Study of VMS Application

By SHAO Haipeng[1] & LIU Haode[2]

[1]Ph.D, Department of Traffic Engineering. Highway College, Chang'an University, Xi'an, China. P.O. Box 487, Middle of South 2nd Ring Road, Xi'an, 710064, P. R. China; PH (8629) 8233-4886; FAX (8629) 8233-4432; Email: shaohp@126.com

[2]Ph.D, School of Transportation Engineering, Tongji University, 4800 Cao'an Road, Shanghai, 200092, China; PH&FAX (8621) 6958-8994; Email: beuture@163.com

Abstract

Based on previous information about drivers' behavior, six types of driver Route Choice Behavior were studied. Using the VISSIM simulator, the effectiveness evaluation models of each traveler type was developed. The simulation-based model evaluated various types of guidance information; including dynamic route guidance using variable message signs (VMS), under the same travel demands. The results indicate that VMS can reduce travel time of related roads and VMS can markedly increase the network speed.

***Key Words*:** Advanced Traveler Information System (ATIS); VMS;Traffic Simulation; Evaluation of Effectiveness; VISSIM;

1. Introduction

The improvements in advanced traveler information systems (ATIS) have led to many applications. But does the information result in more efficient travel? How do the travelers respond to the information? How does the information impact the traffic system? A simulation based method can be used to evaluate the effectiveness of dynamic route guidance.

Kawashima (1991) and Durand-Raucher (1993) conducted research on drivers' behavior downstream of a VMS. Hato (1995) and M. Wardman (1997) compared travelers' response to different VMS information using a stated preference survey, and found that the effectiveness of the information depended on content, environment and traveler's characteristics (age, gender and road network familiarity, etc.).

Analytical models can calculate travelers' behavior. LI (2005) developed a VMS traffic diversion rate model based on Multi-User Class-Dynamic Traffic Assignment, but the analytical model was limited due to the complexity of traffic system. James (2003) developed a macroscopic model using TRIPS, a transportation planning software, where travelers are routed based on minimizing travel time. Because TRIPS is a macroscopic model tool, it is inferior in describing route choice behavior and traffic management.

2. Traveler Types

Our research involved VISSIM, a microscopic traffic simulation software, to model measures of effectiveness of VMT on a road network. The model is based on defined traveler classifications with access to ATIS. Figure 1 describes the various traveler classifications, based on type of ATIS available and information source, type and amount.

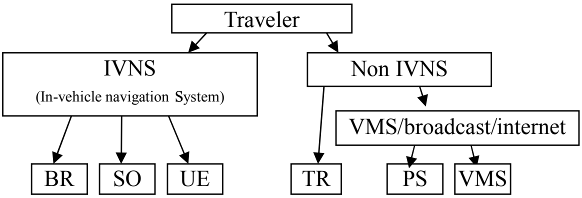


Figure 1: Traveler Classifications

Sun (2006) divided travelers into two groups: those with In-Vehicle Navigation System (IVNS) and those without (Non IVNS). IVNS travelers included 3 types: Bounded Rational (BR), System Optimal (SO) and User Equilibrium (UE). Non-IVNS travelers also included 3 types: TRaditional (TR), Pre-Specified (PS) and VMS.

Our simulation focused on the three types of Non IVNS travelers.

TR traveler ignores all information before and during the trip, so the route choice method is to find the shortest route by experience.

PS traveler receives pre-trip and on route information from media sources, internet, phone updates and VMS along their route so they follow the Time Depended Shortest Path (TDSP) per Sun (2006).

VMS traveler is provided advisory, warning and regulatory information on route. The key in modeling the route choice behavior is a traveler's compliance; therefore a binary model was applied.

3. VISSIM Model

We used VISSIM 4.1 which was available with a component object model (COM) interface, so it was possible for the exterior application to call VISSIM (PTV AG, 2004). The model is appropriate in a high density road network which allows travelers to change routes freely. In the model, the probability of a traveler *(n)* changing routes, whose route contains link (*a)* with a VMS (*c)* is as follows:

$$P(n) = \frac{1}{1 + \exp[\alpha - (\beta \cdot LOF_n + \gamma \cdot LOC_a + \delta \cdot VMS_c + \varepsilon \cdot LOE_n)]} \quad (2)$$

where: $P(n)$——Route transfer probability，$P(n) \in [0, 1]$；

LOF_n——Familiarity with road network by traveler n，$LOF_n \in [0, 1]$；

LOC_a——Status of traffic flow on link a，$LOC_a = v/v_{free}$，$LOC_a \in [0, 1]$；

VMS_c——Confidence of traveler with VMS information c , $VMS_c \in [0, 1]$ ；

LOE_n——Imminence of traveler n to arrive at destination , $LOE_n \in [0, 1]$ ；

$\alpha, \beta, \gamma, \delta, \varepsilon$——coefficient。

3.1 Program Application

The application included three parts: microscopic simulator, control module and route choice module, as shown in Figure 2.

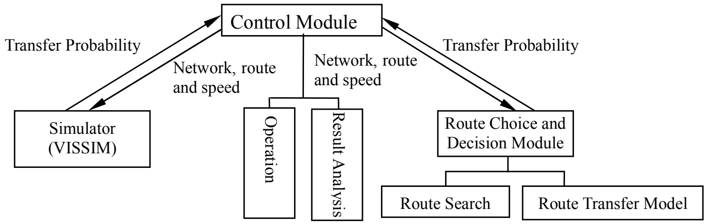


Figure 2: Logic Modules of Evaluation Program

The control module is the centre of the program. It controls the simulator and route choice module, helps exchange data and analyzes the result. Vehicles' routes were initialized, then network speed and link flow rate were calculated. Route choice and decision module receives vehicles' transfer probability from the control module and regenerates vehicles' routes. Route choice and decision module has two functions: searching for the best route according to route guidance information available and calculating the transfer probability. The two functions stay balanced via the control module. The flow chart is shown below.

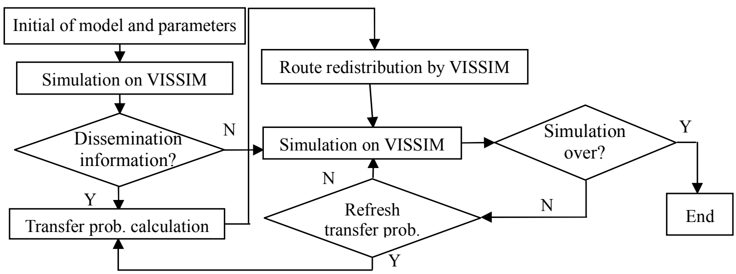


Figure 3: Program Flow Chart

3.2 Indicators

To test how information influences travelers and the traffic system under different traffic flow conditions, the following traffic network performance indicators were calculated (James Y. K, 2003).

$$\text{Network distance} = \sum_{k=1}^{K}\sum_{a} q_a(k) l_a \qquad \text{(km)} \qquad (3)$$

$$\text{Network travel time} = \sum_{k=1}^{K}\sum_{a} q_a(k) t(k) \qquad \text{(h)} \qquad (4)$$

$$\text{Network speed} = \frac{\text{Network distance}}{\text{Network travel time}} \quad (\text{km/h}) \tag{5}$$

where: $q_a(k)$ —— Flow rate of link a at time segment k, (k=1,2,3,…)

l_a —— length of link a (km)

$t_a(k)$ —— travel time on link a at time segment k (h)。

4. Case Study

The case study was based 8 intersections in Hefei where the network allowed vehicles to transfer freely. Field traffic data was collected during the peak hour, 7:30~8:30. Link traffic volumes and VMS location are shown in Figure 4. The network consisted of two major arterials and four road sections. Road sections 1 and 2 are the opposite flows of the east-west arterial, and road sections 3 and 4 are two opposite flows of the north-south arterial.

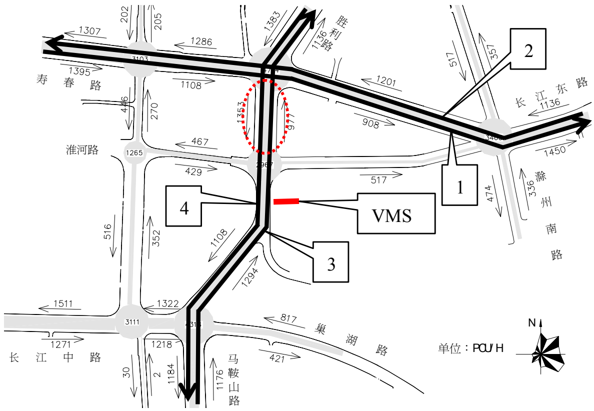


Figure 4: Traffic Volumes on Links and VMS Location

The network model was established in VISSIM, and calibrated with field data (LIU, 2006). Route choice module calculated alternative routes and extracted volume and speed of routes from VISSIM. Together with parameters from the field data, route transfer probability was calculated. Then the flow rate of each alternative route was recalculated according to transfer probabilities and sent to VISSIM simulator.

4.1 Experimental Design

The road network performance indices were calculated and recorded under normal demand. The traffic flow characteristics of two highlighted arterials were studied simultaneously along with the impact of a VMS on the network. Then, the effectiveness of VMS under different travel demands (85%, 90%, 105%, 110% and 120% of normal) was compared to the normal condition to study the sensitivity of road network to VMS. A comparison of travel times on the four road sections with and without VMS is shown in Figure 5.

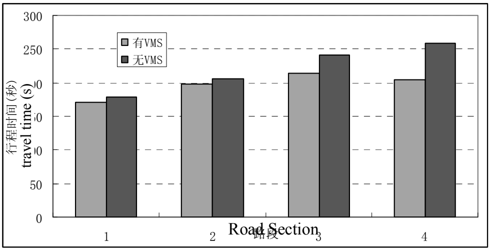


Figure 5: Comparison of Travel Times with and without VMS

4.2 <u>Results</u>

The model results showed that VMS reduced travel time on the arterials, especially in road section 3, where the VMS is located, from 258.5 hr (non-VMS) to 204.1hr (VMS) or by 21%. The travel time on other sections also decreased but by a smaller percentage.

According to Table 1, under the normal traffic demand, the network speed increased from 19.616 to 20.245 (km/h) or by 3.2% with the VMS. In addition, the total network travel time was reduced from 203.792hr to195.837hr or by 3.9%.

Table 1: Operation Indexes with VMS under Different Traffic Demands

No.	Traffic Demand	Difference in speed (%)	Network Speed (km/h)		Total Time(h)		Network Distance (km)	
			VMS	Non-VMS	VMS	Non-VMS	VMS	Non-VMS
1	0.85	0.0315	19.624	19.025	224.493	230.079	4405.451	4377.253
2	0.9	0.0296	20.486	19.897	193.923	198.18	3972.707	3943.187
3	1.0	0.0321	20.245	19.616	195.837	203.792	3964.72	3997.584
4	1.05	0.0345	20.023	19.355	207.852	213.221	4161.821	4126.892
5	1.1	0.0506	19.19	18.266	238.821	248.812	4582.975	4544.8
6	1.2	0.0601	12.074	11.39	387.891	403.545	4683.396	4596.378

When traffic demand increases (1.05, 1.1, 1.2), the average network speed decreases and the total travel time increases. The network speed with VMS is higher than that without VMS under the same conditions. The total network travel time with VMS is lower than that without VMS under the same demand.

When the traffic demand is lower than normal, the percentage difference between the network speed with VMS versus Non-VMS is smaller (3.15% and 2.96%) than when the demand is higher than normal. At traffic demands 1.1 or 1.2 above normal, the percentage difference between the network speed with VMS versus Non-VMS is 5.06% and 6.01% respectively.

In spite of different traffic demands, the dynamic route guidance provided by the VMS always improved the network performance indictors.

5. Conclusions

In this paper, a simulation based method was used to evaluate the effectiveness of various route guidance options and relevant models were developed. The VMS was used as a case study. Conclusions can be drawn as follows:

1) Travelers can be categorized by route choice behavior and information types.
2) Simulation based method can breakthrough the hypothesis for analytical model, and describe the network and travelers' behavior in more detailed.
3) Results with and without VMS indicate that VMS did reduce the travel time on the study arterials.
4) Results with and without VMS under various demand scenarios indicate that VMS did increase the average network speed, with greater benefits occurring as the congestion increased.

References

Durand-Raucher, Y., Yim, Y. and Ygnace, J. (1993). Traffic information and driver behaviour in Paris region. *Proceedings Pacific Rim Transtech Conference*, Seattle ASCE, Vol. 1., pp. 167-169.

Hato, E., Taniguchi, M. and Sugie, Y. (1995) Influence of traffic information on driver's route choice. Paper presented at *World Conference on Travel Research*, Sydney.

James Y. K. Luk and Chao Yang. Comparing Driver Information Systems in a Dynamic Modeling Framework [J]. *Journal of Transportation Engineering*, 2003, 129 (1): 42-50

Kawashima, H. (1991) Informatics in road transport: an overview of examples in Japan and future issues. Paper presented at *4th German-Japanese Seminar on Transportation Systems*

LI Shuo, FAN Bing-quan, MA Liang. Optimal CMS traffic diversion rate model based on MUC-DTA[J], *J. University of Shanghai for Science and Technology*, 2005 Vol.27(2):162-166. (In Chinese)

Liu Haode, Yang Xiaoguang, Sun Jian, Parameter Calibration for VISSIM Using A Hybrid Heuristic Algorithm: A Case Study of a Congested Traffic Network in China, *Proceedings of the International Conference on Applications of Advanced Technologies in Transportation Engineering*, 2006.8

Mahmassani, H.S. *Development and Testing of Dynamic Traffic Assignment and Simulation Procedures for ATIS* [R]. Technical Report DTFH6 1-90-R-00074-FG. 1994.

M. Wardman, P. W. Bonsall and J. D. Shires. Driver Response to Variable Message Signs: A Stated Preference Investigation. *Transpn Res*. C, Vol. 5, No. 6, pp. 389-405, 1997

PTV AG. Innovative Transportation Concepts, Inc. VISSIM COM Description of the COM interface VISSIM 4.00-12, 2004.

SUN Jian, Study on Traffic Network Simulation Key Technology Under ATIS[D], A dissertation for Ph.D degree, Tongji University, Shanghai, 2006.

Mesh Network Applications in First Respond Service, Case Study: Policy Simulation

Saini Yang[1]

[1]Assistant Professor, Cleveland State University, Department of Civil and Environmental Engineering, Stilwell Hall 122, Cleveland, OH 44115; PH (216) 687-2230; FAX (216) 687-6593; email: s.yang27@csuohio.edu

Abstract

The advantages of mesh network provide huge potential for its application in transportation network security and first response service. The applications of mesh network in transportation security will not only improve emergency response but also help the coordination and integration across agencies, regions, and public safety and transportation disciplines. The goal of this paper is to explore the potential applications of mesh network in emergency response service and to analyze the performance of various emergency dispatching policies using simulations. The case study shows that with the help of mesh network and integrated management, promising saving in average response time and reduction in percentage of long response times can be achieved. (110 words)

1. Introduction

A mesh network is a networking technique which routes data, voice and instructions between network nodes. A mesh network has its unique advantages. By sharing access to the higher cost network infrastructure, mesh technique can effectively improve the network efficiency. Since a mesh network is a fully connected network, it allows continuous connections and reconfiguration around occupied, broken or blocked paths. Therefore, the components in a mesh network can reach each other via multiple hops without queueing in traditional network. Another advantage is the high reliability. The network can overcome a node breakdown or a bad connection. These features provide huge potential for its application in transportation network security and emergency response service.

In an emergency response process, the emergency response time is an important index of event's negative impacts, especially for severe ones. The crucial response time can be separated into 4 phases: detection time, preparation time, travel time and treatment time. The applications of mesh network in emergency response system can help integrate across agencies, regions, public safety and transportation disciplines and shorten the response time. For instance, a mesh network application can provide access to operational data/resources. The data can be regional transportation data, the remote video, incident logs, road sensors, multiple state/federal law enforcement criminal databases, etc. With the help of Computer Aided Dispatch (CAD) and advanced GIS capabilities (e.g., incident/user GPS identification and aerial photo overlays), emergency response vehicles can be served in a more prompt and accurate manner.

Some pioneer projects are under development recently, such as the Capital Wireless Integrated Network (CapWIN) and Champaign-Urbana Community Wireless Network (CUWiN) project. These applications are helping police, transportation, and fire personnel access operational information and communicate across jurisdictions and disciplines. Rather than dispatching emergency response vehicles from separated agencies, e.g. police department, fire department, and paramedical department, it brings potential to have integrated fleet management. Further, it provides opportunity to apply new operation policy in emergency respond service.

The simplest emergency dispatching strategy is First-In-First-Out (FIFO) policy. Due to various emergency priorities, traditionally, the emergencies with higher priority are served first based on the Nearest-Origin (NO) policy. When mesh network is available, it is easier for the emergency vehicles to exchange real time information on route and it is possible to apply new policies to the emergency vehicles. For example, when a vehicle is assigned to an incident site, it can be directed to new destination when necessary or an idle vehicle may be routed to other stations to prevent unit deficiency by detailed voice, text or graphic instructions without causing potential confusion.

To test new emergency dispatching policies, simulation is a very useful tool. In the past 40 years, simulation models have been developed and commonly used to evaluate the performance of emergency response systems. Early simulation models (Carter and Ignall, 1970; Fitzsimmons, 1973; Ignall, et al. 1978) are based on the FIFO system. No queue is considered in the simulation and these simplifications greatly reduced the precision of the models. Goldberg et al. (1990) developed a multi-server queuing system on a FIFO basis without considering priority scheduling of calls. Later, they (1991) extended the previous work by allowing stochastic travel times, unequal vehicle utilizations, various call types, and service times that depend on call location. Response time was selected as the performance measure in Zografos (1992, 1994), where FIFO and NO are studied. Because of the real time traffic data can be shared in a mesh network, it is possible to enable dynamic routing feature in the emergency response system. Dynamic shortest path algorithm plays an essential role as the base criterion for on-line vehicle dispatching and routing.

2. Simulation Model

The efficiency of the response system is heavily based on the information. By using mesh network technologies, the fleet surveillance system can take full advantage of the real time traffic information, track the status and location of vehicles and help in developing vehicle assignment plans for new emergency calls. GIS maps can precisely locate the emergency site, and a GIS database can help establish the magnitude of life, property and effort involved, determine the risk zones based on land use data, and build activities in tune with the National Building Code guidelines. Integrated emergency vehicle fleet management can provide better planning of location and fleet, expedite accurate response to an emergency, as well as reduce operation cost. With better planning and operation guidance, the same crew and fleet will be able to handle the responsibility more efficiently and with higher performance levels.

To test the new policies, a simulation model is developed and the conceptual framework is shown in Figure 1. The simulation is driven by both events and fixed time steps (incremental time points). The time stamps of events refer to those times at which a vehicle changes its status or an emergency changes its status. The fixed time steps refer to the times at which the traffic information is updated. The time stamps of events and fixed time points are ranked and the earliest one is selected as the next simulation time point. In each simulation point, the program will update the emergency and vehicle information, such as : the current location, the route to take, the destination, the time stamp of next status change, current status, next proposed status, etc. The vehicle status is tightly related to the emergency status. Furthermore, some vehicle status changes may result in the reconsideration of the dispatching decision. For instance, if a vehicle finishes its task and it is on its way back to station, that means the vehicle is available at this point. In this case, this vehicle can be assigned to an emergency location or relocate it to another location for better coverage. Optimization is needed upon this event. The optimizer module makes decision about the movement of all vehicles according to an assignment strategy and the result of simulation. So it is the key module of the operation, receiving and processing all service calls and controlling all activities. The optimization model is discussed in Yang (2005).

In this simulation, real-time traffic information is assumed to be known, which includes the average traffic flow for non-peak hours, and AM and PM peak hours. For emergencies, the temporal distribution, spatial distribution and the priority distribution are calibrated from the real operation data. Ambulances, fire trucks and police cars are considered. 4 dispatching policies are considered: FIFO, NO, rerouting with and without relocation polices.

To get precise estimates of the system performance measures, it is important to appropriately analyze the simulation output. Because the output process from the steady-state distribution is considered, it is necessary to discard a specific transient time, which is called the warm-up period, in which the state of system is not yet stable. Based on preliminary experimental results, one day is used as warm-up period. To ensure an appropriate statistical analysis from simulation results, a number of simulation replications are necessary. The number of replications needed depends on the specified precision, degree of confidence and sample variance.

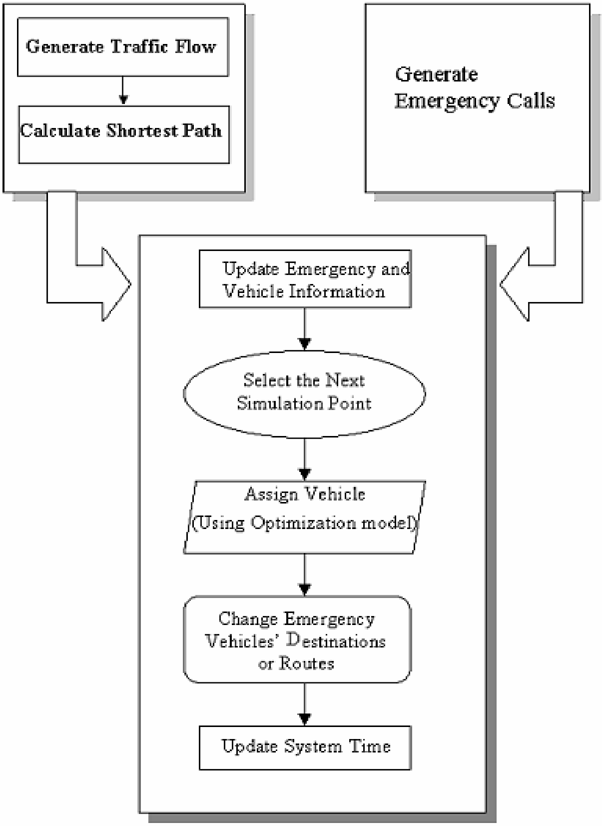


Figure 1. Conceptual simulation flow chart.

3. Case Study

The data used in this study are generated from real-world operational data for the ambulances and the medical units during November and December of 2000. More than 3000 records are analyzed. Each record stands for one dispatched vehicle. A series of variables describe various information associated with the vehicle, including the time at which the emergency call arrived, vehicle identification number, dispatching time, arrival time and call type. The road network of case study area is shown in Figure 2. The locations of emergency calls are matched to GIS locations to obtain the historical spatial distribution of emergency calls. ARENA Analyzer is used to find the best fitted distributions for historical data.

Figure 2. Road network of case study area.

To compare alternative dispatching strategies under different emergency frequency, we consider average response time, maximum response time and the ratio of emergency events whose response time exceeds the pre-specified response time limit to total accidents. The average response time is the main criterion to judge a dispatching strategy since it plays a crucial role in minimizing the adverse impacts. Several sets of problems were solved under four scenarios: (a) FCFS, (b) NO, (c) optimized dispatching without coverage concern, and (d) optimized dispatching with coverage concern. The same series of emergency calls were used in all scenarios. These problems were solved using CPLEX version 9.0. The results of ambulances' performance measures are shown as Table 1.

Table 1. Summary of computation results

Policy	Average Response Time (minutes)	Longest Response Time (minutes)	% vehicles with Response time longer than limit
(a)	8.63	20.10	21%
(b)	3.51	14.37	16%
(c)	3.07	9.78	14%
(d)	3.02	7.90	9%

It is obvious that the results for ambulances under NO is close to the results under optimized dispatching scheme, while under FCFS strategy the response time is much longer than the other scenarios. The average response time under policy (d) is about 2% less than that of policy (c). However, the number of vehicles that arrive at the emergency sites later than the maximum allowable waiting time for the emergencies decreases by 5%. This indicates that vehicle relocation allows more vehicles to reach emergency sites within the desired response time. For fire engines and police cars, similar patterns can be observed.

4. Conclusion

Application of mesh network in transportation network is a unique and challenging task. There are numerous directions for future research. In this study, a simulation model is developed to compare the performance of integrated emergency response fleet management with various emergency dispatching policies. Real world data is used in the case study and the results show promising benefits of potential mesh network application in emergency response. For future study, it is important to improve the security and reliability of the mesh network itself in order to enable real world operation of mesh network. How to efficiently store and pass information among multiple types of user, and how to simplify the interaction among users are another groups of questions waiting for transportation engineers and researchers.

References

1. Carter, G. and E. Ignall, "A Simultion Model of Fire Department Operations,." IEEE System Science and Cybernetics, Vol. 5, pp. 282-293 (1970).

2. Fitzsimmons, J., "A Methodology for Emergency Ambulance Deployment." Management Science, Vol. 19, No. 6, pp. 627-636 (1973).

3. Ignall, E.D., P. Kolesar, and W.E. Walker, "Using Simulation To Develop and Validate Analytic Models: Some Case Studies." Operations Research, Vol. 26, No. 2, pp. 237-253 (1978).

4. Goldberg, J., Dietrich, R., Chen, J., M. Mitwasi, "Validating and Applying a Model for Locating Emergency Medical Vehicles in Tucson, AZ." European Journal of Operational Research, No.49, pp. 308-324 (1990).

5. Goldberg, J. and F. Szidarovszky, "Method for Solving Nonlinear Equations Used in Evaluating Emergency Vehicle Busy Probabilities". Operations Research, Vol. 39, pp. 903-916 (1991).

6. Goldberg, J., and L. Paz, "Locating Emergency Vehicle Bases when Service Time Depends on Call Location, Transportation Science." Vol. 25, No.4, pp. 264-280 (1991).

7. Zografos, K., Douligeris, C. and C. Lin, "A Model for the Optimum Deployment of Emergency Repair Trucks: An Application in the Electric Utility Industry." *Transportation Research Record*, 1358, pp. 88-94 (1992).

8. Zogfrafos, K., Douligeris, C., and C. Lin, "A Simulation Model for Evaluating the Performance of an Emergency Response Fleet." Transportation Research Record, 1452, pp. 27-34 (1994).

9. Yang, Saini, Masoud Hamedi and Ali Haghani, "Online Dispatching and Routing Model for Emergency Vehicles with Area Coverage Constraints." Transportation Research Record, No. 1923, pp. 1-9 (2005).

Research on The Stability of Concrete Mattress Under Wave Action in the Yangtze Estuary Deepwater Channel

WU Su-shu[1], Don Bond[2]

1. College of Traffic, Hohai University, Nanjing 210098, China; email: wusushu@hhu.edu.cn
2. Engineering Services, Department of Public Works, City of Temple, 3210 E. Ave. H, Bldg. A, Temple, TX 76501-8402; email: dbond@ci.temple.tx.us

Abstract

Flume experiments were performed to investigate the needed thickness of concrete mattress under wave action. Hydrodynamic pressures above and below the mattress were measured, and the mattress's stability was observed. Results indicate that the hydrodynamic wave pressure on the mattress accords well with cnoidal wave theory. Based on the theoretical analysis, an expression of the required thickness of a concrete mattress under wave action is proposed, which is mainly determined by wave height *H,* water depth *h* and wave period *T*. The theoretical coefficient of the semi-empirical formula was also predigested, according to the theoretical analysis.

Key words**:** hydrodynamic wave pressure, concrete mattress, stable thickness

1 Introduction

Concrete mattress is a new construction method for scour protection. Compared with jackstone, it adapts better to sea bed deformation, and is conveniently prefabricated and constructed. In the US, concrete mattress has been used to prevent bank erosion and to maintain navigation channels in the Mississippi River (Kearney, et al., 1976). In China, it has been successfully applied in both river and estuary engineering, such as bank protection in the Yangtze River (Xu, et al., 2004) and Yellow River (Xu, et al., 2000). Since 1998, it has been used for riverbed stabilization in waterway regulation of the Yangtze River estuary. Wave action is an important hydrodynamic aspect of the flow in the estuary of the Yangtze River. Therefore the stability of concrete mattress under wave action should be researched. Thus far, there have been several studies on the stability of concrete mattress under wave action. Klein (1996), Pilarczyk (1998), Wang (2000) and Gu (2004) reported that the stable thickness is proportional to wave height. It is also related to a parameter which might be a function of breaker parameter (Klein 1996 and Pilarczyk 1998) or the mattress type (Wang 2000 and Gu 2004). But there is a problem employing this parameter to determine the thickness in the Yangtze estuary deepwater channel. The thickness designed by wave methods is much larger than the thickness, 0.16m or 0.20m, applied in the Yangtze estuary deepwater channel. So it is necessary to research the reasons for this and provide a feasible formula for designing concrete mattress thickness in bottom protection engineering.

In this study, a series of experiments were conducted to measure the hydrodynamic pressures on both the upside and underside of the mattress and observe the stability. Then a semi-empirical formula for concrete mattress thickness is deduced based on flume experiment and theoretical analysis. The formula is based on a mattress situated in front of dike, under wave action. Finally the theoretical coefficient of the semi-empirical formula was predigested.

2 Wave flume experiment

2.1 Experimental apparatus and conditions

A series of experiments were conducted at Hohai University, China, in a wave flume having a length of 80m, a width of 0.8m and a depth of 1.0m. A piston-type wave paddle at one end of the wave flume generates waves which dissipate on an absorbing beach at the other end. Three mattress models were built in front of the dike. Model thicknesses were 0.16m, 0.20m, and 0.24m. Each model was 0.4m long and 0.4m wide. Two pressure gauges were mounted on the model to measure the hydrodynamic wave pressures above and below the mattress (D=0.20m) and to determine the hydrodynamic wave pressure difference. Figure 1 shows the position of the gauges. The water depth varied between 2.5m and 6.0m at the location of the mattress. Only regular waves were tested. Wave heights varied between 1.50m and 3.60m, with a constant period of 8.10s, corresponding to a wave steepness varying between 18 and 30. The experimental layout was designed on a scale of 1:20.

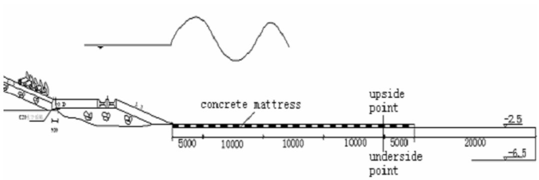


Figure 1. Model of concrete mattress and location of pressure measurements

2.2 Results

Experimental observation shows that all mattresses were stable under wave action. The only failure at wave trough occurred in the 0.16m thick mattress under a wave heights greater than 3.3m. Therefore, the stable thickness of the concrete mattress is between 0.16m and 0.20m. For all other conditions, the stable thickness is less than 0.16m.
According to our observations, the mattress may become unstable under wave troughs. So the hydrodynamic wave pressures above and below the mattress were measured under this condition (Fig.2 and Fig.3). When the relative wave height (*H/h*) is 0.6 and wave period is constant, results show that the hydrodynamic wave pressures always decrease with increasing wave flatness (*L/H*).

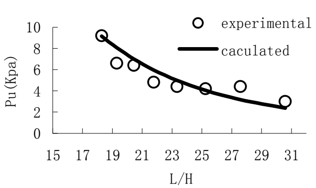


Figure 2. Comparison between calculated and measured hydrodynamic wave pressures above mattress

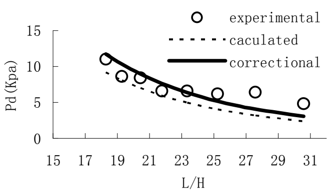


Figure 3. Comparison between calculated and measured hydrodynamic wave pressures below mattress

3 Hydrodynamic wave pressure computation

3.1 Wave theory selection

Since hydrodynamic conditions are very complex and wave behavior is highly nonlinear in beach protection engineering, wave theory used to compute the hydrodynamic pressure should be carefully selected. Wiegel (1960) and Le Mehaute (1976) suggested that wave theory can be chosen according to wave height (H), water depth (h) and wave period (T), or wave length (L). When wave period is constant, relative wave height (H/h) is the critical parameter. In beach protection engineering, concrete mattress usually becomes less stable with increasing relative wave height, when the relative water depth (h/L) is less than 1/10. In this case, cnoidal wave theory and wave flow function are more appropriate. Thus, cnoidal wave theory was selected as the theoretical basis for the following discussion.

3.2 Computation of hydrodynamic wave pressure

Hydrodynamic wave pressure can be expressed as

$$P_u = \gamma_w H\left\{\left[\frac{K(k)-E(k)}{k^2K(k)} + cn^2\left[2K(k)\left(\frac{x}{L}-\frac{t}{T}\right),k\right]-1\right]\right\} \tag{1}$$

where H is wave height, h is water depth, L is wavelength, T is wave period, γ_w is the density of water, $K(k)$ is the complete elliptic integral of the first kind, $E(k)$ is the complete elliptic integral of the second kind, and , $cn^2\left[2K(k)\left(\frac{x}{L}-\frac{t}{T}\right),k\right]$ is the Jacobian elliptic function. The modulus k can be determined by the dispersion relation about the cnoidal wave, which is written as

$$\frac{16}{3}[kK(k)]^2 = \left(\frac{L}{h}\right)^2\frac{H}{h} \tag{2}$$

Figure 2 shows that hydrodynamic wave pressure above the mattress calculated by the cnoidal wave theory fits the experiment points very well with a correlation coefficient of 0.949. Thus the suitability of the cnoidal wave theory is demonstrated.

But the calculated hydrodynamic wave pressure below the mattress does not fit the experimental values satisfactorily, and the correlation coefficient is only 0.809. This may be because the effect on the hydrodynamic wave pressure caused by the concrete mattress is ignored. Since the concrete mattress affects hydrodynamic wave pressure below the mattress, the velocity near the mattress decreases. This increases total wave pressure, which consists of both the hydrostatic and hydrodynamic wave pressures. With hydrostatic wave pressure fixed, variation of total wave pressure causes a hydrodynamic wave pressure change. The direction of hydrostatic wave pressure is opposite that of the total wave pressure under wave trough. It follows that the hydrodynamic wave pressure will decrease with an increase in total pressure. In order to take the interaction with the concrete mattress into account, the coefficient λ is introduced into the primary expression as follows:

$$P_d = \lambda P_u \tag{3}$$

The experimentally determined value of the coefficient λ is 0.78. The modified result is coincident with the experiment value (Fig.3), and the correlation coefficient is 0.950, indicating that the modified method is feasible.

4 Semi-empirical formula for stable thickness of concrete mattress

The main forces acting on the concrete mattress include the hydrodynamic wave pressure difference, underwater gravitational pull on the mattress, and the pull of the cord on concrete mattress.

1） Underwater gravity

$$W=(\gamma_\alpha-\gamma_w)D \tag{4}$$

where D is the thickness of the mattress and γ_a is the density of the mattress.

2） Difference of hydrodynamic wave pressure

The force causing the concrete mattress to float is the hydrodynamic wave pressure difference. According to theoretical analysis and experimental results, it can be computed by the following formula:

$$\Delta P = P_u - P_d = \gamma_w H\,(1-\lambda)\left\{\left[\frac{K(k)-E(k)}{k^2K(k)}+cn^2\left[2K(k)\left(\frac{x}{L}-\frac{t}{T}\right),k\right]-1\right]\right\} \tag{5}$$

where P_u is the hydrodynamic wave pressure above the mattress and P_d is the hydrodynamic wave pressure below the mattress.

3） Pull of the cord on concrete slab

Concrete mattresses are hinged by cords. If an individual slab begins to float, the cord will respond and prevent the slab from floating. For convenience, the pulling force of the cord is neglected in the analysis.

As a result, the stability condition of concrete mattress under wave action can be described as

$$W \geq \Delta P \tag{6}$$

Based on the analysis, a Semi-empirical formula for the stable thickness of concrete mattress under wave action is as follows:

$$D=\frac{\Delta P}{\gamma_\alpha-\gamma_w}=\frac{\alpha H}{\gamma_r'} \tag{7}$$

Where $\alpha=(1-\lambda)\beta$ is the modified coefficient of wave height, β is the theoretical coefficient $\beta=\frac{K(k)-E(k)}{k^2K(k)}+cn^2\left[2K(k)\left(\frac{x}{L}-\frac{t}{T}\right),k\right]-1$, and γ_r' is the relative density of concrete mattress $\gamma_r'=(\gamma_\alpha-\gamma_w)/\gamma_w$.

Equation 7 indicates that the required thickness of concrete mattress is related to wave height H, the modified coefficient of wave height α, and the relative density of the concrete under water γ_r'. The needed thickness increases with increasing wave height H and modified coefficient α, but decreases with increasing relative density of the concrete under water. The modified coefficient of wave height α and the modulus k are functions of water depth h, wave height H, and wave period T. Generally, α and k increase with increasing H and T, and decrease with increasing h.

5 Analysis and predigestion of the theoretical coefficient

Considering the form of the Semi-empirical formula, the theoretical coefficient is the most complex component of the formula. So to be easily applied, the theoretical coefficient will be analyzed and predigested in the following section.

5.1 Analysis of the theoretical coefficient

β is the hydrodynamic wave pressure coefficient based on cnoidal wave theory. It reflects the nonlinearity of waves in the coastal zone. Because the concrete mattress is lifted more easily at wave trough, the theoretical coefficient should be analyzed under this condition. According to the Jacobian elliptic function, $cn^2\left[2K(k)\left(\frac{x}{L}-\frac{t}{T}\right),\ k\right]$ is zero at wave trough. Wiegel, R. L. (1960) and Le Mehaute

(1976) suggested that cnoidal wave theory should be used when the Ursell number is larger than 26. Based on Eq.2, the range of the modulus k is between 0.88 and 1.00. The relationship between the modulus k and the theoretical coefficient β is shown in Figure 4. As the modulus increases, the theoretical coefficient first increases slowly, then decreases sharply. Therefore a maximum value of 0.298 can be found when the modulus is 0.931.

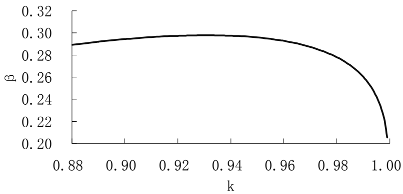


Figure 4. Relationship between modulus k and theoretical coefficient β

5.2 Predigestion of the theoretical coefficient

According to the analysis presented above, the range of the modulus k can be divided two parts. The first range is from 0.88 to 0.931, in which the theoretical coefficient is monotonally increasing. But when the modulus ranges from 0.931 to 1.0, the theoretical coefficient is monotonally decreasing. So the theoretical coefficient should be predigested in two parts accordingly.

(1) Modulus k between 0.931 and 1.0.

Based on the code of hydrology for sea harbors in china, wave length can be computed as follows:

$$L = \frac{gT^2}{2\pi} th \frac{2\pi h}{L} \tag{8}$$

According to Eq.2 and Eq.8, the dispersion relation of cnoidal wave theory can be changed according to the following:

$$\frac{16}{3}[kK(k)]^2 = \left[2\pi \Big/ arcth\left(2\pi L \big/ gT^2\right)\right]^2 \frac{H}{L} \tag{9}$$

When the modulus k determined from Eq.9 is greater than 0.931 and wave height is constant, the modulus decreases with increasing wave length, which may decrease the value of the modulus to 0.931. Choosing conservatively for safety, the modulus should be set as 0.931, corresponding to a theoretical coefficient of 0.298.

(2) Modulus k between 0.88 to 0.931.

When the modulus varies between 0.88 to 0.931, the theoretical coefficient varies from 0.289 to 0.298. The greatest change in the theoretical coefficient is only 0.009 (3%). Again, choosing conservatively, the difference is ignored. Therefore a modulus k of 0.931 can be used to design the concrete mattress.

In summary, the theoretical coefficient is equal to 0.298 when the modulus k ranges between 0.88 and 1.0. So the semi-empirical formula can be predigested as follows:

$$D = \frac{\alpha H}{\gamma_r'} = 0.298 \frac{(1-\lambda)}{\gamma_r'} H \tag{10}$$

The results computed by Eq.10 and Eq.7 are given in Table 1, showing that the results calculated by the predigested coefficient are similar to the original. The greatest difference is only 0.01m.

Table 1. Results for the original and predigested formulas (m)

H	Semi-empirical formula	Predigested formula
3.96	0.19	0.19
3.35	0.17	0.17
2.19	0.10	0.11
1.13	0.04	0.06

6 Conclusion

（1） In the flume experiment, the hydrodynamic wave pressures above and below the mattress were measured. According to theoretical analysis, the cnoidal wave theory was chosen to calculate the hydrodynamic wave pressure. Considering the covering effect of the concrete mattress, the coefficient λ was introduced to calculate the hydrodynamic wave pressure below the concrete mattress. The calculated results fit the experimental points very well, which demonstrates the feasibility of this modified method.

（2） Based on the stressed state analysis of the concrete slab, the required thickness of the mattress under wave action was obtained. Theoretical analysis indicates that the thickness of the concrete mattress is related to wave height H, the modified coefficient of wave height α, and the relative density of the concrete under water γ_r', with wave height being the key factor.

（3） According to the analysis of the theoretical coefficient, the maximal theoretical coefficient is 0.298, if the modulus k is between 0.88 and 1.0. Based on the character of the theoretical coefficient, the semi-empirical formula is predigested. The results computed by the predigested formula are similar to the original.

Reference

Kearney, F. and Prendergast, J. (1976). "STUDY OF ARTICULATED CONCRETE REVETMENT MATTRESS: TEST AND ANALYSIS EM DASH RESULTS OF FY 1975 PROGRAM". Technical Report United States Army Corps of Engineers.

Klein Breteler, M. (1996). "Alternatieve open taludbekledingen (Alternative open slope revetments, draft report, in Dutch)". Delft Hydraulics, H1930.

Le Mehaute (1976). "An Introduction to Hydrodynamics and Water Waves". Springer-Verlag.

Pilarczyk, K. (1998). "Dikes and Revetments Design Maintenance and Safety Assessment". A.A. Balkema, 413-416.

Wang, J. C. (2000). "Technical Specification of Geotextiles Application". China Planning Press, 48-50.

Gu, Y. (2004). "Design and study on the application of Geotextiles in the Yangtze Estuary Deepwater". Master Thesis, Hohai University, China.

Wiegel, R. L. (1960). "A presentation of cnoidal wave theory for practical application". Fluid Mech., 7(2), 273-286.

Xu, Y. X. and Li, X. N. (2000). "Geotextile Application of Water Conservancy". Yellow River Press.175-182.

Xu, X. R., Tang, H.W., Zong, J. and Ye, S. (2004). "New technique for bank protection on Nanjing reaches of Yangtze River". Advances in Science and Technology of Water Resources. 24(4), 26-28.

On Financing for Port Projects by Asset-Backed Securitization in China

Shao Ruiqing[1]

(Shanghai Lixin University of Commerce, Shanghai 201620, China)

Abstract: Based on the fact that the speed of growth of the port capacity lags behind that of economic growth in China because of the shortage of investment in port facilities, the author deems that financing for port projects by asset-backed securitization (ABS) is a practical way to promote the investment. With the operation features of ABS taken into consideration and an analysis of the properties of port project in China made, a tentative design has been presented of the feasible modes of the financing for port projects by ABS in this paper.

Keywords: Port projects, Asset-backed securitization, Financing modes

1. Introduction

China is a booming economic power, with 18,000 islands, over 18,000 kilometers of coastline and more than 100,000 kilometers of inland waterway channel. However, the shortage of investment in port facilities leads to the fact that the speed of growth of the port capacity lags behind that of economic growth. ABS has been an international development trend in financing for projects. Port assets are characterized by their huge capacity to absorb capital and stable economic benefits while the number of listed companies are still insufficient, which put a limit to the expansion of ports through financing in the stock market. It remains an essential issue to seek an innovative way in the development of ports under the background of the high speed economic liftoff of China. As is shown by the experience of other countries, ABS can serve as a supplementary mode of financing for China.

2. Structure modes of ABS

ABS is meant the process of constructing and converting the uncirculated stock assets of the originator, i.e., the seller, into marketable and circulated financial products. In terms of its operations, there should be a special purpose vehivle (SPV) participant in securitization so as to realize the purposes of ABS. The basic structure of ABS is a financial activity participated by three parties. Its simplified model is shown in Fig. 1.

---

[1] Shao Ruiqing(1957—), Ph.D, Professor and Vice President of Shanghai Lixin University of Commerce. Research fields: Investment, Financing and Accounting in Transportation Industry,. Tel: 13501992359; E-mail: rqshao@shmtu.edu.cn

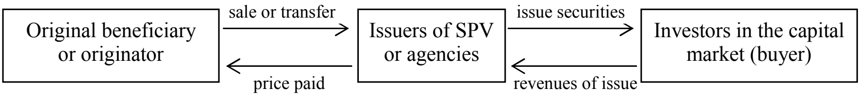


Fig. 1 Basic structure of ABS

The securitized assets have the market as their basis. Thus the capacity to make profits should be the premise of port facilities that have all or part of the properties of public goods. The gains of port facilities normally come from the charges collected from the users, with their content and forms varying from country to country. But most of such gains are stable and can be predicted, which serves as the premise for the securitization of port assets.

3. ABS financing and its participants of port projects in China

The financing body of a project is normally its original beneficiary or originator. The infrastructural and public port facilities is mainly invested by the central or local government in China. In view of the fact that the owner of assets is usually the financing body of port facilities, government should act as the financing body in the securitization of infrastructural and public port facilities while private individuals can serve as that in the securitization of commercial assets. The securitization financing of port facilities, whether by the government or private individuals, is different from the issuing of government or company bonds. ABS requires the establishment of an asset pool which serves as the mortgage of financing. The establishment of an asset pool for the securitization of ports depends on the financing target such as size and maturity. The port assets and expected return of the first phase of the project can serve as the asset pool of the financing of the second phase. In a similar way, financing can thus be conducted from the society.

In China, the authorized agencies of securitization financing consist of investment banks, securities agencies, and trust and investment companies. The investors of ABS for port projects may come from various sources. There are three major objects of financing such as companies related to ports, transportation or logistics, financial institutions, and individuals.

4. Assumptions on three ABS modes for port projects in China

Based on the analysis above three ABS modes which are suitable for port projects in China are designed as follows.

4.1 ABS mode for newly-built independent single port project

The precondition of this mode is that the construction of a new port is entirely the development of virgin soil, with no existent infrastructure or auxiliary facilities available in the construction zone. This mode can be applied to projects under central government as well as projects under local governments.

With regard to the financing of a new port project, the infrastructure input in the prophase of port financing should be the task of governments, either central or local. If the government is short on funds, ABS can fulfill the task. The tentative mode designed can be applicable by departments or organizations concerned in governments such as Ministry of Communications, Ministry of Finance or Department of Land Administration. In that case, with the usufruct of state-owned land as the target assets, and income from anticipated future auction as earnings, the securitization is aimed to discount once for all the transfer income of land usufruct auctioned by stages so that the investor can gain excessive returns from the accretion of the land. A simplified model is showed as Fig. 2.

Detailed procedures of the implementation of the model are as follows:

(a) Departments or wholly-funded state-owned enterprises set up by governments shall carry out feasibility demonstration to the project, mark out the basic layout of the port, separate would-be government-invested land under construction for public facilities with non-governmental land for construction, and then consign assessment institutions to evaluate land price, base price for auction and auction income of non-governmental land for construction. Meanwhile, publicity should be given to construction of port project in order to maximize land price in auction.

(b) Non-governmental land for construction will be divided into lots by category according to the planning, and SPV will be entrusted to conduct packaging and popularization, and predict reasonable total current auction price.

(c) The scale of securities to be issued will be determined by predicted auction price and the securities' rate of return will be determined by the rate of return of the financial market. Meanwhile, derivatives of this basic product, such as option, can be issued to gain option income.

(d) SPV will be entrusted to auction land by stages and by lots according to the construction plan and earn auction income by stages.

(e) After making issuance payment to SPV, governments gain issuance income then.

(f) Issuance income will be invested to infrastructure construction of the port by stages.

(a) With the development of port construction pushing forward and the land price rising, the excess income over the anticipated auction income can repay principle and interest of the issued securities.

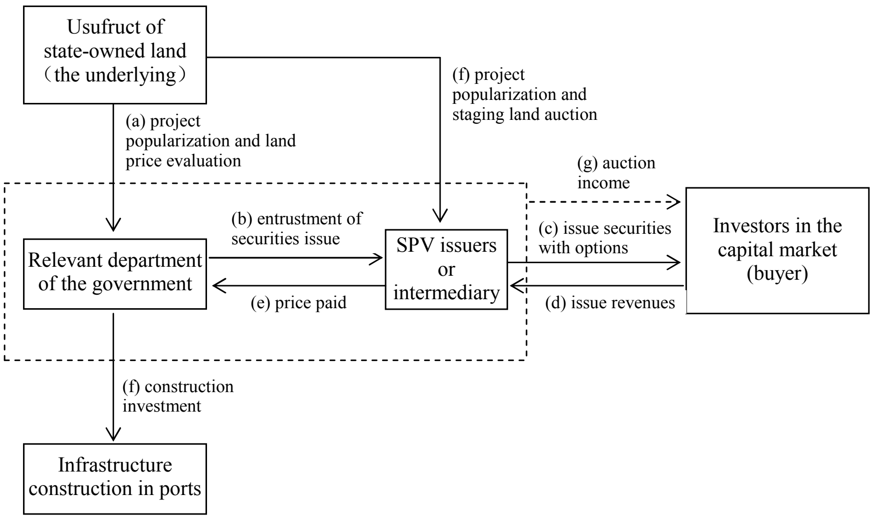


Fig. 2 Financing mod for a newly-built independent single port project

Securitization financing with usufruct of state-owned land as underlying is a brand-new and well-grounded assumption, because port land is scarce resource, and land usufruct has its own internal value. Land or its usufruct value climbs year after year and it enjoys much higher appreciation rate than market rate. This earnings-time limit framework of land assets caters to current investment anticipation of Chinese investors. Besides, this financing mode can achieve win-win results in that governments can gain the future income of land auction once for all and invest it to port infrastructure construction; investors can gain risk free return and excess return of land accretion by way of buying securities and additional option.

4.2 Financing mode of constructing the second-phase project based on securitization of the first-phase project assets

The first mode can further financing but specific implementation shall be determined by the development of the first-phase financing. The following factors should be taken into account:

(a) The rate of land appreciation, which will influence and determine the repayment of capital and interests of the first-phase issued securities;

(b) Construction and planning of supplemental land, which will also influence the cash flow of future land auction;

(c) If there would be anticipation mistakes, the first-phase port assets should be transferred on the market in order to repay the capital the issued securities. After the assets being listed on the market, the holding rate of state-owned share will influence the cash flow of port assets rent.

When the first-phrase project is completed, and the capital and interests of the first-phrase securities are paid out, the project assets and its anticipated income held by the state can be taken as the basic assets for securitization financing of second-phase project. A detailed description of this model is shown in Fig. 3.

As for the implementation of the second-phase assets securitization, the issue scale should be in accordance with the final status of the first infrastructure assets and anticipation of assets income. The difference is that cash flow of repayment of the principle and interest doesn't come from land auction, but from port assets finally hold by the state after first-phase construction and anticipated assets income. Therefore, the implementation of the second-phase securitization is based on the success of the first-phase securitization. If everything is all right, the mode can continue till a mature port is built.

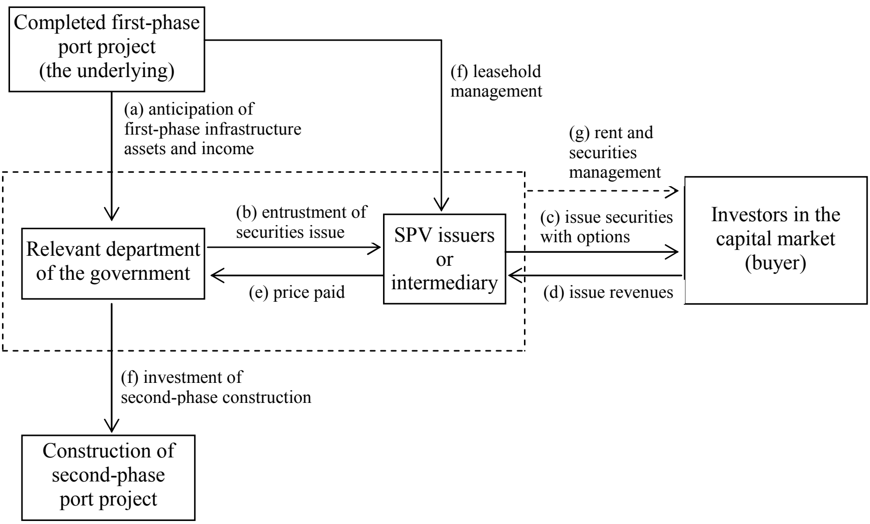


Fig. 3 Financing mode of constructing the second-phase project based on the first-phase project

4.3 Financing mode of constructing new port project based on comprehensive securitization financing planed by the country

The precondition of this mode is that the sponsor has the ownership or usufruct of several port assets, and the anticipated social benefit of newly-built port will be great. Therefore, this mode is applicable to national port construction project planed by central government, such as port construction project of shipping center. A detailed model of assets securitization is showed as Fig. 4.

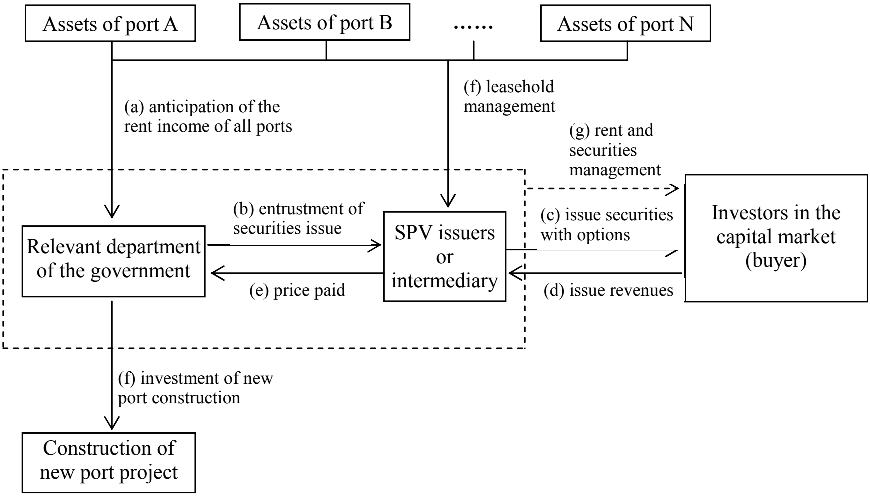


Fig.4 Financing mode of the construction of a new port project through national comprehensive asset securitization

Suppose that the central government has completed infrastructure construction on N ports by way of the former two financing modes or other modes, and gained certain usufruct of each port assets, further return brought by the right can be taken as basic assets for new port construction financing.

Securities issued by this financing mode have great capability to resist market risks. Competitions among ports play an endogenous role within the whole national port management system, so that there won't be market risks caused by unstable operation income of single port assets, which is a great advantage of this comprehensive financing mode.

5.Conclusions

Port assets, a type of high quality assets in China, are in a complete non-current condition at present. Port construction demands greater investment. Port projects can be invested in the form of stock assets. The limited amount of capital is far from satisfying the demands of the development of national economy and foreign trade on port development. ABS provides a new financing solution to the bottleneck of the development of national economy and foreign trade, and is therefore a preferable mode of financing in accordance with the national conditions in China.

Bibliography

1．P· A ·Samuelson. *The Pure Theory of Public Expenditure*. Review of Economics and Statistics. 1954

2．Gordon, H.S. The Economic Theory of a Common-Property Resource: The Fishery, Journal of Political Economy . 1954

3．Olson, Mancur, 1965. *The Logic of Collective Actions*. Trans. by Chen Yu.. Shanghai Hongkong Joint Publishing Company, 1995

4．Wang, Kaiguo, *On Assets Securitization*. Shanghai University of Finance and Economics Press, 1999

5．Wu Wei, *Analysis of the Efficient Provision of Public Goods from the Perspective of Economics*, doctoral dissertation, Northwestern University, 2004. 4

6．Liu, Junlin, Li, Yanbao. *Exploration of Capital Structure Theory on Current Port Financing Mode in China*. Port Economy

7．Huang, Yong, Ma, Ning, *Research on Multifinancing Model of Port Construction*. Journal of Beijing Jiaotong University. 2006. 9

8．Jin, Song, Feng, Haifeng, Zhang, Meiqing. *Actuality and Developing Trend of Investment and Financing Direction of the Waterage in China*. World Shipping. 2006. 2

9．*Shao Rui qing , The Reference on the Financing of the Waterage Projects in Europe, US and Japan. China Investment. 2002. 12*

10．Yu, Jing. *Exploration of Financing Modes of Port Infrastructure*. Port Economy. 2004. 3

RESEARCH ON CAPACITY AT URBAN INTERSECTIONS WITHOUT TRAFFIC SIGNALS

Haiyuan LI and Wei DENG

School of Transportation, Southeast University, No.2, Sipailou, Nanjing, P. R. China, 210096; PH: 86-25-83795857; Fax: 86-25-83794102; email: timber2001@163.com

Abstract
Two-Way Stop-Controlled (TWSC) intersections are used widely. Capacity of vehicular traffic streams at TWSC intersections has been studied. This paper provides an improved procedure or model for determining TWSC intersection capacity based on traffic conflicts, for intersections with shared lanes and flared entries. The developed model was calibrated for conditions in China and can also assess the influence of non-motorized road users. The model results suggest that impact of non-motorized users on capacity of vehicular traffic can be appreciable.

Key words: Two-Way Stop-Controlled (TWSC) intersection, Capacity, Traffic conflict

1 Introduction

Two-Way Stop-Controlled (TWSC) establishes a type of intersection operation with widespread use both in urban and rural environments. The priorities given to pedestrian, bicycle and vehicle movements at TWSC intersection in traffic laws of different countries vary. Lots of efforts were devoted to analyze the capacities at TWSC intersections. A research project sponsored by Ministry of Science and Technology, P. R. China, was carried out by the authors. As a result, a modified procedure was developed to estimate the capacities of traffic streams at TWSC intersections based on the field data and conflict technique method.

2 Conflict Technique Method

2.1 Capacities of streams in a departure sequence

Conflict technique method does not claim to be exact mathematics. It simplifies the operation of intersection by a plausible analogy to queuing systems. A conflict group (a departure sequence) is formed in the same conflict area. Each conflict group involves many conflict points, which are close to each other and can be occupied by only one vehicle at any time. One conflict group usually contains traffic streams from several directions (see Figure 1). It is assumed that every vehicle of stream i occupy the conflict area for exactly t_{Bi} seconds. All streams in a conflict group can use 3600 seconds all together in an hour. If the traffic flow of stream j, Q_j, is known, which has the right of way over stream i, the probability of stream j occupying the conflict area is given by following equation. (Note: Refer to Table 1 for description of symbols used in all equations.)

$$P_{Bj} = \frac{Q_j \cdot t_{Bj}}{3600} \tag{1}$$

The probability that the conflict area is not occupied by a vehicle of movement j is given as follows.

$$\overline{P}_{Bj} = 1 - P_{Bj} \tag{2}$$

For a waiting minor vehicle, the conflict area is also occupied, if a priority vehicle is approaching the conflict area. Assuming that the priority stream gaps are exponential distributed, the probability that the conflict area is not occupied by an approaching vehicle is estimated by following equation (Brilon, 2005).

$$P_{aj} = e^{-\frac{Q_j \cdot t_{Aj}}{3600}} \tag{3}$$

A vehicle of minor stream can only enter the conflict area if both above conditions are fulfilled simultaneously. The probability that both conditions are fulfilled is given as follows.

$$P_{0j} = \left(1 - P_{Bj}\right) \cdot P_{aj} \tag{4}$$

The maximum capacity of movement i (C_{maxi}) is the maximum number of vehicles that can pass through the conflict area without influence of priority streams.

$$C_{\max i} = \frac{3600}{t_{Bi}} \tag{5}$$

Then the actual capacity of movement i can be expressed as follows (Brilon, 2005).

$$C_i = C_{\max i} \cdot \prod_{j \in D_k} P_{0j} \tag{6}$$

2.2 Capacities of streams in more than one departure sequences

A vehicle of minor streams at TWSC intersections has to decelerate or stop. Only when all conflict areas required to pass are not occupied by other priority streams, minor vehicle can then enter the intersection. Figure 2 is the case that minor stream 3 has to pass through two conflict areas, A and B. According to Brilon and Miltner (2005), the capacity of the stream i can be determined as follows.

$$C_i = C_{\max i} \cdot \prod_{k \in D_{ni}} P_{0ki} \tag{7}$$

3 Capacity Models of Streams at TWSC Intersections

Assuming that each turning movement (12 vehicle movements, 8 pedestrian movements and 8 bicycle movements) has its own traffic lane at approaches, more than one conflict area has to be examined for each movement. All these conflict areas can be arranged into 12 groups of conflicts, according to the graph theory and the types of conflicts (see Figure 3). The vehicle of stream i can enter the intersection only when all relevant conflict groups are free of other priority streams. In such case the capacity of vehicle stream i will be as follows.

$$C_i = C_{\max i} \cdot \prod_{k \in D_{ni}} P_{0ki} = \frac{3600}{t_{Bi}} \cdot \left\{ \prod_{k \in D_{ni}} \left[1 - \frac{1}{3600} \cdot \sum_{j \in D_k} \left(A_{ji} \cdot Q_j \cdot t_{Bj} \right) \right] \right\} \cdot e^{-\frac{1}{3600} \cdot \sum_{j=1}^{12} \left(A_{ji} \cdot Q_j \cdot t_{Aj} \right)} \tag{8}$$

When pedestrian movements, bicycle movements and vehicle movements are conflicted to each other, it is defined that $A_{ij} = 1$ for movement i having priority over movement j, or $A_{ij} = 0$ for movement i having to give way to movement j. Since the conflicts among pedestrians and the conflicts among bicyclists are minor, these two types of conflicts are not taken into account. In addition, conflict coefficient can be modified by adopting the priority degree proportion for traffic streams at observed intersections.

4 Group Calculation of Non-motorized Movements

Since pedestrians and bicyclists usually get through the intersections group by group, so group volume and group occupation time of pedestrian and bicycle movements should be adopted in the calculation models.

4.1 Group calculation of pedestrian movements

In order to determine group volume and group occupation time, the analysts must observe in the field or estimate the platoon size of pedestrians waiting to cross the intersections (TRB, 2000).

$$N_{ci} = \frac{V_{pi} \cdot e^{V_{pi} \cdot t_{ci}} + V_{pvi} \cdot e^{-V_{pvi} \cdot t_{ci}}}{\left(V_{pi} + V_{pvi}\right) \cdot e^{\left(V_{pi} - V_{pvi}\right) \cdot t_{ci}}} \tag{9}$$

With $t_{ci} = \dfrac{W_{Bi}}{S_{Pi}} + 1.1$ (10)

$$V_{pvi} = \sum_{j \in D_k} \left(A_{ji} \cdot Q_j \right) \tag{11}$$

Then the spatial distribution of pedestrians can be obtained as follows (TRB, 2000). If no platooning is observed, spatial distribution of pedestrians is assumed to be 1.

$$N_{pi} = INT\left[\frac{0.75\cdot(N_{ci}-1)}{W_{Ei}}\right]+1 \tag{12}$$

Group occupation time of pedestrian movement i can be determined as follows.

$$t_{Gpi} = t_{ci} + 2\cdot(N_{pi}-1) \tag{13}$$

Group flow rate of pedestrian movement i can be expressed as follows.

$$n_{pi} = INT\left(\frac{V_{pi}}{N_{pi}}\right) \tag{14}$$

4.2 Group calculation of bicycle movements

As the analytic method of pedestrian group, group volume and group occupation time of bicycle movements can be obtained. The platoon size of bicyclists who wait to cross the intersection can be obtained through observation or estimation methods.

$$N_{si} = \frac{V_{bi}\cdot e^{V_{bi}\cdot t_{bi}} + V_{bvi}\cdot e^{-V_{bvi}\cdot t_{bi}}}{(V_{bi}+V_{bvi})\cdot e^{(V_{bi}-V_{bvi})\cdot t_{bi}}} \tag{15}$$

With $t_{bi} = \frac{W_{Bbi}}{S_{bi}} + 2.5$ (16)

$$V_{bvi} = \sum_{j\in D_k, k\in D_{ni}} (A_{ji}\cdot Q_j) \tag{17}$$

Then the spatial distribution of bicyclists can be obtained as follows.

$$N_{bi} = INT\left(\frac{N_{si}-1}{W_{bi}}\right)+1 \tag{18}$$

$$W_{bi} = \bar{D}_b \cdot N_{bfi} \tag{19}$$

$$N_{bfi} = \frac{W_{Zi}}{D_0} \tag{20}$$

Group occupation time of bicycle movement can be determined as follows.

$$t_{Gbi} = t_{bi} + 2\cdot(N_{bi}-1) \tag{21}$$

Group flow rate of bicycle movement i can be expressed as follows.

$$n_{bi} = INT\left(\frac{V_{bi}}{N_{bi}}\right) \tag{22}$$

5 Capacity Calculations of shared traffic lane and flared approach

5.1 Capacity calculation of shared traffic lane

According to Wu (2000) the capacity of the shared traffic lane will be as follows.

$$C_{sh} = \frac{\sum_i Q_{si}}{\sum_i X_{si}} \tag{23}$$

With $\sum_i Q_{si}\cdot t_{Bsi} + (Q_F\cdot t_{BF}) + (Q_b\cdot t_{Bb}) \le 3600$ (24)

5.2 Capacity calculation of flared approach

To make sure that the flared lane can offer the function of the right-turn lane, the maximum capability of the flared lane can be determined as follows (TRB, 2000).

$$n_{\max} = \max_i (L_{si}) + 1 \tag{25}$$

With $L_{si} = \frac{d_{si}\cdot V_{si}}{3600}$ (26)

Then the actual capacity of the approach with flared traffic lane, according to TRB (2000), will be as follows.

$$C_{sf} = \left(C_{sz} - C_{sw}\right) \cdot \frac{n}{n_{\max}} + C_{sw}\text{, if } n \le n_{\max} \tag{27}$$

$$C_{sf} = C_{sz}\text{, if } n > n_{\max} \tag{28}$$

6 Calibrations of Parameters in the Models

According to the method proposed by Brilon and Miltner (2005), parameters were calibrated based on the data from the eight intersections observed. Moreover, based on the approach of Kyte et al. (1996), the observed capacities (C) can be estimated as follows. The values of the occupancy time parameters (t_{Bi}, t_{Ai}) used are given in Table 2.

$$C = \frac{3600}{t_1 + t_{mv}} \tag{29}$$

7 Evaluations of the Capacity Models

The capacities observed at the intersections using Kyte et al. (1996) method compared well with the model results (see Figures 4 and 5). In addition, Figure 6 shows that the model results compared well with capacities from the conventional gap acceptance method as proposed by Gao (1999). Figure 7 indicates that the non-motorized road users significantly influence the capacities of the vehicle movements.

8 Conclusions

The paper presents a procedure and model, based on traffic conflicts, for determining capacities of traffic streams at TWSC intersections. The model results match well with observed capacities and capacities obtained from Gap Acceptance Method. The model results demonstrate that impact of pedestrian and bicycle movements to vehicle movement capacities can be significant.

Acknowledgements

This research has been funded by a national project from Ministry of Science and Technology (Project no. 2006BAJ18B03) and a grant for excellent doctor dissertation from Southeast University, P. R. China.

References

Brilon, W. and Miltner. T. (2005). "Capacity at intersections without traffic signals." Transportation Research Record 1920, Journal of Transportation Research Board, National Research Council. Washington, DC, pp. 32-40.

Gao, H. (1999). The analysis methods of capacity at highway intersections. Ph.D. Dissertation, Southeast University, Nanjing, P.R. China.

Kyte, M. et al. (1996). Capacity and Level of Service at Unsignalized Intersections, Final Report Volume 2 - All-Way-Stop-Controlled Intersections, NCHRP Web Doc 6, TRB, National Research Council, Washington, DC.

Miltner, T. (2003). "Level of service at unsignalized urban intersections." Institute for Transportation and Traffic Engineering. Ruhr-University Bochum. No.27.

TRB (2000). Highway Capacity Manual. Transportation Research Board, National Research Council, Washington, DC.

Wu, N. (2000a). "Determination of Capacity at All-Way Stop-Controlled (AWSC) Intersections." Transp. Res. Rec 1710, Journal of Transportation Research Board, National Research Council. Washington, DC, pp. 205-214.

Wu, N. (2000b). "Capacity at All-Way Stop-Controlled and First-In-First-Out Intersections." In: Proceedings of the 4th International Symposium on Highway Capacity. Transportation Research Circular E-C018, TRB, Washington, DC.

Table 1. Symbols and Descriptions

Symbols and Descriptions	Eq #
t_{Aj} is average time during which an approaching vehicle of priority movement j is occupying the conflict area in advance of its arrival.	1
D_k is the set of priority streams in conflict group k related to stream i.	6
P_{0ki} is the probability that conflict area k is not occupied by priority streams for stream i.	7
k is the number of conflict areas related to stream i.	7
D_{ni} is the set of conflict areas that stream i needs to pass through.	7
D_k is the set of conflict streams in conflict group k; A_{ji} is the conflict coefficient.	8
N_{ci} is size of a typical pedestrian crossing platoon for pedestrian movement i.	9
V_{pi} is flow rate of pedestrian movement i.	9
V_{pvi} is flow rate of vehicle movements conflicting with pedestrian movement i.	9,11
t_{ci} is average time for a single pedestrian crossing conflict area.	9,10
W_{Bi} is total width of one-way lanes and bicycle path that pedestrian movement i has to cross.	10
S_{Pi} is average pedestrian walking speed; W_{Ei} is effective crosswalk width.	10,12
N_{si} is size of a typical bicyclist crossing platoon for bicycle movement i.	15
V_{bi} is flow rate of bicycle movement i.	15
V_{bvi} is flow rate of vehicle movements conflicting with bicycle movement i.	15,17
t_{bi} is average time for a single bicyclist crossing conflict area.	15,16
W_{Bbi} is total width of one-way lanes and bicycle path that bicycle movement i has to cross.	16
S_{bi} is average speed of bicyclists.	16
W_{bi} is actual width of bicycle occupying road when bicyclists ride through intersections.	18,19
$\overline{D}_b$ is actual width of a bicycle occupying road when a bicyclist rides through intersections.	19,20
N_{bfi} is platoon size of bicyclists at the first row.	19,20
W_{Zi} is effective width of bicycle path; D_0 is immobile width of a bicycle occupying road.	20
Q_{si} is traffic volume of vehicle stream i at shared lane.	23,24
X_{si} is the degree of saturation of vehicle stream i, that is $X_{si} = Q_{si}/C_{si}$; C_{si} is the capacity of vehicle stream i.	23
t_{Bsi} is the average time that vehicles of vehicle stream i occupy the conflict area.	24
Q_F and Q_b are the traffic volumes of pedestrian and bicycle streams.	24
t_{BF} and t_{Bb} are the average time of pedestrians and bicyclists occupying the conflict area.	24
L_{si} is average queue length of vehicle stream i, under the condition that right-turning stream uses the right-turn lane, while through and left-turning streams share a traffic lane or not.	25,26
d_{si} is the average delay of vehicle stream i; V_{si} is the flow rate of vehicle stream i.	26
C_{sz} is the capacity of the approach with the right-turn lane.	27,28
C_{sw} is the capacity of the approach without the flared traffic lane.	27
n is the actual capability of the flared traffic lane; C is the estimated capacity.	27,29
t_1 is the required time for a vehicle in the first position of the queue to pass the intersection.	29
t_{mv} is the time for a vehicle to move up from the second to the first position of the queue.	29

TABLE 2. Proposal values of the model parameters

Traffic Movements i	1,7	2,8	3,9	4,10	5,11	6,12
t_{Bi} (s)	3.3	2.2	2.4	3.6	3.5	3.2
t_{Ai} (s)	3.2	3.5	3.0	4.0	4.3	3.8

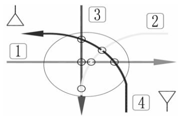


Figure 1. Four streams in a conflict group

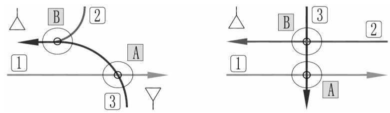


Figure 2. A stream in several conflict groups

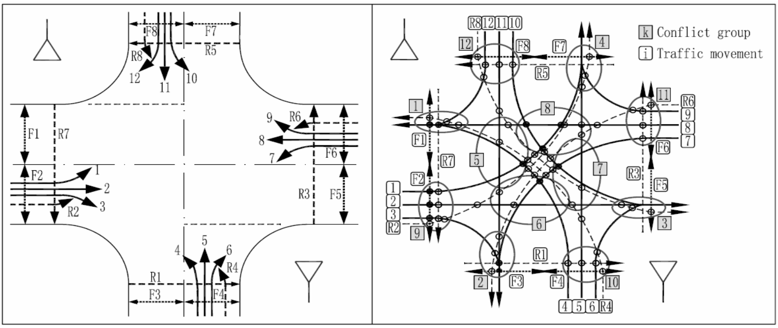


Figure 3. Arrangement of conflict groups at an TWSC intersection including pedestrians and bicyclists

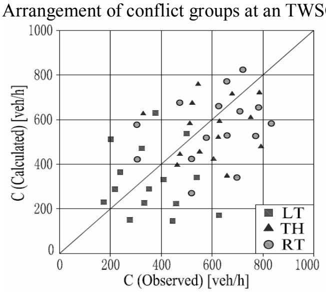


Figure 4. Comparison of observed capacities and capacities calculated by conflict technique with absolute priority.

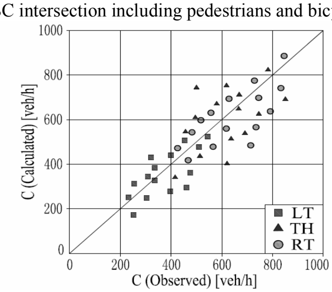


Figure 5. Comparison of observed capacities and capacities calculated by conflict technique with observed priority.

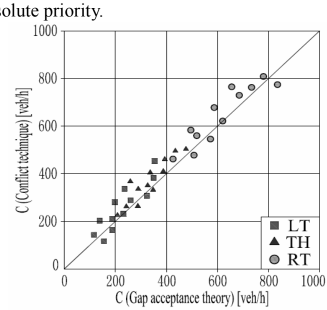


Figure 6. Comparison of capacities from different methods

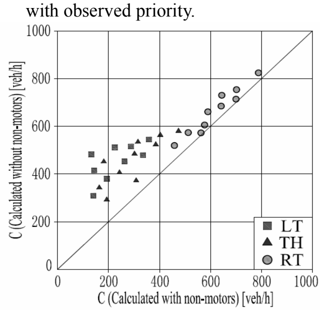


Figure 7. Comparison of capacities with and without non-motorized users.

A Model of Parking Choice and Behavior

Yanjie Ji[1], Wei Wang[2], Wei Deng[3], and
Jean-Daniel Saphores[4]

[1] School of Transportation, Southeast University, No. 2, Sipailou, Nanjing, P. R. China; 210096; Tel: 13813996939; Fax: 025-83794102; Email: jiyanjie@seu.edu.cn
[2] Idem; Tel: 025 -83794101; Fax: 025-83794102; Email: wangwei@seu.edu.cn
[3] Idem; Tel: 025 -83794103; Fax: 025-83794102; Email: dengwei@seu.edu.cn
[4] Institute of Transportation Studies, University of California, Irvine, USA, 92697; Tel: (949) 856 4454; Fax: (949) 824-8385 Email: saphores@uci.edu

Abstract

With growing auto traffic and scarce land in most Chinese cities, parking guidance systems (PGS) are becoming increasingly valuable for reducing congestion and making efficient use of existing parking facilities. This paper contributes to the growing parking choice literature by proposing a two-phase parking model for pre-trip PGS with an application based on a parking behavior survey conducted in Nanjing, China. Our two-phase parking choice model simulates parking choice behavior in real-world situations; it searches all available parking options based on maximum acceptable walking distance to the driver's destinations, and selects optimal routes based on five criteria including distance, safety, convenience, cost, and accessibility. Our modeling approach can help drivers select optimal parking spaces with the benefit of detailed pre-trip route information.

Introduction

Pre-trip parking choice can be simplified to two kinds of choices: choosing a parking space and selecting a travel route. When neither pre-trip nor en-route parking information are provided, drivers typically search for parking spaces based on their previous parking experience. At a given point in time, this search could have three outcomes. Drivers could be: 1) parked after reaching destination; 2) still looking for parking at destination; or 3) stuck in traffic on the way. Inefficient parking searches can significantly increase traffic congestion, decrease environmental quality, and lower productivity. A solution to more efficient parking is to use pre-trip parking information through internet, mobile phones or other information technologies. A scientific parking choice model should therefore be developed to jointly manage traffic and parking information, calculate optimal choices, and distribute the resulting information to users.

Substantial progress has been made in simulating drivers' parking behavior. A review of parking models (Young, 1991) identified a number of modeling approaches that have been employed to understand and replicate parking choice behavior. The CLAMP model has been used to investigate a wide range of parking policies in European cities (Polak, 1989). Thompson and Richardson (1998) modeled the parking search behavior of motorists; they considered access, waiting, direct and egress costs components, and accounted for uncertainty in car parking through queue sizes and departure rates. They found that long term experience does not necessarily lead to better choices. Dell'Orco (2003) presented a model based on Possibility Theory that simulates parking choice behavior for different parking policies. However, to-date few published papers are concerned with pre-trip decisions of parking guidance systems.

This paper aims at developing a parking choice model for pre-trip parking guidance systems that can assist drivers seeking optimal parking and travel routes in terms of availability, distance, safety, convenience, cost, and accessibility. To simulate the search of available options and then deciding among them, we propose a two-phase parking choice model.

Behavior Characteristic of Parking Choice

A random survey of drivers' parking behavior was conducted using face-to-face interviews at six major car parks in the Nanjing commercial area between 10:00 AM and 18:00 PM on a Saturday in April 2006. About 20% of the drivers who parked in these six car parks were interviewed about their parking choices after leaving their vehicles, for a total of 560 respondents.

Results of this survey indicate that the important factors influencing parking choice are (for most to least important) walking distance from parking lot to destination, parking safety, parking convenience, parking fee, parking accessibility, and others (Figure 1).

Figure 1. Distribution of important factors affecting parking choice

Let us start with walking distance, which is the most important factor for our respondents. About 45% of drivers reported they would accept walking 100 to 200 meters from a parking lot to their destination, while about 95% drivers do not want to walk more than 300 meters (Figure 2).

Parking safety is related to parking lot type and management measures. Our survey indicates that underground parking lots are perceived to be the safest (Figure 3).

Parking convenience is also related to parking lot type. Overall, about 44% of drivers interviewed stated that curbside parking is most convenient for them (Figure 4).

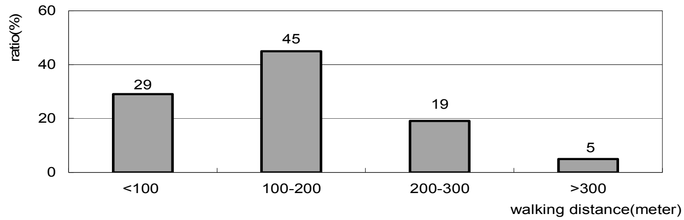


Figure 2. Distribution of acceptable walking distance

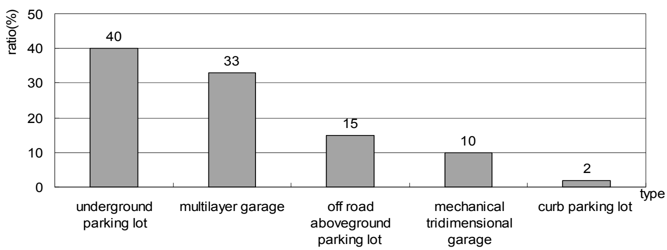


Figure 3. Distribution of perceived safety by parking type

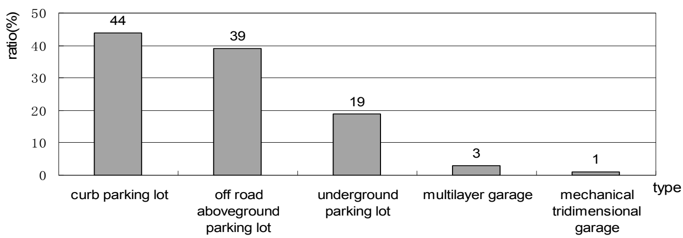


Figure 4. Distribution of perceived convenience by parking type

Two-Phase Parking Choice Model

Basic Assumptions

(1) Walking distance is assumed to be the simple linear distance between a parking lot and the destination of the driver.

(2) Users are assumed to optimize when choosing parking lots and travel routes.

(3) Drivers will not queue for available parking spaces when there are no available spaces in a parking lot.

First Phase

(1) Walking distance: According to the behavioral survey of drivers' parking choice, 95% of drivers are not willing to walk more than 300 meters from a parking lot to their destination. We thus chose parking lots within 300 meters of a driver's destination as his/her initial parking options. Walking distance from the parking lot P_i

(i=1, 2, ···, n) to the destination v_t may be represented by the equation $L_{P_i}=\sqrt{(x_{P_i}-x_{v_t})^2+(y_{P_i}-y_{v_t})^2}$, where (x_{P_i},y_{P_i}) , (x_{v_t},y_{v_t}) respectively represent coordinates of the parking lot P_i and the destination v_t. When L_{Pi} is less than or equal to L_{max} (the maximum acceptable walking distance from the parking lots to the destination, which herein equals 300 meters), parking lot P_i can be included in the parking choice set.

(2) Parking availability: The availability of parking spaces is time dependent so we must take into account the travel time from the trip origin v_s to the candidate parking lots P_i. If the travel time is T_{pi} and t_0 is the departure time, the arrival time is t_0+T_{pi}. To predict the availability of parking spaces, we rely on the method developed by Yang and Chen (2003). Let $B(P_i)$ denote the number of available parking spaces at time (t_0+T_{pi}). If $B(P_i)$ is equal to 0, P_i should be eliminated from the set of candidate parking lots. Otherwise, parking lot P_i can be considered as an option for parking.

Second Phase

(1) Modeling approach: candidate parking lots compose the decision-making alternatives set P = {P_1, P_2, ···, P_m}. Five criteria, including walking distance from parking lots to destination, parking safety, parking convenience, parking cost and parking accessibility, are used as optimization parameters and are included in the object vector set G={L, S, C, F, T}. If the driver only considers one factor among those five indexes (distance, safety, convenience, cost, and accessibility), then the problem simplifies from a multi-objective to a single-objective optimization problem, which is easier to solve. If a driver chooses two or more criteria, it is difficult to ascertain the weight of each sub-objective. In this paper, a simple and practical method for multi-objective decision-making under fuzzy-preference is adopted.

Step 1: Set up decision-making alternatives set P = {P_1, P_2,···, P_m}, objective vector set G = {G_1, G_2,···, G_n} and decision matrix $A_{m\times n}=(y_{ij})_{m\times n}$, where y_{ij} represents the attributes of P_i to a sub-objective G_j.

Step 2: Transform the decision matrix $A_{m\times n}$ to the standard matrix $Z_{m\times n}=(z_{ij})_{m\times n}$.

Step 3: Compute the optimal weight set for multi-objective W= (w_1, w_2, ···, w_n).

Step 4: Compute the synthetic attribute of every candidate parking lot: $D_i(w)$, i=1,…,m.

Step 5: The optimal parking lot can be obtained through the sequence of synthetic attribute of every candidate parking lot.

(2) Quantification of indexes: Walking distance, parking cost and driving time are all quantitative benefit criteria, while parking safety and parking convenience, which are decided by parking lot type, are qualitative criteria. To quantify parking safety and convenience, we used the coordinate system shown in Figure 5 to assign quantitative values to different categories. Assignment results for each parking lot type are shown in Table 1.

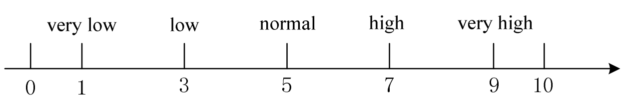


Figure 5. Quantification of fuzzy benefit indexes

Table 1. Valuations of parking safety and convenience

Parking lot type	Curb parking lot	Off road aboveground parking lot	Underground Parking lot	Multilayer garage	Mechanical tridimensional garage
Parking safety	3~5	5~7	7~9	7~9	5~7
Parking convenience	7~9	7~9	5~7	5~7	3~5

(3) Data standardization: the decision matrix $A_{m\times n}= (y_{ij})_{m\times n}$ shows the indices of candidate parking lots; this matrix can be used to develop the basic information of the decision-making problem. The decision matrix can be standardized. The three cost criteria L (distance), F (parking cost), and T (parking accessibility) may be standardized using the function $Z_{ij}=\left(y_j^{\max}-y_{ij}\right)/\left(y_j^{\max}-y_j^{\min}\right)$, where i=1,2,⋯,m. Conversely, the two benefit criteria C (convenience) and S (safety) may be standardized using $Z_{ij}^{'}=\left(y_{ij}-y_j^{\max}\right)/\left(y_j^{\max}-y_j^{\min}\right)$, where again i=1,2,⋯,m. In the above, $y_j^{\max}$=max $\{y_{1j}, y_{2j},\cdots,y_{mj}\}$ and $y_j^{\min}$=min$\{y_{1j}, y_{2j},\cdots,y_{mj}\}$.

(4) Finding the weight of every sub-objective: the weight of every sub-objective is composed of the weight vector W=(w_1, w_2, ⋯, w_n), which satisfies the constraint condition $\sum_{j=1}^{n} w_j^2=1,\ w_j\geq 0$, j=1, 2, ⋯, n. Normally, if an attribute shows little variation across parking lots, it has little influence on parking choice.

With the standardized matrix Z = $(z_{ij})_{m\times n}$, for sub-objective G_j, the difference between candidate parking lot P_i and other candidate parking lots may be defined as: $v_{ij}(w)=\sum_{k=1}^{m}\left|z_{ij}-z_{kj}\right|w_j$, (i=1,2,⋯,m; j=1,2,⋯,n), where, $v_j(w)$ represents the sum of dispersions among the candidate parking lots: $v_j(w)=\sum_{i=1}^{m}v_{ij}(w)=\sum_{i=1}^{m}\sum_{k=1}^{m}\left|z_{ij}-z_{kj}\right|w_j$, (i = 1,2,⋯,m; j=1,2,⋯, n).

The objective function may then be represented by:

$$\max F(w)=\sum_{j=1}^{n}v_j(w)=\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{m}\left|z_{ij}-z_{kj}\right|w_j .$$

The solution is then given by: $w_j^*=\left(\sum_{i=1}^{m}\sum_{k=1}^{m}\left|z_{ij}-z_{kj}\right|\right)\Big/\left(\sum_{j=1}^{n}\sum_{i=1}^{m}\sum_{k=1}^{m}\left|z_{ij}-z_{kj}\right|\right)$.

(5) Decision Ordering: Through the steps mentioned above, the objective weight vector under fuzzy preferences can be obtained; it is denoted by $W^*= (w_1^*,w_2^*,\cdots,w_n^*)$. According to the simple additive weighting method, the function $D_i(w)=\sum_{j=1}^{n}z_{ij}w_j^*$,i=1,2,⋯,m, represents the multi-objective comprehensive evaluation value. The bigger the multi-objective comprehensive evaluation value is, the better is the optimum parking lot.

Evaluation of our Two-Phase Parking Choice Model

The proposed parking choice model considered factors that significantly affect parking choice behavior. Drivers' preferences are taken into account by assigning different weights to different factors. In addition, by modeling parking choice behavior using a two-phase procedure (identifying alternatives and then making a decision), our model not only simulates real-world parking decision-making, but it also shortens the search process and provides drivers with timely advice. In general, the proposed model can reduce search time for parking spaces and it can make a more efficient usage of existing parking facilities.

The model presented above optimizes parking and route choice at the user level rather than at the system level. However, system optimization is essential for improving the transportation system. Further research addressing system optimization is very much needed.

References

Dell'Orco, Ottomanelli and Sassanelli (2003). *Modelling uncertainty in parking choice behavior*. Presented at the 82th Annual Meeting, Transport Research Board, Washington DC.

Polak, J. and Axhausen, K. W. (1989). *CLAMP: a macroscopic simulation model for parking policy analysis*. Presented to the 68th Annual Meeting of the Transport Research Board, Washington DC.

Thompson, R. G. and Richardson A. J. (1998). *A parking search model*. Transportation Research Part A 32(3), 159-170.

Young, W. (1991). *PARKSIM 1.1 users manual*. Department of Civil Engineering, Monash University.

Yang Zhao-Sheng, Chen Xiao-Dong (2003). *Research on the estimation for effective parking space of the intelligentized parking guidance system*. Journal of Transportation Systems Engineering and Information Technology, 12-15 (in Chinese).

CAPACITY OF ALL-WAY STOP-CONTROLLED INTERSECTIONS DERIVED BY CONFLICT TECHNIQUE

Baohong HE[1], Haiyuan LI[2], and Amiy VARMA[3]

[1]College of Transportation, Kunming University of Science and Technology, Kunming, P. R. China, 650093; PH: 86-871-3802298; email: haoziyy@sina.com

[2]School of Transportation, Southeast University, No.2, Sipailou, Nanjing, P. R. China, 210096; PH: 86-25-83795857; email: timber2001@163.com

[3]Department of Civil Engineering, North Dakota State University, Fargo, ND, USA.

Abstract

All-Way Stop-Controlled (AWSC) intersections are used widely. Capacity of vehicular traffic streams at AWSC intersections has been studied. This paper provides an improved procedure or model for determining AWSC intersection capacity based on traffic conflicts, for intersections with shared lanes, short lanes, and flared entries. The developed model was calibrated for conditions in China and can also assess the influence of non-motorized road users. The model results suggest that impact of non-motorized users on capacity of vehicular traffic can be appreciable.

Key words: All-Way Stop-Controlled (AWSC) intersection, Capacity, Traffic Conflict

1 Introduction

All-Way Stop-Controlled (AWSC) intersections have been used worldwide in both urban and rural environments. All departing vehicle streams at AWSC intersections have similar priority. However, the priorities given to pedestrian, bicycle and vehicle movements at AWSC intersection in traffic laws of different countries vary. A modified procedure or model was developed to estimate the capacities of traffic streams at AWSC intersections based on field data and conflict technique method.

2 Conflict Technique Method

2.1 Capacities of streams in a departure sequence

Since all streams at AWSC intersections are considered to be equal in the hierarchy of the priority for departure, the vehicles of different streams have to enter conflict areas alternately. As a result, a conflict group (a departure sequence) is formed in the same conflict area. Each conflict group involves many conflict points, which are close to each other and can be occupied by only one vehicle at any time. One conflict group usually contains traffic streams from several directions (see Figure 1). It is assumed that every vehicle of the stream i occupy the conflict area for exactly t_{Bi} seconds. Under overload or oversaturated conditions, when traffic flows exceed their capacities, the capacities of all streams in a conflict group have the same value of (Wu, 2000a, 2000b).

$$C_i = C = \frac{3600}{\sum_{i=1}^{n} t_{Bi}} \tag{1}$$

If all streams are under non-overload or under-saturated condition, and the traffic flow of the stream j, Q_j, is known, the probability of the stream j occupying the conflict area is given by following equation. (Note: Refer to Table 1 for description of symbols used in all equations.)

$$P_{Bj} = \frac{Q_j \cdot t_{Bj}}{3600} \tag{2}$$

The probability that the conflict area is not occupied by the movement j is given as follows.

$$P_{0j} = 1 - P_{Bj} \tag{3}$$

The maximum capacity of the movement i (C_{maxi}) is the maximum number of vehicles that can pass through the conflict area without influence of other traffic streams.

$$C_{\max i} = \frac{3600}{t_{Bi}} \tag{4}$$

According to Brilon and Miltner (2005), the actual capacity of the movement i under non-overload or unsaturated condition can be expressed as follows.

$$C_i = C_{\max i} \cdot \prod_{j=1, j\neq i}^{n} P_{0j} \tag{5}$$

If some traffic flows other than the subject stream are under the condition of partial overload, then the remaining capacities of partial overload traffic flows can be distributed by other overload traffic flows. According to Wu (2000a, 2000b) the capacity of overload stream i will be as follows.

$$C_i = \frac{3600 - \sum_{j\in D_m} \left(Q_j \cdot t_{Bj}\right)}{\sum_{k=1}^{n} t_{Bk} - \sum_{j\in D_m} t_{Bj}} \tag{6}$$

2.2 Capacities of streams in more than one departure sequences

Figure 2 is the case that stream 3 has to pass through two conflict areas, A and B. If all streams in the departure sequences are non-overloaded, then according to Brilon and Miltner (2005), the capacity of the stream i can be determined as follows.

$$C_i = C_{\max i} \cdot \prod_{k\in D_{ni}} P_{0ki} \tag{7}$$

If all streams in the departure sequences are overloaded, then the capacity of stream i can be given as follows.

$$C_i = \frac{3600}{\sum_{j\in D_{ni}} t_{Bj}} = \frac{3600}{\sum_{k\in D_{ni}, h\in k} t_{Bkh} - (n_i - 1)\cdot t_{Bi}} = \frac{1}{\sum_{k\in D_{ni}} \frac{1}{C_{ik}} - \frac{(n_i-1)\cdot t_{Bi}}{3600}} \tag{8}$$

In fact, if the capacity of the stream i passing through the conflict area k is adopted as the value of C_{ik} under the conditions of partial overload and all overload, then the condition that streams are partially-overloaded has been taken into account. The capacity of the stream i in several departure sequences should be the maximum value in these three cases.

3 Capacity Models of Streams at AWSC Intersections

Assuming that each turning movement (12 vehicle movements, 8 pedestrian movements and 8 bicycle movements) has its own traffic lane at approaches, more than one conflict area has to be examined for each movement. All these conflict areas can be arranged into 12 groups of conflicts, according to the graph theory and the types of conflicts (see Figure 3). The vehicle of the stream i can enter the intersection only when all relevant conflict groups are free of other streams. In such case the capacity of vehicle stream i will be as follows.

$$C_i = C_{\max i} \cdot \prod_{k\in D_{ni}} P_{0ki} = \frac{3600}{t_{Bi}} \cdot \prod_{k\in D_{ni}} \left[1 - \frac{1}{3600} \cdot \sum_{j\in D_k} \left(A_{ji} \cdot Q_j \cdot t_{Bj} \right) \right] \tag{9}$$

Since all vehicle movements at AWSC intersections have the equivalent right of way, all A_{ij}-values equal to 1 among relevant vehicle movements. When pedestrian movements, bicycle movements and vehicle movements are conflicted to each other, it is defined that A_{ij}=1 for the movement i having priority over the movement j, or A_{ij}=0 for the movement i having to give way to the movement j. Since the conflicts among pedestrians and the conflicts among bicyclists are minor, these two types of conflicts are not taken into account. In addition, conflict coefficient can be modified by adopting the priority degree proportion for traffic streams at observed intersections. If a stream has priority over the other streams, the service time of this stream must be subtracted from the total time. The capacity of the stream i can then be determined as follows.

$$C_i = \max\begin{cases} \dfrac{1}{\sum\limits_{k\in D_{ni}} \dfrac{1}{C_{ik}} - \dfrac{(n_i-1)\cdot t_{Bi}}{3600}} \\ \dfrac{3600}{t_{Bi}}\cdot \prod\limits_{k\in D_{ni}}\left[1-\dfrac{1}{3600}\cdot \sum\limits_{j\in D_k}\left(A_{ji}\cdot Q_j\cdot t_{Bj}\right)\right] \end{cases} \quad (10)$$

With

$$C_{ik} = \max\begin{cases} \dfrac{3600-\sum\limits_{g\in D_{pb}}\left(A_{gi}\cdot Q_g\cdot t_{Bg}\right)}{\sum\limits_{j\in D_v} t_{Bj}} \\ \dfrac{3600-\sum\limits_{g\in D_{pb}}\left(A_{gi}\cdot Q_g\cdot t_{Bg}\right)-Q_j\cdot t_{Bj}}{\sum\limits_{h\in D_v, h\neq j} t_{Bh}}, j\in D_v, j\neq i \\ \ldots\ldots \\ \dfrac{3600-\sum\limits_{g\in D_{pb}}\left(A_{gi}\cdot Q_g\cdot t_{Bg}\right)-\sum\limits_{j\in D_v, j\neq i}\left(Q_j\cdot t_{Bj}\right)}{t_{Bi}} \end{cases} \quad (11)$$

4 Capacity Calculation at 422 AWSC Intersections

4.1 Capacity calculation of approaches without flared traffic lane

In the case of intersections with no flared approach for the right-turning stream, all streams (through, right-turning, and left-turning) at each approach share a traffic lane. According to Wu (2000a, 2000b) the capacity of the shared traffic lane will be as follows.

$$C_{sh} = \frac{Q_{sL}+Q_{sT}+Q_{sR}}{X_{sL}+X_{sT}+X_{sR}} \quad (12)$$

With $\left(Q_{sL}\cdot t_{BsL}\right)+\left(Q_{sT}\cdot t_{BsT}\right)+\left(Q_{sR}\cdot t_{BsR}\right)+\left(Q_F\cdot t_{BF}\right)+\left(Q_b\cdot t_{Bb}\right)\leq 3600$

4.2 Capacity calculation of approaches with flared traffic lane

The right-turning stream at this kind of approach can use the flared lane to turn, which increases the capacity of the approach. The actual capacity of the approach with flared lane should be higher than the one without flared lane, and be lower than the one with right-turn lane. To make sure that the flared lane can offer the function of the right-turn lane, the maximum capability of the flared lane can be determined as follows (TRB, 2000).

$$n_{\max} = \max_i\left(L_{si}\right)+1 \quad (13)$$

With $L_{si} = \dfrac{d_{si}\cdot V_{si}}{3600}$ (14)

In the case of intersections having flared approach for right-turning vehicles, the capacity of the shared lane for through and left-turning streams can be assessed as follows.

$$C_{sLT} = \frac{Q_{sL}+Q_{sT}}{X_{sL}+X_{sT}} \quad (15)$$

With $\left(Q_{sL}\cdot t_{BsL}\right)+\left(Q_{sT}\cdot t_{BsT}\right)+\left(Q_F\cdot t_{BF}\right)+\left(Q_b\cdot t_{Bb}\right)\leq 3600$

Then the actual capacity of the approach with flared traffic lane, according to TRB (2000), will be as follows.

$$C_{sf} = \left(C_{sz} - C_{sh}\right) \cdot \frac{n}{n_{\max}} + C_{sh}, \text{ if } \ n \le n_{\max} \tag{16}$$

$$C_{sf} = C_{sz}, \text{ if } \ n > n_{\max} \tag{17}$$

With $C_{sz} = C_{sLT} + C_{sR}$

5 Calibrations of Parameters in the Models

For practical use of the model it is necessary to know the values of the occupancy time parameters for different traffic movements. According to the method proposed by Brilon and Miltner (2005), parameters were calibrated based on the data from the eight intersections observed. Moreover, based on the approach of Kyte et al. (1996), the observed capacities (C_0) can be estimated as follows. The values of the occupancy time parameters (t_{Bi}) used are given in Table 2.

$$C_0 = \frac{3600}{t_1 + t_{mv}} \tag{18}$$

6 Evaluations of the Capacity Models

The capacities observed at the intersections using Kyte et al. (1996) method compared well with the model results (see Figures 4 and 5). In addition, Figure 6 shows that the model results compared well with capacities from the conventional motorcade analysis method as proposed by Gao (1999). Figure 7 indicates that the non-motorized road users significantly influence the capacities of the vehicle movements.

7 Conclusions

The paper presents a procedure and model, based on traffic conflicts, for determining capacities of traffic streams at AWSC intersections. The model has been calibrated for Chinese conditions. The model results match well with observed capacities and capacities obtained from Motorcade Analysis method. The developed model extends capability by incorporating pedestrian and bicycle movements in determination of capacities at AWSC intersections. The model results demonstrate that impact of pedestrian and bicycle movements to vehicle movement capacities can be significant.

References

Brilon, W. and Miltner. T. (2005). "Capacity at intersections without traffic signals." Transportation Research Record 1920, Journal of Transportation Research Board, National Research Council. Washington, DC, pp. 32-40.

Gao, H. (1999). The analysis methods of capacity at highway intersections. Ph.D. Dissertation, Southeast University, Nanjing, P.R. China.

Kyte, M. et al. (1996). Capacity and Level of Service at Unsignalized Intersections, Final Report Volume 2 - All-Way-Stop-Controlled Intersections, NCHRP Web Doc 6, TRB, National Research Council, Washington, DC.

Miltner, T. (2003). "Level of service at unsignalized urban intersections." Institute for Transportation and Traffic Engineering. Ruhr-University Bochum. No.27.

TRB (2000). Highway Capacity Manual. Transportation Research Board, National Research Council, Washington, DC.

Wu, N. (2000a). "Determination of Capacity at All-Way Stop-Controlled (AWSC) Intersections." Transp. Res. Rec 1710, Journal of Transportation Research Board, National Research Council. Washington, DC, pp. 205-214.

Wu, N. (2000b). "Capacity at All-Way Stop-Controlled and First-In-First-Out Intersections." In: Proceedings of the 4th International Symposium on Highway Capacity. Transportation Research Circular E-C018, TRB, Washington, DC.

Table 1. Symbols and Descriptions

Symbols and Descriptions	Eq #
C_i is the capacity of stream i, C is the average capacity of all streams, and n is the number of traffic streams in the conflict group.	1
D_m is the set of the non-overload streams in the conflict group.	6
P_{0ki} is the probability that conflict area k is not occupied by the other streams for stream i	7
k is the number of conflict areas related to stream i	7
D_{ni} is the set of conflict areas that stream i needs to pass through.	7, 10
t_{Bkh} is the time that a vehicle of stream h occupies the conflict area k related to stream i	8
n_i is the number of conflict areas related to stream i	8
C_{ik} is the capacity of stream i passing through the single conflict area k	8
D_K is the set of conflict streams in conflict group k	9
A_{ji} is the conflict coefficient.	9
D_{pb} is the set of pedestrian and bicycle movements in conflict group k related to vehicle stream i;	11
D_v is the set of vehicle movements in conflict group k related to vehicle stream i;	11
A_{gi} is the conflict coefficient between pedestrian or bicycle movement g and vehicle movement i;	11
Q_g is the volume of pedestrian or bicycle movement g;	11
t_{Bg} is occupation time of pedestrian or bicyclist.	11
Q_{sL}, Q_{sT} and Q_{sR} are the traffic volumes of left-turning, through and right-turning streams at the shared lane respectively;	12
t_{BsL}, t_{BsT} and t_{BsR} are the average time that vehicles of left-turning, through and right-turning streams occupy the conflict area, respectively;	12
Q_F and Q_b are the traffic volumes of pedestrian and bicycle streams, respectively	12
t_{BF} and t_{Bb} are the average time of pedestrians and bicyclists occupying the conflict area, respectively	12
X_{sL}, X_{sT} and X_{sR} are the degrees of saturation of left-turning, through and right-turning streams.	12
C_{sL}, C_{sT} and C_{sR} are the capacities of left-turning, through and right-turning streams	12
L_{si} is the average queue length of vehicle stream i, under the condition that right-turning stream uses the right-turn lane, while through and left-turning streams share a traffic lane.	13, 14
d_{si} is the average delay of vehicle stream i.	14
V_{si} is the flow rate of vehicle stream i.	14
n is the actual capability of the flared traffic lane.	17
t_1 is the required time for a vehicle in the first position of the queue to pass the intersection;	18
t_{mv} is the required time for a vehicle to move up from the second to the first position of the queue.	18

Table 2. Occupancy Times (t_{Bi}) for Different Traffic Movements

Traffic Movements i	Left-turning	Through	Right-turning	Pedestrian	Bike
t_{Bi} (s)	3.5	3.3	3.1	5.1	3.5

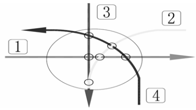


Figure 1. Four streams in a conflict group

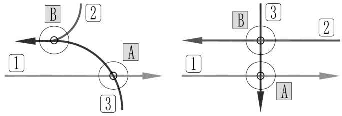


Figure 2. A stream in two conflict groups

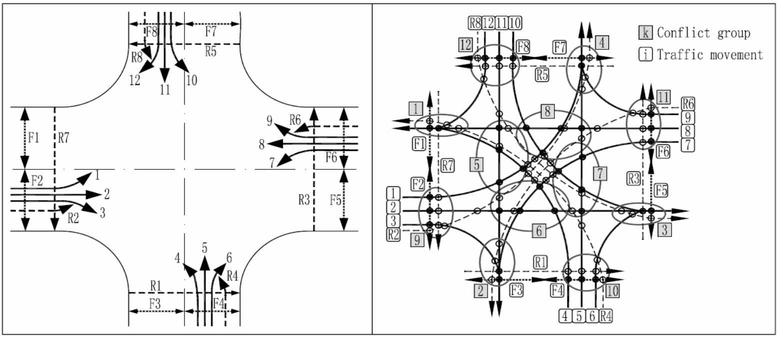


Figure 3. Arrangement of conflict groups at an AWSC intersection including pedestrians and bicyclists

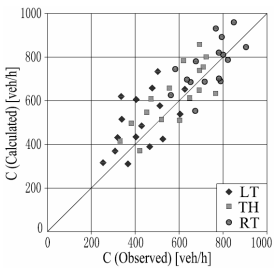


Figure 4. Comparison of observed capacities and capacities calculated by conflict technique with absolute priority.

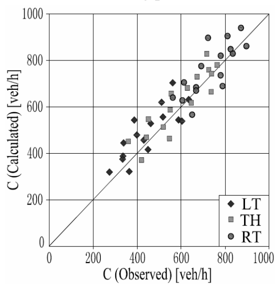


Figure 5. Comparison of observed capacities and capacities calculated by conflict technique with observed priority.

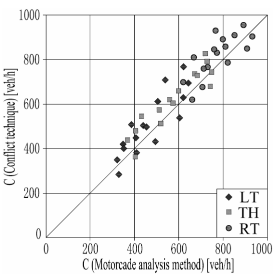


Figure 6. Comparison of Capacities from different methods.

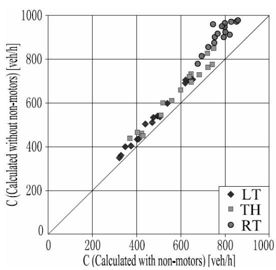


Figure 7. Comparison of Capacities with and without non-motorized users.

Sensitivity Analysis of Large Vehicle Mix Rates on Traffic Flow State Parameters

Lianyu Wei[1], Weizheng Liu[1], Hongjun Cui[1], and Deogratias Eustace[2]
[1]School of Civil Engineering, Hebei University of Technology, Tianjin, China 300401; email: newlwz1981@126.com
[2]Department of Civil Engineering, University of Dayton, Dayton, OH 45469, USA; email: deo.eustace@udayton.edu

Abstract: Mixed traffic is a remarkable character in our country, and therefore, a study on traffic composition is always one of the critical issues in the national traffic engineering field. In this paper, a concept of a large vehicle is used on a basic freeway section, investigated the real-time traffic flow to acquire parameters of vehicle type model. With the help of simulation software, we carried a study on how the 3 traffic flow parameters namely, density, speed and volume, change with the different large vehicle mix rates to find out the sensitivity analysis of large vehicle mix rate to traffic flow state parameters. Finally, we were able to determine the most unsafe large vehicle mix rate in a basic freeway section. The results from this paper are considered as a technical support to speed limits on freeway and traffic management decision system.
Keywords: mixed traffic, large vehicle, mix rate, traffic flow state parameter, and sensitivity analysis.

1 Introduction

Mixed traffic, i.e., different vehicle types, is a remarkable character in our country, and therefore, a study on traffic composition is always one of the critical issues in the national traffic engineering field. Compared to other vehicle types, a large vehicle has quite distinct attributes of shape, dynamic force, and characteristic operations that lead to significant effects on the traffic flow. This results in decreased highway capacity and level of service (LOS), traffic jams, and loss of spatio-temporal resources of roadway networks and high traffic costs. In this paper, a new concept of a large vehicle is defined, firstly, to build up a large vehicle model; and secondly, is used to a basic freeway section to investigate the real-time traffic flow to acquire parameters of vehicle type model. By the help of simulation, we carried out a study using the three traffic flow parameters, namely density, speed and volume to find out the sensitivity analysis of large vehicle mix rates to traffic flow state parameters, and finally determined the most unsafe large vehicle mix rate on a basic freeway section.

2. Analysis of Large vehicle impact on traffic

2.1 Slowing down the velocity of traffic flow

Based on system dynamics and diffusion theory, traffic flow could be considered as an open self-organized system under equilibrium. With different mix rates of large vehicles, the differences between individual vehicles will bring out the inner resistance of the traffic flow system similar to viscous force in hydrodynamics.

This resistance exists between the individual vehicles and it makes them hold down with each other to equilibrium. The increased mix of large vehicles would slow down the traffic flow velocity and reduce the highway capacity. Huber (1986) studied the effect of large vehicles when he developed the passenger car equivalent (PCE) values for trucks.

2.2 Moving bottleneck

Gazis and Herman (1992) note that a cluster of several vehicles would queue behind a slow moving truck. The moving bottleneck definition was put forward by Lighthill and Whitham (1955), which states that there is a function between traffic flow and density, and their model was built up based on the following assumptions:

(1) A slow vehicle runs at the speed v_1 on dual lanes, and the upstream vehicles run at v_0, where $v_0 > v_1$;
(2) The upstream vehicles approach the slow one and follow it in a platoon, and then all upstream vehicles run at the same speed v_1. Both lanes have the same platoon and chance to overtake.

Newell (1998) formulated the moving bottleneck theory by studying the traffic flow near the slow vehicle based on Lighthill and Whitham (1955) thought. Recently Xunchu (2006) studied the road bottleneck, the disorder of traffic at the road bottleneck, platoon, echo waves and traffic jam. In this paper, a model of a moving bottleneck caused by a large vehicle was built up based on the assumption described below.

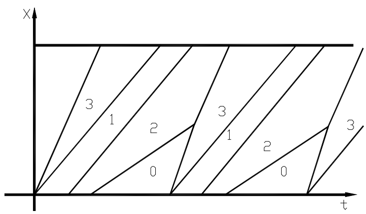


Figure 1 Spatio-temporal impact of moving bottleneck by large vehicle

In Figure 1, a freely flowing traffic runs at the normal state 3. After a large vehicle enters the traffic flow at a speed v_1 the upstream traffic will slow down into state 1, that is, will run at speed v_1. The upstream vehicles will decelerate and follow the large vehicle at state 2 in which the velocity fluctuates to some extent. As the echo wave forms a platoon, the vehicles upstream of the platoon run at the speed of state 0. When the upstream vehicles overtake successfully, they return to state 3. This generates a discharge wave from state 0 to 3. If there is enough space, the platoon wave (2-0) and the discharge wave (0-3) will compress into one traffic wave (2-3), which will run at the speed of the large vehicle, v_1.

Analysis of the average traffic velocity can help to demonstrate the impact of a moving bottleneck. At the bottleneck of a large vehicle, there are only two states, 2 and 3. From the relationship between traffic flow, density and velocity, several definitions can be derived from Figure 1. If A_i is the area of state i, q_i is the traffic flow of state i, and k_i is the density of state i, where i = 1, 2, 3. Then the characteristic variables of the traffic flow of the large vehicle are:

$$q = \frac{A_2 q_2 + A_3 q_3}{A_2 + A_3} \qquad k = \frac{\mathrm{A}_2 k_2 + A_3 k_3}{A_2 + A_3} \qquad v = \frac{A_2 q_2 + A_3 q_3}{A_2 k_2 + A_3 q_3}$$

As a result of the impact of a moving bottleneck by a single large vehicle, the velocity at the bottleneck is up to the free primary traffic flow q and the velocity of the large vehicle, v_l. When the mix rate of large vehicles is low, the impact of a moving bottleneck by large vehicles can be neglected, but when the mix rate is high, the disturbance between the platoon wave and discharge wave will make the impact of a moving bottleneck more complicated.

2.3 Impact on overtake

The overtaking ratio is defined as the ratio of overtaking traffic flow, Q^*, to the total traffic flow, Q, which indicates the overtaking frequency as shown in Eq. 1.

$$\lambda = \frac{Q^*}{Q} \qquad (1)$$

The velocity of a large vehicle is much lower than that of a car; its size is much larger than that of a car. Both of these lead to poor sight conditions when a car follows a large vehicle. When the traffic flow is low, the disturbance between them is negligible. Based on driving behavior theory and experience, the car following behind a large vehicle is likely to overtake it. A successful overtaking event depends on the driver, the acceptance gap, sight distance, the difference in velocity and the power performance of the vehicles.

There are two assumptions to consider: (1) on the one-way dual lane, the overtaking vehicle moves on to the left lane and then return back on to the right lane, and (2) that the vehicles' interference with each other is negligible for low traffic flow, i.e., overtaking is regarded as an independent incident.

If P is defined as the probability of a single-vehicle overtaking, f is the distribution function of large vehicles on the roadway, and F is the cumulative frequency distribution function of overtaking on the lanes, then,

$$F = \int F(P, f) = \sum F(P, f) \qquad (2)$$

For ideal traffic flow, the probability of overtaking on a dual-lanes highway is likely to increase with the mix rate of large vehicles. The cumulative frequency distribution function of the overtaking ratio depends on the distribution of large vehicles on multi lanes of the same direction. The size of a large vehicle reduces the sight conditions of subsequent vehicles, which increases the difficulty of overtaking. Obviously, the overtaking of car-to-car is easier than that of car-to-truck and truck-to-truck. As shown in Figure 2, the car-overtaking track is shorter when overtaking a car than when overtaking a truck.

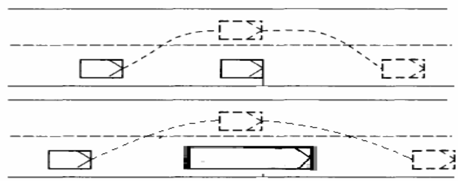


Figure 2 Comparison of different overtaking maneuvers

For the same traffic flow rate, the higher the large vehicles mix rate, the more difficult in overtaking, thus, the lower the overtaking ratio. The moving bottleneck will increase with the mix rate of large vehicles. In other words, it is more probable to overtake at the bottleneck, and the overtaking ratio will increase with the mix rate of large vehicles. Also, since overtaking a large vehicle is more difficult than a car, events of overtaking a large vehicle will decrease with the increased mix of large vehicles.

3. Simulation study

Since traffic accidents occur in flashes both spatially and temporally, research efforts on traffic accidents are mainly based on historical data. The collected data is always limited and partial because the instantaneous state of traffic flow can not be obtained. That is why in this study the simulation software, VISSIM, is used. VISSIM is a microscopic, behavior-based traffic simulation software. In order to simplify the study, we assume a traffic stream composed of cars with different mix rates of large vehicles on a basic section of a one-way highway with dual lanes, 0% gradient, and 1000 meters long. Data collection points are set at 100 m-intervals.

4. Result analysis

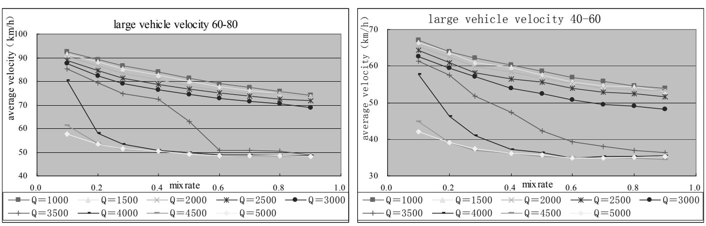


Figure 3 Relationship between the average velocity and mix rate of large vehicles

Figure 3 above shows the impact of the large vehicle mix rate on the average velocity for different flow volumes. When traffic flow is beyond 4000 veh/h, the traffic becomes congested. Both the large vehicles and cars run at slower velocity. This reduces the impact of large vehicles in high density traffic flows. But in the non-congested situation, with the mix rate of large vehicles in the range of 0.2 to 0.4, it is obvious that the average traffic velocity changes with the mix rate of large vehicles.

Figure 4 shows that, in the non-congested traffic flow, the velocity variance of traffic flow changes with the mix rate of large vehicles. It first increases and then decreases. In the congested situation, however, it decreases steadily with the mix rate of large vehicles. That is to say, in the non-congested situation, the mix of large vehicles increases the frequency of velocity change, from free to blocked. However, when the mix rate of large vehicle increases to some extent, about 0.3, there are so many large vehicles that the highway capacity falls significantly and the free cars are blocked by large vehicles into a platoon. It can be said that the higher the difference between large vehicles and cars, the higher the impact of large vehicles mix rate has on discreteness of traffic velocity.

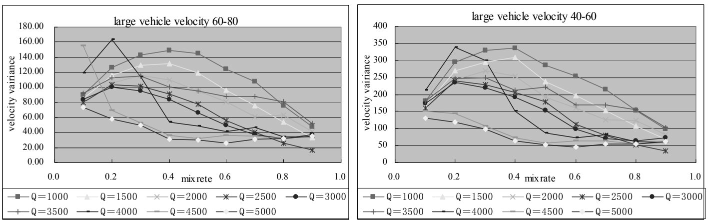


Figure 4 Relationship between the variance of velocity and mix rate of large vehicle

Figure 5 shows that the overtaking ratio first increases with the mix rate of large vehicles, and then decreases. The turning point is at the mix rate of large vehicles of about 0.3. It can be concluded that accident occurrences are most probable at the mix rate of large vehicles of about 0.3. Thus, the lower the traffic flow, the larger the overtaking ratio, i.e., in free flow traffic, it is easier to overtake. But as the large vehicle mix increases, traffic become more congested and do not meet conditions favorable for overtaking, i.e., in terms of velocity, headway, and sight distance.

Figure 6 shows that the large vehicles impact density variance in the same way that they impact velocity variance and overtaking ratio.

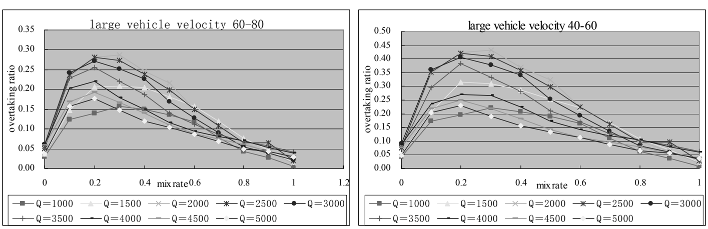


Figure 5 Relationship between the overtaking ratio and mix rate of large vehicle

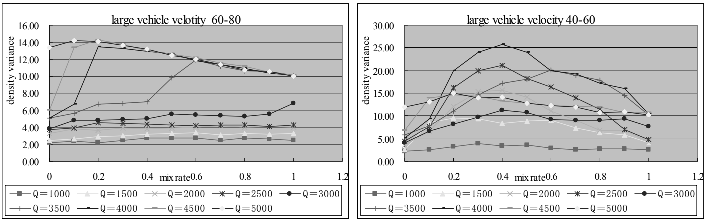


Figure 6 Relationship between density variance and mix rate of large vehicle

5. Conclusions

Based on the simulation results of the study presented in this paper, the following conclusions can be drawn:

The mix rate of large vehicles can reduce the velocity of a traffic flow. The mix of large vehicles makes the velocity to be discrete. The velocity discreteness increases with the increasing mix rate of large vehicles and then decreases. At about 0.3 mix rate, it is most probable to overtake. But, with increasing traffic flow, it becomes difficult to overtake.

The mix rate of large vehicles generates a moving bottleneck as a result of non-uniformity in density distribution. When large vehicles and cars run at quite different speeds, the density discreteness first increases and then decreases with the mix rate of large vehicles.

In a non-congested situation, with a mix rate of large vehicles of about 0.3, the large vehicles have the most impact on traffic flow. In this case the traffic flow is most unsteady and this becomes more obvious when the large vehicles and cars are running at quite different speeds. The results form this study can be used as a support into vehicle-type research in traffic design and for special management to large vehicles in our country.

6. References

Gazis, D.C., Herman, R. (1992) "The moving and phantom's bottleneck". *Journal of Transportation Science*, 26, 223-229.

Huber, M. J. (1986) "Estimation of passenger car equivalents of trucks in traffic stream." *Transportation Research Record,* 869, 60-70.

Lighthill, M. J. and Whitham, G. B. (1955). "On kinematic waves II. A theory of traffic flow on long crowded roads." *Proc. Roy. Soc. of London*, London, UK, Series A, 229, 317-345

Newell, G. F. (1998) "A moving bottleneck." *Trans. Research Part B*, 32 (8), 531-537.

Xunchu, X.. (2006). "Planning thought of urban road network system and examples analysis." *Urban Transport of China*, 4(1), 30-35.

Geotechnical and Geometric Considerations in Highway Design and Construction in Hilly Terrain

Vishnu Diyaljee[1], F.ASCE

[1]Managing Director, GAEA Engineering Ltd, 2408 -113 A St, Edmonton, Alberta, T6J 4Y2; diy001@gaeaeng.com

ABSTRACT

Some of the primary geotechnical and geometric issues considered in upgrading of a 12 km section of Hwy 40:34 situated in hilly terrain are presented and discussed. Steep grades, erosion, substandard curves, sharp shoulders, rock slopes, soft ground, wetlands, and the proximity of natural lakes to the highway were some of the features addressed in ensuring a satisfactory design and end product. This case study illustrates the importance of addressing geotechnical, geometric and construction issues simultaneously at the design stage, especially for highways in hilly terrain.

INTRODUCTION

The 12 km section of Hwy 40:34 in question is situated along Hwy 40 near the Town of Grande Cache, Alberta, located 143 km northwest of Hwy 16 near Hinton (Figure.1). This section extends 6 km east of the Smoky River Bridge and 6 km west of the Town of Grande Cache. Being close to the foothills of the Rocky Mountains, this highway is also characterized by steep (6 to 7%) grades on either side of the Town, requiring climbing lanes to be constructed to accommodate truck traffic. To the east of the Town, the roadway alignment traverses between rock slopes and Victor and Grande Cache Lakes, which are popular tourist and recreation facilities in the Province of Alberta. To the west of the Town, the alignment is bordered by steep gravel and shale slopes rising to about 60 m above the existing roadway (Figs.2 and 3).

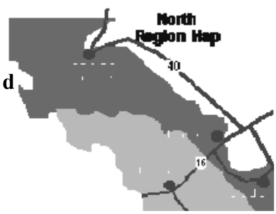


Figure 1. Project Location

Figure 2. Steep Gravel Slope

Figure 3.Steep Shale Slope

GEOMETRIC DESIGN CONSIDERATIONS

An engineering evaluation and assessment was undertaken of various recommendations for this section of highway made in a 1997 Planning Report by others. From this review, it was determined that overlaying the existing pavement would result in a 7.3 m wide pavement which would not meet the 3R/4R (Geometric Design Standards for Existing Paved Roads) suggested minimum criterion of a pavement width of 10 m (Highway Design Guide 1999). A 10.8 m width was desired at the overlay stage to be compatible with the section north of the Smoky River bridge which was under construction.

Measurements north of the Smoky Bridge showed that the base course was 13-14 m wide and that this width could be attained by widening on both sides of the existing highway in the majority of instances. Widening on one side, desirable from a construction aspect, was also evaluated but was not considered economically feasible as a result of steep side slopes and steep and high backslopes (up to 60 to 70 m in height) that encroached the highway in many locations. The steep gravelly hillside and rock slopes required widening of the roadway to be selectively undertaken resulting in widening on one side, both sides, and alternate sides in order to achieve maximum widths without creating serious construction problems.

For the East hill climbing lane, the recommendation made in the 1997 Planning Report was to construct this from Station (Sta.) 29+570 to Sta. 31+530 on the right of centerline. This location was reviewed in the field and minor adjustments were made to suit the existing ground conditions near Victor Lake and Grande Cache Lakes. For the West hill climbing lane, important constraints were (a) the existence of steep embankment sideslope approximately between Sta.33+136 and Sta.32+806 and, (b) past slope stability problems at about km 35.0. On the basis of the field review it was decided to locate the climbing lane between Sta.32+786 and Sta.34+786. Fig. 4 shows the grade widening strategies adopted for the final design.

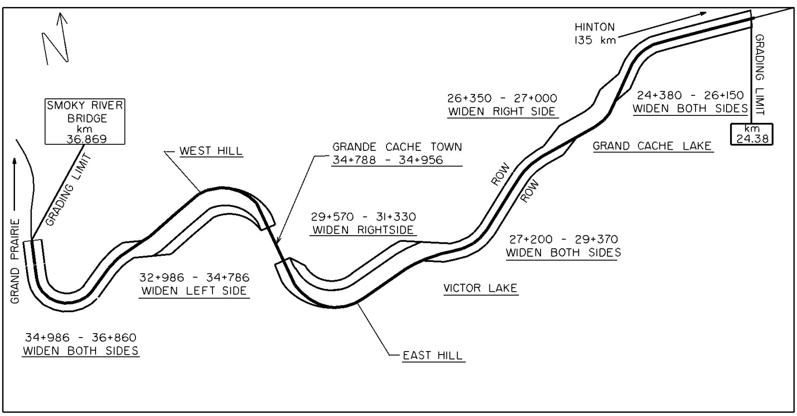


FIG.4. Plan of Route Showing Grade Widening Strategies (km 24.38 same as Sta 24+380 etc)

GEOTECHNICAL DESIGN CONSIDERATIONS

A geotechnical investigation of the site was undertaken with a backhoe as the soils were observed to be predominantly either rock or gravel in nature. Testpits were dug at locations in and within the areas where construction activity was to take place. Backslopes and sideslopes were tested to determine the materials type. Rock areas were logged as these were visible from the exposed cuts alongside of the highway. The rock areas corresponded to those identified in the 1967 construction report of the existing highway. In that report the rock was excavated by blasting. Moisture content and Atterberg Limit tests were undertaken to provide data for contractual purposes. From the soil survey, the materials to be obtained from the cuts along the route were generally found to be satisfactory road building materials.

A geotechnical review of the project was undertaken in relation to the design cross-sections, ground conditions and soils information. Slopes in the rock cut and earth cut areas were modified, where appropriate, to accommodate the characteristics of the materials. Steeper slopes were identified for areas which would otherwise result in larger volumes of excavation than needed.

One particular area between Sta.28+400 and Sta.29+300 on the East Hill contained muskeg and wet ground on both sides of the roadway adjacent to Victor Lake. The testpits showed generally silty material overlying rock with a relatively small muskeg thickness. From a review of past construction records of the existing highway, it was noted that the grade had to be raised in the muskeg area and a pad had to be built before heavy earthmoving equipment could be used to build the grade. Since the widening would be partially within the existing embankment for a major length of the muskeg stretch, the existing conditions were considered to be satisfactory.

For the excavation of rock, blasting was made mandatory in hard rock areas. Special provisions were written to have the rock blasted or ripped into sizes which were not large and were manageable to utilize in the fill areas. The use of very large rock sizes rock in the grade construction was considered undesirable since such material would not be properly compacted without allowing fine material to occupy the void spaces to produce a dense soil mass as required by the Specifications.

The need for large volumes of fill or the construction of retaining walls was identified at two locations along the alignment. The first location was between Sta. 32+980 and Sta.33+080, where a culvert led to a ravine that flowed towards the Smoky River. The second location was between sta. 33+840 and sta.33+900, where an extension of the slope of a 20 m high embankment fill would impact existing power lines and environmentally sensitive wetlands at the toe of slope.

Gabion retaining walls averaging 3 m at the high fill location and 6 m in height at the culvert location were designed for both these locations. The provision of the retaining wall at the culvert site allowed an indirect saving of the cost of the pipe extensions which would have been required otherwise. The design of the retaining wall at the high embankment location obviated fill extending into a wetland area. Further, a need to relocate power poles at the toe of the slope was also avoided, which would have been required otherwise to accommodate the fill extension.

A helical pile anchoring system was designed for the gabion wall at the high fill location to provide additional resistance against sliding. The wall at the culvert location was designed with Tensar geogrid reinforcement to support the new backfill.

There were two critical rock areas in this project on the east side of the Town Grande Cache. The first one was near Grande Cache Lake (approximately from Sta.26+580 to Sta.26+700) and the other one was near Victor Lake (approximately from Sta. 29+310 to Sta. 29+520). From a field reconnaissance, the existing slopes were observed to be stable and inclined at 1V: 1H. Since the existing rock slopes were stable, similar 1:1 cut slopes were adopted in the new design. Special provisions were incorporated in the tender document for conducting blasting operations through these sections in a controlled manner such that the excavation does not go beyond the existing right-of-way limits and the rock is broken into smaller sizes which can be utilized into the new grade. It was also stipulated that precautions should be taken (a) to avoid any harmful effects to people, animals and existing housing structures around the lakes as a result of the operations, and (b) to install protective devices to prevent flying rock, rock fragments or boulders from entering the lakes and from spilling onto the existing asphalt surfaced roadway.

Erosion problems were anticipated where steep grades existed. In addition, a potential for polluting the lakes adjacent to this highway was also considered possible, as a result of the construction activity. Strawbales and silt fences were identified to be used along the ditches and around the culverts to trap the sediments and prevent then from being taken to the water bodies. This was especially pertinent along the East Hill in the vicinity of Grande Cache Lake, Victor Lake and at different culvert inlet and outlet ends on the West Hill that discharge run-off in the Smoky River.

The preliminary design cross sections were reviewed in the field to determine where sections required any changes to be more practically developed during construction in relation to the topography and soils condition. The extensions of these sections in relation to the right-of-way were also examined and modifications made to allow the roadway to fit within the present right-of-way limits at all locations. Where necessary and feasible, ditch widths were reduced in steep cut areas.

RELEVANT CONSTRUCTION ISSUES

The highway upgrading was started in May 1999 and was completed in November 1999, taking a total of 110 effective working days. The Prime Contractor managed various components of work operations through sub-contractors. Rainy weather prevailed for a major part of the construction time resulting in slow progress generally. Work operations were undertaken on one side of the roadway at a time allowing for the traffic accommodation on the existing road.

Construction of the two gabion walls was undertaken by a sub-contractor experienced in such work. Appropriate sizes of the rock excavated within the project limits were used to fill in the gabion baskets. Figures 5 and 6 show the gabion wall in the high fill location, during and after construction. Similarly, Figures 7 and 8 show the Tensar reinforced gabion wall at the culvert location.

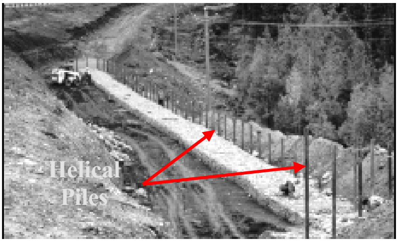


Figure 5. Gabion Wall Construction

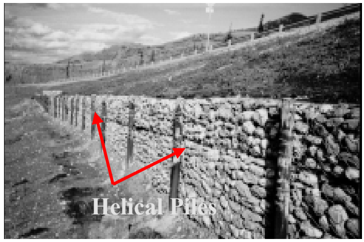


Figure 6. Completed Gabion Wall

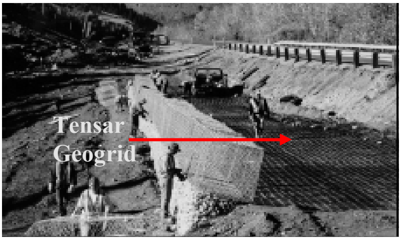


Figure 7. Tensar Gabion Wall

Figure 8. Completed Gabion Wall

Controlled rock blasting was done to remove rock cuts from most of the rock slope area on the east hill project section between sta. 26+578 and sta. 26+710. At other locations, the rock was jack hammered and loosened with a rock breaker which was also used to reduce oversize rock. The Contractor also posted signs informing the travelling public. The traffic was stopped until the blast was done and at least one lane of traffic was cleared of rock that might have covered the existing road.

For the muskeg section, the use of a geotextile for padding across was debated at the time of design but was not considered necessary because of the raise in the grade at the time of original construction. However, the grade widening portion was found to be quite soft at the time of construction because of wet weather conditions. As a result, a woven geotextile was used as reinforcement and a slow rate of construction was followed as a precaution to mitigate foundation failures.

Figure 9. Laying of Geotextile

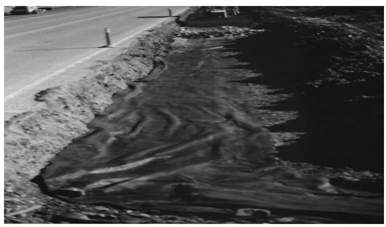


Figure 10. Backfilling on Geotextile

Temporary erosion control measures in the form of silt fences (Figure 9) were implemented during the construction in the ditches and at the culvert invert and outlet ends to prevent soil erosion in the ditches and minimize sediment transportation towards the water bodies. For permanent erosion control measures, the Contractor opted to use a patented synthetic permeable sedimentation barrier called "Enviroberm"(Cascade Geotechnical Inc.)(www.cascade.ab.ca). These barriers were constructed of two panels of high density, extruded UV resistant polyethylene pinned down with "M" pins in a single row, complete with an erosion control matting on the underside. Figure 10 shows the "Enviroberm" in ditch location.

Figure 11. Silt Fence in Ditch

Figure 12 Enviroberm in Ditch

GENERAL DISCUSSION AND SUMMARY

The Contractor used equipment that was wider than the designed width of widening, resulting in extra width of the grade widening and use of base course material in excess of the design requirements. The design and construction of this project provided a learning experience under two scenarios: (a) the need for better planning coordination between ideal design concepts and actual conditions on a project, and (b) the need to consider potential repercussions of circumstances at the time of construction, if different from the design assumptions.

Under the first scenario, it is highly desirable before starting a detailed design to explore what are the practical constraints in a project and design the details of the project to suit those constraints, since a good design depends on the understanding of various construction aspects. Typical examples of issues to be considered are: (a) site specific constraints, (b) availability of resources-personnel, equipment and funding, and constructability. Under the second scenario, although not necessarily likely to happen in every project, it is desirable to consider at the design stage itself various situations that can arise at the time of construction and provide appropriate contingency solutions or precautions in the designs. Typical examples of the issues to be considered are: (a) variation in weather conditions, and (b) mobilization of equipment different from the type originally anticipated at the time of design.

REFERENCES

Alberta Infrastructure. (1999). "Highway Geometric Design Guide" www.infratrans.gov.ab.ca

Optimizing Highway Alignments Based On Improved Particle Swarm Optimization and ArcGIS

Tu Shengwen[1,2], Guo Xiucheng[1],Tu Shengwu[3]
[1]School of Communication, Southeast University, Nanjing 210096, China;
[2]School of Highway Engineering, Changsha University of Science & Technology, Changsha 410076, China;PH(086)15951856796; e-mail: tusw2004@163.com
[3]Jiangsu Transportation Research Institute, Nanjing 210000, China; PH(086)13601403651; e-mail: tsw@ti.js.cn

Abstract: Optimizing highway alignments is a very complex engineering problem. The factors that should be considered in the design process are complex and interrelated. Up to now a variety of methodologies have been applied to the optimization of highway alignments, but no single method had yet been developed that presented itself to be universally accepted as an effective model for the problem. In this paper, A method that integrates geographic information systems(GIS) with improved particles swarm optimization(PSO) for simultaneously optimizing three-dimensional highway alignments is presented. With improved PSO introduced, the complex nonlinear alignments optimization problem becomes very simple to carry out. A GIS system —ArcGIS and its object-oriented data model — Geodatabase are exploited to store relevant geographic information and highway alignment elements. A real life example is presented to demonstrate that the proposed method has good efficiency and better convergence compared with existing methods.

1 Introduction

For more reliable and realistic applications highway alignment optimization processes should consider many factors. Recently, a solution approach (Jong and Schonfeld, 2003) based on genetic algorithms (GAs) for three-dimensional highway alignment optimization has been developed. Another model integrating geographic information systems (GIS) with GAs has also been developed (Jha and Schonfeld, 2000a, 2004).

Many studies show that genetic algorithms for solving complex engineering optimization problems can be effective and efficient. However, they also require more on development of operators appropriate to the problems and their algorithms are fairly complex to be programmed. It is desirable that a model for optimizing highway alignment should directly exploit a GIS database because most spatial information is becoming available in such computer-readable form. But most study ignored the fact that topographic maps are generally stored in CAD format.

In this paper, we present a method that integrates geographic information systems with improved particles swarm optimization(PSO) for simultaneously optimizing three-dimensional highway alignments.

2 Highway costs and design constraints

2.1. Highway cost components

In this study, highway costs are classified into different categories according to their relations to alignment characteristics for easy formulations. The complete cost function involves complex computations. The detailed formulations are given in Jong (1998). Each cost component is briefly discussed below.

2.1.1. Location-dependent costs

Highway costs such as land acquisition and soil stabilization costs that depend on the location of the alignment are classified as location-dependent costs.

2.1.2. Length-dependent costs

Highway costs that vary linearly with the total alignment length are categorized as length dependent costs.

2.1.3. Earthwork cost

The "average-end-area" method (Wright,1996) is adopted in this study for estimating earthwork volume of a highway project.

2.1.4. User costs

The major components of user costs usually include vehicle-operating costs, the value of travel time, and accident costs. Vehicle-operating costs may include fuel and oil consumption, tire wear, and vehicle depreciation.

2.2. Design constraints

The most important constraints in highway designs are maximum gradient, minimum radius and minimum length of vertical curve. Maximum gradient depends on the importance of the road and the topography over which the highway is built. Minimum radius is a function of design speed, superelevation, and coefficient of side friction. Minimum length of vertical curve is controlled by the sight distance and the change of road grades. All constraints are handled through external penalty functions that are added to the total cost function.

3 The optimization problem

Our highway alignment optimization problem is to find an alignment that minimizes overall cost, subject to constraints. In our problem the horizontal alignment is represented as a combination of tangent sections and circular curves. The vertical alignment is represented as a combination of tangent sections and parabolic curves. The coordinates of intermediate points through which the alignments cross are treated as decision variables. For this purpose, first a number of orthogonal cutting lines (for horizontal alignment) or planes (for vertical alignment) are constructed at equally spaced intervals between the given end points (see Fig. 1). Those lines are orthogonal to the line connecting the start and end points. (For 3-D alignments, we consider orthogonal cutting planes.) The intermediate points lie somewhere along these cutting lines (or planes). While the precision of alignment thus generated depends on number of intermediate points. The optimization problem is thus formulated as

$$\min_{xp_1,yp_1,zp_1,xp_2,yp_2,zp_2\ldots xp_n,yp_n,zp_n} \quad C_T=C_O+C_U+C_P \qquad (1)$$

s.t. $x_l \leqslant xp_i \leqslant x_u \quad \forall i=1,2,3\cdots\cdots n;$ (2a)

$y_l \leqslant yp_i \leqslant y_u \quad \forall i=1,2,3\cdots\cdots n;$ (2b)

$z_l \leqslant zp_i \leqslant z_u \quad \forall i=1,2,3\cdots\cdots n;$ (2c)

$|g_i| \leqslant G_{max} \quad \forall i=1,2,3\cdots\cdots n;$ (2d)

$L_{vi} \geqslant l_{min} \quad \forall i=1,2,3\cdots\cdots p;$ (2e)

where C_T is total cost, C_O, C_U, and C_P are operator, user, and penalty costs, (x_l, y_l, z_l) and (x_u, y_u, z_u) are lower and upper limits on the decision variables (xp_i, yp_i, zp_i)'s. n and p are the number of intermediate points and vertical curves, respectively. Eqs. (2a),(2b) and (2c) impose upper and lower limits on the decision variables and Eqs. (2d) and (2e) limit the maximum gradient and minimum vertical curve length.

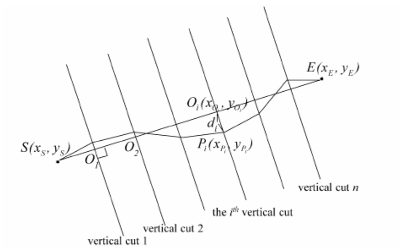


Fig. 1. Alignment representation for a two-dimensional case.

The two algorithms for generating horizontal and vertical alignments are designed to mimic a manual engineering design process (Jong,1998). At any intersection point where the intersection angle is not zero, a circular curve (for horizontal alignment) or a parabolic curve (for vertical alignment) is inserted. Thus, the problem reduces to finding the Pi's, which are treated as the decision variables.

4 GIS application

Recent developments in computing, the growth of the Internet, advances in DBMS technology, object-oriented programming, mobile computing, and wide GIS adoption, have led to an evolving vision and role for GIS. The ArcGIS product was built to satisfy these evolving requirements to deliver a scalable, comprehensive GIS platform. In ArcGIS, Geodatabase—a new object-oriented data model is provided. The geodatabase is a data model for representing geographic information using standard relational database technology.

In this study, we employ ArcGIS as foundation platform of highway alignments optimization and Geodatabase to store basic topographic information. However, as we know, topographic maps are generally stored in CAD format among engineering domain. So how to convert CAD maps into GIS system becomes very crucial. In this study, Visual C++ and the customized tools of ArcGIS—ArcObjects are exploited to realize data transition. Fig.2 shows the procedure of data processing.

5 Solution algorithm

The particle swarm optimization (PSO) algorithm is a member of the wide category of swarm Intelligence methods for solving global optimization problems.

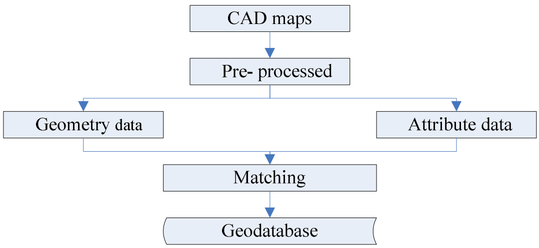


Fig.2. Procedure of data transition

PSO can be easily implemented and it is computationally inexpensive compared with genetic algorithms, since its memory and CPU speed requirements are low. The basic PSO algorithm is developed by Kennedy and Eberhart(1995).

As swarm Intelligence method, PSO also has premature convergence, especially in complex multi-peak search problems. Many researchers are devoted into this field to handle this problem. In this paper, we propose an improved particle swarm optimization (Improved PSO) to perform alignments optimization. In Improved PSO, a population of points sampled randomly from the feasible space. Then the population is partitioned into several sub-swarms, each of which is made to evolve based on PSO. At periodic stages in the evolution, the entire population is shuffled and points are reassigned to sub-swarms to ensure information sharing. The Improved PSO strategy is presented below.

Step 1: Initializing. Select $p \geqslant 1, m \geqslant 1$, where, p is the number of sub-swarms, m is the number of points in each complex. Compute the sample size $s = pm$. Sample s points $X_1;\ldots;X_s$ in the feasible space. Compute the function value f_i at each point X_i.

Step 2: Ranking. Sort the points in order of increasing function value. Store them in an array $E=\{X_i, f_i \mid i=1,2,\ldots,n\}$.

Step 3: Partitioning. Partition E into p sub-swarms $A_1, A_2, \ldots; A_p$, each containing points m, such that: $A^k=\{X_j^k, f_j^k \mid X_j^k = X_{k+p(j-1)}, f_j^k = f_{k+p(j-1)}, j=1,\ldots,m\}, k=1,\ldots,p$.

Step 4: Evolving. Evolve each complex A^k using particle swarm optimization separately.

Step 4.1: Initializing. Select q, T, where, q is the population size of PSO,T is the maximal iterated generation.

Step 4.2:Selecting.Choose q distinct points Y_1^k,…, Y_q^k from A^k according to the function values to construct a sub-swarm. Better points in A^k have more probability to be selected. Store them in $F^k=\{Y_i^k, V_i^k, u_i^k \mid i=1,\ldots,q\}$, where V_i^k is the velocity for particle, Y_i^k and u_i^k is the corresponding function value. Find out the best previously visited position of each particle P_i^k and the position of the best individual of the whole swarm G^k.

Step 4.3: Comparing. Compare the function value between each particle Y_i^k and P_i^k ,if Y_i^k is better than P_i^k ,then $P_i^k = Y_i^k$ .Compare the function value between each particle Y_i^k and G^k, if Y_i^k is better than G^k, then $Y_i^k = G^k$.

Step 4.4: Renewing. According to the formulation of basic PSO , renew the position and velocity of each particle.

Step 4.5: Iterating. Iterate by repeating Step 4.3 and Step 4.4 T times, where T is a user-specified parameter which determines how fast each complex should evolve.

Step 5: Sub-swarms shuffling. Replace A^1, … , A^p into E. Sort E in order of increasing function value.

Step 6: Check convergence. If the convergence criteria are satisfied, stop. Otherwise, return to Step 4.

Improved PSO combines the strengths of the particle swarm optimization, competitive evolution and the concept of complex shuffling. It greatly enhances survivability by a sharing of the information gained independently by each complex. In Improved PSO, each member of a complex is a potential parent with the ability to participate in a process of evolution.

6 A numerical example

In this section, a numerical example in which the optimal alignment backtracks is employed to investigate the effectiveness of the proposed model and evolutionary program. The start and end points of the proposed highway alignment are located at the corner of study area.

The proposed highway is a two-lane rural road and its length is about 2 km. The maximum gradient is assumed to be 5% and the design speed is 40 mph due to the difficult terrain. The maximum gradient restricts the configuration of the resulting vertical profile and affects earthwork cost. The design speed is important in determining minimum radius on horizontal curves and sight distance along vertical curves, thus eventually influencing user costs.

For this problem, 2 intersection points are used to depict the alignments and an initial population of size 150 is randomly generated. The program is run for 100 generations. Since the solution algorithm is probabilistic and the result depends on the random seed, we run 5 program replications. The standard deviation of objective values is about 10% of the mean. This means that the proposed algorithm produces consistent results. The best of these 5 replications yields a total cost 6.35×10^6 Chinese yuan. The graphical representations of the optimized horizontal and vertical alignments are displayed in Figs. 3 and 4.

7 Conclusions

Highway alignment optimization is a complex problem. It has an infinite number of alternatives to be evaluated in a continuous search space. Moreover, cost functions are very difficult to formulate, are non-differentiable, noisy and implicit. Due to the nature of the problem faster and efficient search algorithms are needed rather than

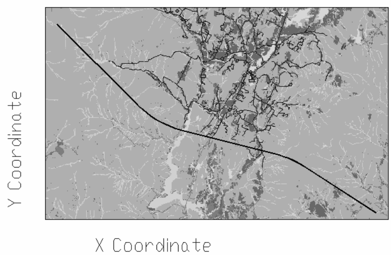


Fig. 3. The optimized horizontal alignment

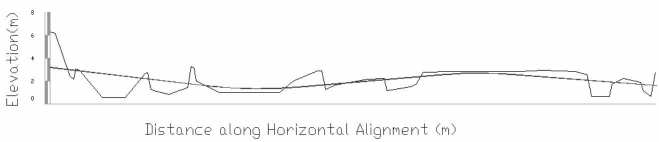


Fig. 4. The optimized vertical alignment

conventional methods.

This study presented a improved PSO approach for highway alignment optimization, one of the swarm intelligence techniques. The proposed approach is implemented in an test example, which indicates that substantial improvement in computing efficiency can be achieved with the improved approach. The approach also improves the solution (i.e., an economical alignment is obtained) compared to the traditional one stage approach.

References

Jong,J.C.,Schonfeld, P., 2003. An evolutionary model for simultaneously optimizing three-dimensional highway alignments. Transportation Research, Part B 37 (2), 107–128.

Jha, M.K., 2000. A geographic information systems-based model for highway design optimization. Ph.D. dissertation, University of Maryland, College Park.

Jha, M.K., Schonfeld, P., 2004. A highway alignment optimization model using geographic information systems. Transportation Research, Part A 38 (6), 455–481.

Eberhart R, Kennedy J. New optimizer using particle swarm theory [C]. In: Proceedings of the sixth international symposium on micro machine and human science. Japan: Nagoya; 1995. p. 39 – 73

Jong, J.-C., 1998. Optimizing Highway Alignments with Genetic Algorithms. University of Maryland, College Park, Ph.D. dissertation.

Wright, P.H., 1996. Highway Engineering. John Wiley & Sons Inc., New York.

Research on Highway Three-Dimensional Dynamic Sight Distance Based on Drivers' Characteristics

LIAO Jun-hong[1, 2], ZHANG Wei-han[2], WU Hong-bo[2], Larry Gillen[3]
[1]Corresponding Author; Engineer; Address: 8 Xitucheng Rd, Beijing, China. 100088; Tel.: +86-10-62357218; Fax: +86-10-62356515;
E-mail: ynwpl@hotmail.com.
[2]Key Laboratory of Road Safety Ministry of Communications PRC, Research Institute of Highway Ministry of Communications, Beijing 100088, China.
[3]Editing Author; Engineer; Address: GILLENgineering, PO Box 122, Marengo, IN 47140, USA; E-mail: lggillen@gillengineering.com.

Abstract: Based on the analyses of research findings from both domestic and abroad, Highway Three-Dimensional Dynamic Sight Distance (H3DSD) and methods for calculating it are put forward in this paper. Key techniques and basic characteristics for H3DSD are illustrated in detail. With the characteristics of second-order continuity, localization, flexibility, etc., cubic B-spline curve, as the prerequisite of calculation of H3DSD, is used to represent highway alignment. Meanwhile, the method for calculating highway geometry elements from GPS data is also provided. Using the relation of space geometry and vector, the calculation of H3DSD is done in 3D space. In the calculation, the drivers' dynamic visual angle, the limitation of the vehicle's headlights at nighttime, and the combination of horizontal and vertical alignment, etc. are taken into account. And a new method is provided to evaluate the consistency of highway alignment much more scientifically.

Key words: 3D Dynamic Sight Distance; Spline curve; Driver's characteristics; Dynamic visual angle; Sight distance.

1 Literature review

Driving behavior is closely related to drivers' visual information, and one of the important measurements for drivers' visual information is sight distance. From the review of the research contents and characteristics from both domestic and abroad, much research has been conducted on the calculation and analysis of highway sight distance, but, many limitations (LIAO 2007) still exist. For instance, some influencing factors are not considered in the calculation. Also, current technical limitations make some suggested approaches difficult or impossible to implement in some regions. Based on existing research findings, improvements on the current methods for calculating sight distance are considered for application.

2 Definition of H3DSD

2.1 Definition

Highway Three-Dimensional Dynamic Sight Distance, as shown in figure 1, is defined as the dynamic and farthest distance from the observation point to the farthest point along the 3D spatial alignment a driver can discern during driving.

H3DSD is a driver's dynamic and actual sight distance and is mainly influenced by operating speed and the combination of horizontal and vertical alignment. With different operating speeds, a driver has different fields of vision, and it is described by transverse visual angle and vertical visual angle in this paper, and specially, when at nighttime, the angle is limited by the headlamp range of the headlight of the operating vehicle. Note that the farthest point a driver can discern is the point determined by searching from the nearest point of the center line or edge line of the lane a driver can discern, until the first point that can't be discerned along the extended direction of the roadway.

In figure 1, the observation point P is the position of a driver's eyes; sight direction is the tangent direction at the point P in 3D spatial alignment; α and β each represents a driver's maximum visual angle of the direction parallel and perpendicular to the roadway surface; Point C is the farthest point of roadway center line or edge line a driver can discern; length (PC) is a driver's H3DSD at observation point P.

2.2 Key Techniques

The key techniques in the calculation and application of H3DSD mainly include:

(1) 3D representation of highway alignment

The core of the calculation of H3DSD is the geometric relationship of spatial points and vectors. Moreover, the precondition of the spatial operation is to compute the 3D coordinate of the spatial point. Thus, using the 3D coordinate of the control-point to construct spline curves representing highway alignment is put forward (BEN-ARIEH 2004). With many prominent characteristics, such as the localization (YANG 2003), the second-order continuity of the joint of two neighbor segment curves (LI 2001), etc., a cubic B-spline curve is chosen as the 3D representation of highway alignment.

(2) Relational model between operating speed and driver's transverse visual angle and vertical visual angle.

According to existing research findings, a driver's dynamic vision is greatly influenced by the operating speed. Generally, a driver's dynamic vision at high speed will reduce 10%-20%, even 30%, compared to the field of vision when stopped. All in all, a driver's field of vision is quite different at different speeds. Therefore, dynamic vision, one of the most important influencing factors of sight distance, must be involved in the calculation of sight distance. The relational model between a driver's visual angle and the farthest distance can be determined by experiments. Then dynamic vision angle can be used as one of the limitations in the sight distance calculation, and the value of sight distance will be much more practicable.

3 Method for Calculating H3DSD

The calculation and analysis steps of H3DSD are:

(1) Computing control-point coordinates

In order to calculate the coordinate, the start point of the highway alignment is considered as the reference point. And the coordinate of this point is set to (0, 0, 0) or assigned the actual coordinate. Using the coordinate of each point of intersection post in the lineal table and curvilinear table, x and y coordinate of each point of roadway centerline can be calculated, and the z coordinate can be calculated utilizing grade and the length of the hill segment. After error treatment, the control-point sequence of the spline curves is determined.

(2) Generating highway alignment

Using the coordinate of the control-point sequence, the cubic B-spline curve can be fitted. And the curve should pass through the start and the end point of the objective highway alignment for better accuracy. With a moving window of four points at a time, x, y and z value of each point can be calculated by feeding the step length t into the multinomial determined by those four points. Then the spatial 3D alignment is developed. Suppose these four points, $P_{j-1}P_jP_{j+1}P_{j+2}$ are the control sequences, the curve between $P_j(x_j, y_j, z_j)$ and $P_{j+1}(x_{j+1}, y_{j+1}, z_{j+1})$ is given by the following equation. In this equation, parameter t is grater than or equal to 0 and less than or equal to 1. The spatial point sequence of the curve between those two points can be generated through changing the value of t from 0 to 1 regularly. Meanwhile, with the characteristics of spline curves, the corresponding tangent vector expression of each point can be easily obtained.

$$\left[x(t), y(t), z(t)\right] = \frac{1}{6}\left[t^3 t^2 t^1 1\right]\begin{pmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{pmatrix}\begin{pmatrix} x_{j-1} & y_{j-1} & z_{j-1} \\ x_j & y_j & z_j \\ x_{j+1} & y_{j+1} & z_{j+1} \\ x_{j+2} & y_{j+2} & z_{j+2} \end{pmatrix}$$

(3) Computing H3DSD

Considering the demand of calculation precision and the limitation of the calculation efficiency, the value of parameter t can be determined. Using the cubic b-spline curve equation, the spatial coordinates at point A, the observation point, and point B, the availability of which needed to be checked, can be calculated. Each checking computation point like point B should be checked in 3D space to ensure whether it's in the driver's vision area or not, from the following aspects: horizontal curve, vertical curve, driver's dynamic vision angle (α 、 β), the vertical angle of vehicle head light beam (β_1), etc. If the current point meets the demand of the above constraints, then assign point B as the next point. Otherwise, the H3DSD value of observation P, length (AB) along the roadway, is retrieved. Figure 5 shows the general computing flow of H3DSD. The core of the computation of H3DSD is to use spatial vector relation in 3D space to judge the point whether it meets the demand of horizontal and vertical dynamic visual angle and the vehicle head light beam, etc. or not. Details of these processes are:

1) Meet the demand of the vertical dynamic visual angle?

In this part, the vertical visual angle can be categorized into two classes, crest curve and sag curve.

(a) Crest curve

Figure 2 shows the situation of a crest curve. ϕ_{PB}— The angle made by line PB with the horizontal plane. ϕ_{BC}— The angle made by line BC with the horizontal plane. ϕ_B— The angle made by the tangent at point B with the horizontal plane. ϕ_A— The angle made by the tangent at point A with the horizontal plane.

When calculating the sight distance, length (AC) between observation point A and current point C along the road should be compared with l_0 (the farthest distance along the road surface the driver can discern with uninterrupted view of that surface without the limitation of roadway environment) first. If length (AC)>l_0, the H3DSD of the observation point is l_0. Otherwise, the following approximation is used:

$$|| \phi_{PB} - \phi_A | - \beta | < \delta$$

For some value of tolerance δ, if it meets this condition, the distance from P to the current point along the road is the sight distance. If it doesn't, continue to the next condition judgment with

$$| \phi_{PB} - \phi_B | < \delta$$

If it doesn't meet this condition, the length of AB along the road is the sight distance. Otherwise, point C is determined when the following condition is met: the angle made by line PB with the horizontal plane should coincide with that made by line BC, namely.

$$| \phi_{PB} - \phi_{BC} | < \delta$$

If that condition is true, then the distance between the current point and the observation point is the sight distance. If not, continue to the next step.

(b) Sag curve

Figure 3 shows the situation of sag curve. ϕ_{PB}— The angle made by line PB with the horizontal plane. ϕ_A— The angle made by the tangent at A with the horizontal plane. β— The vertical angle of the driver's dynamic vision; when at nighttime, assign the maximal vertical angle of the vehicle head light beam, β_1, as β.

When calculating the sight distance at this situation, length (PB) between observation point P and current point B along the road should be compared with l_0 first. If length (PB)>l_0, the H3DSD of the observation point is l_0. Otherwise, the following condition should be checked:

$$| \phi_{PB} - (\phi_A + \beta) | < \delta$$

If this condition is not satisfied, calculate the sight distance. Otherwise, continue to the next step.

2) Meet the demand of transverse dynamic visual angle?

In this model, only the limitation of the driver's transverse visual angle is considered, and the limitation of roadside obstacles or roadside clearance are not. Figure 4 shows the calculation of transverse dynamic visual angle. The check of transverse dynamic visual angle is done in the plane of the road using the relation of spatial vector. In figure 4, α is the maximal transverse angle driver can discern, and χ is the angle made by line AB with the center line of the driver's sight which is the tangent at A. Using the following approximation:

$$|\alpha - \chi| < \delta$$

If α coincides with χ, the distance between A and B along the road is the sight distance. Otherwise, the iterative procedure is continued. Updating the value of t, 3D coordinates of the next point will be calculated using the equation representing the highway alignment. Then, continue the judging procedure.

(4) Drawing the sight distance map and analyzing the sight distance

Using the available H3DSD of each point, the sight distance map can be drawn, and then segments with poor sight distance, or segments with hidden features, can be determined according to *Technical Standard of Highway Engineering*.

4 Conclusions

As one of the measurements of visual information a driver obtains, sight distance is closely related to traffic safety. In order to evaluate the consistency and continuity of highway alignment more impersonally, highway 3D dynamic sight distance, a new manner of describing driver's sight distance, is put forward on the basis of analyzing and studying relative research from both domestic and abroad. After that, the processes and methods for calculating are expounded in detail. H3DSD is a new theory and concept for evaluating the continuity of highway alignment. This method has been applied in highway safety audit. In the future, more research on the calculation of H3DSD will be done by considering roadside obstacles. The innovations of this paper lie in:

(1) Cubic B-Splines are used to represent highway alignment; moreover, it is applied in the calculation of H3DSD. This paper especially emphasizes the importance of calculation in 3D space, and the solution of calculating highway alignment geometry elements by GPS data is also provided.
(2) The drivers' characteristics are emphasized in the calculation. Furthermore, the influence of dynamic visual angle as a function of vehicle speed is taken into account.
(3) The limitation of the vehicle's head light beam at nighttime is taken into account in the calculation.
(4) The detail calculating flow and the method for calculating H3DSD is provided.

References

BEN-ARIEH, D., CHANG. , RYS, M., and ZHANG, G. (2004). "Geometric Modeling of Highways Using Global Positioning System Data and B-Spline Approximation." *J. Transp. Eng.* 130(5), 632-636.

LIAO Jun-hong, ZHANG Wei-han. (2007). "Research on Highway Three-dimensional Dynamic Sight Distance" *J. Journal of Highway and Transportation Research and Development*, 24(7), 99-106.

LI Qing-yang, WANG Neng-chao, YI Da-yi. (2001). "Numerical Analysis" M, Beijing: Tsinghua University Press.

YANG Xiao-jing. (2003). "Methods for Generation of B-spline Surfaces from Triangle Meshes". D. Beijing University of Technology.

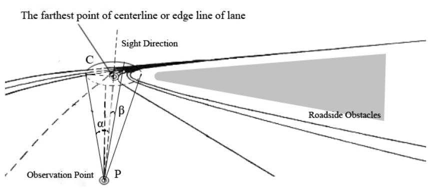


Figure 1 Definition of H3DSD

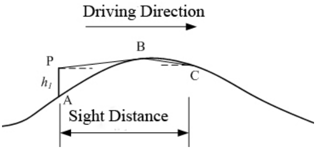


Figure 2 Sketch of crest curve calculation

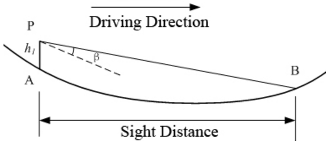


Figure 3 Sketch of sag curve calculation

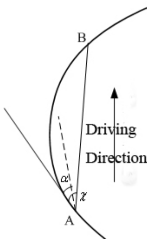


Figure 4 Sketch of calculation of transverse dynamic visual angle

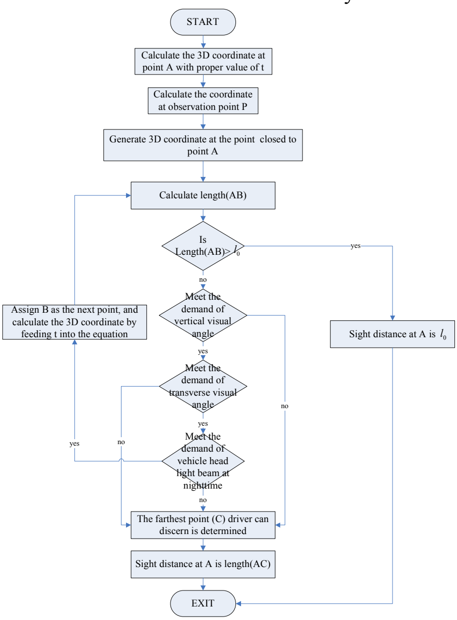


Figure 5 Flowchart for H3DSD calculation

Consistency Evaluation of Interchange Alignments

Shengwen Tu[1,2],Xiucheng Guo[1], Peter Johnston[3],Jianming He[4],Shengwu Tu[5]
[1]School of Communication, Southeast University, Nanjing 210096, China;
[2]School of Highway Engineering, Changsha University of Science & Technology, Changsha 410076, China;PH(086)15951856796; e-mail: tusw2004@163.com
[3]Graef, Anhalt, Schloemer, and Associates, Inc
[4]China Highway Engineering Consulting Group LTD, Beijing 100097, China
[5]Jiangsu Transportation Research Institute, Nanjing 210000, China;

Abstract: The concept of design consistency evaluation provides highway designers with a proactive tool for roadway safety improvement. Previous research has focused on the development of evaluation models for basic highway segments, while little research has considered interchange ramps. Interchange ramps differ from major roadways in the complexity of vehicles' acceleration and deceleration activities on them. The objective of this paper is to develop a practical consistency evaluation method for interchange alignments based on the theory of operating speed. To begin with, by giving a profound analysis of the moving characteristic of vehicles on ramps, the speed-change process on ramps is divided into five phases. Next, operating speed along highway ramps and speed differential between successive elements of ramps are calculated based on regression models developed by previous research. Three criterions are then established to evaluate the design consistency of ramp alignments. Lastly, a real life example is presented to illustrate the proposed method.

Introduction

Attempts to learn about and to provide information to the designer regarding design consistency and driver expectancy have been the subject of several major research projects and reports.(1)(2)(3)In general, they can be grouped into the following areas: vehicle operations-based consistency, roadway geometrics-based consistency, driver workload, and consistency checklists. Among the different existing measures for design consistency evaluation, the operating speed (V_{85}) approach is the most efficient and quantified measure.(4)

Operating speed is defined as the 85th percentile of the distribution of speeds selected by drivers in free-flow conditions on a particular location of the highway alignment. Although much research has been carried out to predict operating speed and to evaluate design consistency of highways, most of them concerned basic highway segments. Little research has considered consistency evaluation of interchange alignments. Interchange ramps differ from major roadways in the complexity of vehicles' acceleration and deceleration activities on them. The objective of this paper is to develop a practical consistency evaluation method for interchange alignments based on the theory of operating design.

Moving characteristic of vehicles on ramps

Moving speed of vehicles on ramps is usually lower than that of major roadways due to limitations including turning traffic volume, site terrain, located objects, etc. Therefore, drivers leaving a highway at an interchange need to reduce their speed as they exit to a ramp. On the other hand, drivers entering a high speed road facility from an entrance ramp are required to accelerate until the desired speed is reached. These speed-change activities are the principal differentiators between ramps and major roadways. According to Yang(5), this speed-change process can be divided into five phases which are illustrated in Fig. 1.

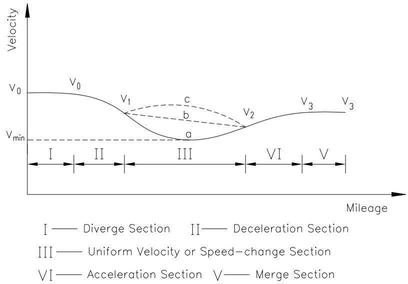


Fig.1 Speed-change Process of Vehicles on Ramps

1. Phase I: Diverge

In this phase, vehicles leave the through traffic of the major road and diverge to deceleration lanes. The speed of vehicles in this phase is close to that of the through traffic on the major road, v_0.

2. Phase II: Deceleration

This phase refers to the process when vehicles diverge from the major road and start to decelerate until exit from the major road. It happens on the full length of the deceleration lane. The beginning speed of deceleration is close to the speed of the through traffic(v_0), and the final speed is the speed of exit(v_1).

3. Phase III: Uniform Velocity or Speed-change Process

Vehicles move on the ramp proper in this phase. This is the procedure that vehicles undertake when exiting a major road and entering the connecting road. Theoretically vehicles should be running with a uniform velocity v_1 in this phase. However, speed-change usually happens during this phase due to the influence of ramp geometric conditions and other factors. So with a beginning speed of v_1, vehicles reach a final speed of v_2 at the end of the ramp proper after an acceleration or deceleration process.

4. Phase VI: Acceleration

This phase happens on the full length of the acceleration lane. The beginning speed of acceleration is the speed of entrance(v_2), and the final speed is close to the speed of the through traffic(v_3) on the other major road.

5. Phase V: Merge

In the last phase, vehicles leave the acceleration lanes and merge into the through traffic of the other major road. The speed of vehicles in this phase is close to v_3.

Consequently, the process that vehicles complete on ramps can be divided into five phases, with a speed-change procedure of $v_0 \rightarrow v_1 \rightarrow v_2 \rightarrow v_3$.

Operating speed prediction

In the past 50 years, a number of models have been developed to predict the operating speed at curved sections along major roadways. But most of them only considered basic highway segments. Vehicle behavior on ramps is complicated due to the diverge, merge, acceleration, and deceleration phases discussed above. Therefore, operating speeds along ramps should be analyzed differently than those speeds on a main road segment.

In the following, a ramp is divided into three parts to predict operating speeds on each part:

1. Deceleration lane

In China, the deceleration lane is usually designed in a direct taper form as shown in Fig. 2. The beginning speed of deceleration lane v_0 is hypothesized to be close to the design speed of the main road. For example, if the design speed of the main road is greater than 80 km/h, v_0 equals 84 to 88 percent of it. On the other hand, if the design speed of the main road is less than 70 km/h, v_0 equals 90 to 94 percent of it. (6)

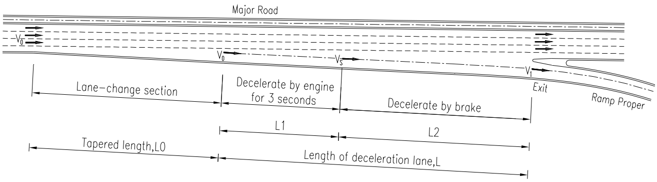


Fig.2 Illustration for Direct Taper Deceleration Lane

The length of deceleration lane is given below according to AASHTO.(6)

$$L=L1+L2=\frac{v_0 t_1}{3.6}-\frac{1}{2}\alpha_1 t_1^2+\frac{1}{25.92\alpha_2}\left[(v_0-3.6\alpha_1 t_1)^2-v_1^2\right] \quad (1)$$

where: α_1 :deceleration rate by engine(m/s^2);

α_2 : deceleration rate by brake(m/s^2);

t_1: deceleration time by engine, normally taking 3 seconds;

v_1: the speed at the end of the deceleration lane(km/h).

Therefore, the speed v_1(km/h) at the end of deceleration lane can be obtained by:

$$v_1=\sqrt{(v_0-3.6\alpha_1 t_1)^2-25.92\alpha_2(L-\frac{v_0 t_1}{3.6}+\frac{1}{2}\alpha_1 t_1^2)} \quad (2)$$

2. Ramp proper

The prediction of operating speeds on the ramp proper could be treated the same as basic highway segments. But the geometric characteristics of ramps are more complicated than that of basic highway segments. For example, most ramp elements are curves rather than tangents, and the combination of horizontal and vertical alignments on ramps are both common and complicated.

Numbers of regression equations for operating speed have been developed based on a number of research papers in China(7). Some of them are listed in Table.1. Before these equations are used, the analyzed highway sections should be separated into segments according to horizontal radius, tangent length, grades, and the combinations of horizontal and vertical alignments. Each segment which has similar geometric properties is treated as an independent prediction unit. The speed at the entrance, middle section, and exit of the prediction unit are then calculated. Spiral transitions, which were proved not to significantly affect the operating speed of vehicles traversing horizontal curves, are regarded as parts of curves connecting to them.

Table 1. Regression Equations for Operating Speeds on Horizontal Curves in China (for passenger cars)

Curves connecting from	Equation
Entrance tangent-curve	v_{middle}= -24.212+0.834v_{in}+5.729lnR_{now}
Entrance curve -curve	v_{middle}=1.277+0.924v_{in}+6.19 lnR_{now} -5.959lnR_{back}
Exit curve-tangent	v_{out}= 11.946+0.908 v_{middle}
Exit curve- curve	v_{out}=-11.299+0.936 v_{middle} -2.0601 lnR_{now} +5.203lnR_{front}

where:

v_{middle}: speed at the middle of curve(km/h); v_{in}:speed at the entrance of curve(km/h);
R_{now}: radius of current curve(m); R_{back}:radius of next curve(m);
v_{out}: speed at the exit of curve(km/h); R_{front}: radius of previous curve(m);

3. Acceleration lane

The speed v_2 at the entrance of the acceleration lane is equal to the speed at the exit of the ramp proper, which is determined by the models mentioned above. The speed v_3 at the exit of the acceleration lane is given by:

$$v_3 = \sqrt{25.92\alpha L + v_2^2} \tag{3}$$

where: α : acceleration rate (m/s^2);
L: the length of acceleration lane(m);
v_2: the speed at the entrance of acceleration lane(km/h).

Therefore, the speed profile along the full length of ramps can be developed by analyzing three stages. This speed profile will be used for the consistency evaluation below.

Speed differential

In China, the current approach to calculate the speed differential is based on calculating the operating speed of the drivers on a particular curve and the previous or next curve and then subtracting these two values and naming it as the speed differential value ($\triangle V_{85}=V_{85\,i}-V_{85\,i-1}$). However, this methodology is based on assuming that speed distribution on the successive elements is the same and this assumption is not necessarily accurate(8)(9). According to the two referenced studies, speed distributions at the curved and tangent sections are not the same, and thus the simple subtraction of the operating speed values should not be performed.

Following this point of view, McFadden and Elefteriadou(9) introduced a new parameter, 85th percentile of maximum speed reduction (85MSR) to describe the speed differential. Additionally, Misaghi1(4) introduced another parameter, $\triangle_{85}V$ to represent speed differential and relevant prediction models were developed. The statistic "$\triangle_{85}V$", the 85th percentile speed differential, is defined as the differential speed not exceeded by 85% of the drivers traveling under free-flow condition. Based on the observations of these studies(4)(8)(9), it was concluded that the simple subtraction of operating speeds at the approach tangent and the middle of the curve underestimates the real values of speed differential.

What should be stated is that all these studies focused on the case that vehicles moved from a fairly long tangent to a curve. But the horizontal alignments of ramps are mostly made up of curves which are connected together one by one. Long tangents are rarely used in ramp alignments. As a result, parameters and models mentioned above may not be suitable to interchange ramps. Giving consideration to this point, we simply use subtraction of operating speeds between successive elements of ramps($\triangle V_{85}$) to describe speed differential.

Design consistency evaluation

Based on the theory of operating speed, three evaluation criterions for consistency of ramp alignments are given as below.

1.Moving characteristic consistency

This criterion means that the speed profile developed by the above methods should comply with the moving characteristic of vehicles on ramps as shown in Fig.1. Among three cases in Fig.1, case a and case b are preferred. Case c should be avoided as best as one can because vehicles' accelerating ahead of schedule may result in higher accident frequency at the exit of ramp.

2.Design speed consistency

Design speed consistency is evaluated based on the difference between the 85th percentile speed(V_{85}) and the design speed(V_D) for individual curve segments as $|V_{85}-V_D|$.This measure aims to ensure that the design speed is in balance with actual driver speeds and behavior.

3.Operating speed consistency

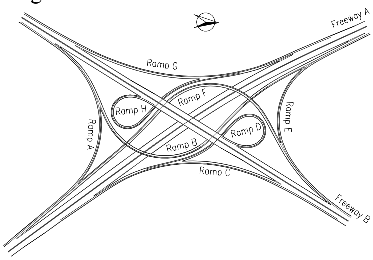
Operating speed consistency is based on the speed reduction between two consecutive elements. This safety measure provides a consistency evaluation for successive elements along a roadway rather than individual elements. As mentioned above, subtraction of operating speeds between successive elements ($\triangle V_{85}$) is used to describe the speed reduction. Table 2 illustrates the cut-off points for "Good", "Fair" and "Poor" designs applicable to operating speed-based consistency measures.

Table 2. Operating Speed Based Criterion

Evaluation	Design speed consistency	Operating speed consistency
Good	$\lvert V_{85}-V_D \rvert \leq 10$ km/h	$\triangle V_{85} \leq 10$ km/h
Fair	10km/h$<\lvert V_{85}-V_D \rvert \leq 20$ km/h	10km/h$<\triangle V_{85} \leq 20$ km/h
Poor	$\lvert V_{85}-V_D \rvert > 20$ km/h	$\triangle V_{85} > 20$ km/h

A real life example

The proposed measure is applied to a real life example: a system interchange between two freeways. The interchange has 8 ramps providing service to all turning traffic. It is located in a flat rural area with original ground elevation between 1.0 and 3.2 meters. The horizontal layout of the interchange is shown in Fig.3. Because of the limited space of paper, only the evaluation results of ramp B are given below.



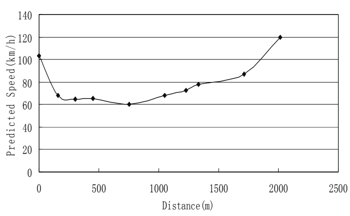


Fig.3. Horizontal Layout of The Real Life Example Fig.4. Predicted Speed Profile for Ramp B

Ramp B is a half-directional ramp which has a design speed of 60 km/h. Fig. 4 illustrates its speeds predicted along the full length of ramp including deceleration lane and acceleration lane. The result indicates that the speed profile is approximately in accordance with the moving characteristic of vehicles on ramps as shown in Fig.1. Moreover, differences of operating speeds between successive elements ($\triangle V_{85}$) are all less then 10km/h, which demonstrates that the operating speed consistency of the ramp is good. However, the $|V_{85} - V_D|$ of some sections is greater than 10km/h, especially at the exit of ramp. Therefore, part of ramp alignments need to be adjusted in order to get better consistency.

Conclusions

The concept of design consistency evaluation provides highway designers with a proactive tool for roadway safety improvement. Although much research has been carried out to predict operating speed and to evaluate design consistency of highways, most of them aimed at basic highway segments. The moving characteristic of vehicles on interchange ramps are complicated, owing to their inherent deceleration and acceleration behavior. Development of specific evaluation measures for alignments consistency of interchange ramps is therefore needed.

Based on the analysis of moving characteristic of vehicles on ramps, the procedure for predicting operating speed and three evaluation criterions for alignments consistency of ramps are described in this paper. A real life example indicates that the proposed measure has good practicality and can give the consistency of ramp alignments an objective evaluation.

Reference

1. Fitzpatrick K., Ingrid B. Anderson, Karin M. Bauer.*Evaluation of Design Consistency Methods for Two-Lane Rural Highways, Executive Summary*. FHWA-RD-99-173, FHWA U.S. Department of Transportation, 2000

2. Fitzpatrick K., Mark D. Wooldridge, Omer Tsimhoni,etc. *Alternative Design Consistency Rating Methods for Two-Lane Rural Highways*. FHWA-RD-99-172, FHWA U.S. Department of Transportation, 2000

3.Fitzpatrick K., L. Elefteriadou, D. Harwood, J. Collins, etc. *Speed Prediction for Two-Lane Rural Highways*.Report FHWA-RD-99-171, FHWA U.S. Department of Transportation, 2000

4. P. Misaghi1,Y. Hassan. *Modeling Operating Speed and Speed Differential on Two-Lane Rural Roads*. Journal of Transportation Engineering, 2005, 131(6):408-417

5.Yang Shaowei. *The Characteristic and Plane Alignment of Automobiles Moving at Interchange Ramps*. Journal of Xi'an Highway University, 1998,18(3):164-167

6.AASHTO.2004.A Policy on Geometric Design of Highways and Streets. American Association of State Highway and Transportation Officials. Washington D.C.

7.Ministry of Communications of Peoples Republic of China, 2004. *Safety Evaluation Manual for Highway Project,* Beijing

8.Hirsh, M. . *Probabilistic approach to consistency of highway alignment*.J. Transp. Eng., 1987,113(3): 268–276.

9.McFadden, J.,and Elefteriadou,L..*Evaluating horizontal alignment design consistency of two-lane rural highways: Development of new procedure*. Transportation Research Record, 1737, Transportation Research Board, 2000,Washington, D.C., 9–17.

Construction Techniques for the Wuhekou Cable-Stayed Bridge

Yongan Wang[1], Zhusheng Jiang[1], Shitong Liu[1], Zhengqun Xue[2] and Scot Parkhurst[3]

[1]Jiangsu Expressway Construction Department, Nanjing, China, [2]hydraulic Research Institute, Nanjing, China, [3]WSP Group PLC, United Kingdom.

Abstract: The Wuhekou Cable-Stayed Bridge has a central span of 370m, which features two towers, two cable planes, and a floated deck system. The 38.6m monolithic deck structure is the widest in China. The standard deck section is a fully pre-stressed concrete box girder, which is strengthened by spaced cross beams and longitudinal stringers, providing enough bending and torsion stiffness for this wide long-span girder. The deck was constructed using a balanced cantilever method with specially designed travelers, supported with a cable on each cycle during casting. Epoxy coated parallel strands and high performance concrete were used to improve the bridge's durability.

1. Introduction

The Wuhekou Bridge is a large cable stayed structure spanning several navigable waterways. The bridge is part of the highway linking YanCheng to Xuzhou and Nanjing to Lianyungang in JiangSu province. The overall length of the bridge is 2,062m which includes a 370m central span and two 152m intermediate cable stayed spans. The bridge construction comprises a two towers twin cable stayed, two-side-box, fully floated, pre-stressed concrete cable-stayed structure. The 854m long north approaches and the 525m south approaches are partial pre-stressed concrete continuous box girders. The bridge was constructed on a horizontal alignment curve of R=5,500m and vertical alignment curves of R1=40,000m R2=20,000m R3=23,075m. The 'H' shaped cable towers are 137.1m high (above from the pile cap). The cable tower foundations are supported on pile groups, each containing 46 number 250cm diameter cast in-boring piles. The main deck structure is 38.6m in width Figure 1 show the elevation of the bridge.

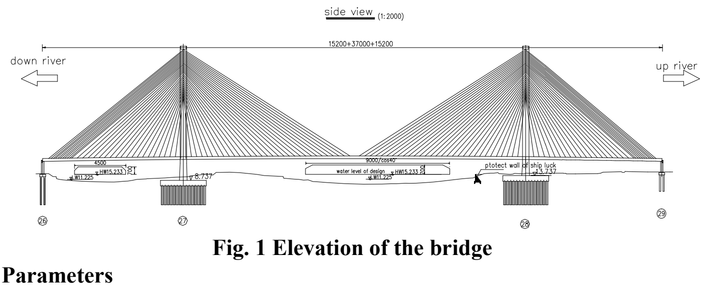


Fig. 1 Elevation of the bridge

2. Parameters

Vehicle speed: 120km/h, six lanes bi-directional. Designed load: lane load-I, truck -super 20, and trailer-120. Earthquake intensity: peak acceleration 0.1g.

Navigable clearance: 7m net height above navigation water level, net width is no less than 90m. Designed flood frequency: 1/300 years. Vertical action: dead load, lane load-I.

Wind action: combined wind and automobile loading are calculated as 26.3m/s on

the deck.
Temperature effects: the designed closure temperature is 15 to 20 ^{O}C, with design temperature change of +20 ^{O}C or -25 ^{O}C.

3. Technical Challenges

The foundation group pile for main pier includes 92 cast in-situ boring piles, each 2.5m in diameter and between 90m to 95m in length, penetrating into the underlying complex geological formation which comprises Quaternary sediments. The bridge contains 177k cubic meters of pumped concrete and 20.6k tons of steel, and the cable stays comprise 1.95k tons in total. Flow high performance concrete was cast monolithically and self compacting was used to improve the bridges durability and achieve high early strength. The main central span is 370m in length with a deck cross-section width of 38.6m. The two-side-box pre-stressed concrete deck structure is complex. Deck segments of the main girder are constructed by a cantilever method using a specially designed traveler with its fore-end stayed on a cable. The pylons were constructed to a height of 137.1m.

4. The structural constructions

The internal forces of the friction piles under foundations were calculated by the "m" method. 90m long piles were adopted for the north pylon (No.27), and 95m long piles for the south pylon (No.28). The pile cap for each pylon required 9,830 cubic meters of concrete. Each pylon pile cap contains three upper layers of 32mmΦ rebar and four layers on the lower surface.

The cable tower was built as a pre-stressed concrete structure. Within the concrete structure a rigid skeletal frame to assist rebar fixing and formwork support. The upper tower is 77.0m in height, which has a box section. The column has a longitudinal width of 7.0m and a 1.2 thick wall, while these figures are to 4.5m and 0.8m transversally. Internal walls have a fillet of 0.5×0.5m and at each hatch corner formed a circular fillet of R=0.3m. The middle tower is 47.0m in height. The longitudinal width of the column varies from 7.0m at the bottom of the upper transverse beam to 8.0m at the center of the lower transverse beam. The outer side of the wall inclines to the central line of the bridge with a gradient of 12:1, as the transverse width of the column covers the same value as the upper tower; however, at the bottom it varies to 1.7m longitudinally and 0.8m transversally, with a 1.0m×0.5m fillet inside and a circular fillet of R=0.3m outside. The lower tower is 13.1m in height. It comprises a four-cell structure in the shape of square cells. Thickness of the external wall tends to be 1.3m at the bottom and 1.0m, 0.8m.in the middle. The transverse widths are 11.0m. The gradient attached to the wall is 2.8:1 inside and 6.1:1 outside. At the anchorage cable zone of the upper tower, annulus tendons were installed to balance the horizontal distributed forces induced by the stayed cable, which were set by strands of 19-φ^{j}15.24 and 12-φ^{j}15.24. The upper transverse beam has a box section with a dimension of 6.0m×6.0m×0.7m and a fillet of 0.5m×0.5m inside while the lower beam covers the same with an exception for its 6.8m width. transverse and longitudinal caging devices on the towers, was supported with rubber bearings on the lateral face of the main girder, while rubber stoppers at the lower transverse beam of the tower, prevent impact movement on both sides. The main girder is a pre-stressed concrete structure. The section of the standard segments and tower zone segments are shown in Fig 2.

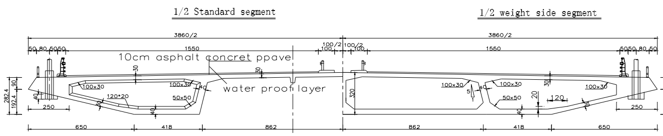


Fig. 2 Dimension of typical cross section

Floors between the side boxes of standard segments are omitted for clarity. Thickness of roofs and webs are gradually increased to 0.6m in the tower zone segments, where floors are installed, without termination until they meet the standard sections. Segments in the intensive cable zone of the side span have a length of 2.5m and are boxed in section were cast with iron-sand concrete to form a weight density of 3.5t/m^3. To accommodate the internal forces, there are two small stringers located under the deck of the main girder. The bridge is constructed from 105 blocks, 86 of which are cast on travelers; among them 82 standard segments of girder each 1s 6.0m in length. Fourteen segments (44m) are located in the tower zone and two in the intensive cable zone (side spans 31.6m), were cast-in-place on brackets. Three on the closure zone of side and main span were cast on hangers utilizing a cantilever method with a length of 2.0m. Transverse beams of eight types are set up wherever stayed cables met the main girder. Fringe transverse beams are located in the end of main gird. The longitudinal pre-stressing of the main girder can be divided into two kinds: thick finish rolling deformed bar in deck and strands as well as thick finish rolling deformed bar in boxes.

The cables are intensively displaced on the altitude ranged from 83.7m to 146.0m. Each branch has 31 pairs of cables, made of ovm250 epoxy coated strands and corresponding anchorages. There are five kinds of cables in all. The cables have four protection layers. The 1st is epoxy-coated layer (and a grease layer for unbounded tendons), single layer of hot-extruding HDPE, as well as HDPE sheathing of whole cable. The configuration of a cable is shown in Fig. 3.

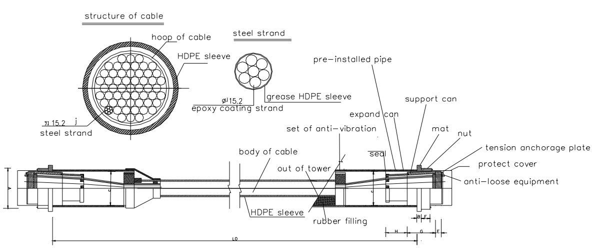


Fig. 3 Stayed cable

5. Materials

The towers are made of C50 grade high performance concrete and the girder is C60. The concrete is composed of 5mm to 25mm coarse limestone aggregate, and a medium sand fineness module of 2.6. The concrete also has some fly ash of level 1. The concrete's three day age compressive strength exceeds 85% strength of the 28 day designed, and its elastic modulus is more than 3.92 x10^{4}mpa while its 90 day degree of creep less than $21x10^{-6}$mpa, table1 and table2 show their properties. Concrete was nearly self-flowing, to improve penetration through the rebar under minimum compaction with vibration. C30 grade concrete was used in pier body, abutment, cap and C25 concrete was used in the pile cap of the approach, collar beam and foundation piles. Stress amplitude of the strands is no less than 200mpa， which complies with the relevant specifications for low-relaxation strands in GB/T 5224-1997.

Table 1 Mixed proportion of C50 high performance concrete in towers

Cement (kg/m^3)	Fly ash (kg/m^3)	Sand (kg/m^3)	Aggregate (kg/m^3)	Water (kg/m^3)	Additive (kg/m^3)
419	Level I : 74	676	1057	180	UC-□ : 4.93

Table 2 Mixed proportion of C60 high performance concrete in main girder

No	Mixed proportion	Cement (kg)	Fly ash (kg)	Additive(kg)		Water (kg)	Slump 21±2cm	Slump <2cm/h	Diffusion 45±5cm	Strength (mpa/7d)
				uc-2	Mabe					
1	1:1.3:2.04:0.316:0.013	464	52	N /A	6.71	163	21	h，3cm	50	59.4
2	1:1.21:1.9:0.316:0.013	490	54	6.8	N/A	172	20	2	43	65.5

6 The construction method

The pylons are on Quaternary sediments, which comprise soft soils, liquefiable soil and hard clay. The steel pile casing required varying lengths of 4.0m, 6.5m, 8.0m and 12.0m respectively, depending on geological conditions. During construction, the casing was firstly buried 2.0m deep by hand and then sank into place using DZ 150 vibration hammers. Reinforcement cages were manufactured segmental (12m×7+12.96m=96.69m) to facilitate installation. The main pile reinforcing bars were jointed using cold extrusion sleeves.

The concrete of the main pier cap was cast in two lifts. Firstly a 3.3m high, $5407m^3$ concrete lift is poured, followed by a 2.7m high, $4424m^3$ concrete lift. The 27^{th} pier cap cofferdam construction was used to brace the foundation pit, and comprised steel sheet piles and 630mmØ x 3mm locked steel pipe piles. The foundation pit for the 28^{th} pier cap was excavated to a depth of about 8.8m. The construction of the pile cap took place following draining of surface water and dewatering. Dewatering of the foundation pit was achieved through a two-stage well point system, which was supplemented with deep well points work as required. A low heat of hydration concrete was used the concrete slump constant is 16-20cm and initial setting time was not less than 30h. During construction, shortening the initial setting time of the last layer concrete following the first pouring and the second pours would control the temperature of mass concrete.

The construction technology of turn over forming was adopted in the casting of concrete of main pylons. The transverse beam was built by cast-in-site concrete technique with steel pipe brackets. The pylons and their corresponding transverse beams were constructed simultaneously. Two 125t.m tower cranes were located on the east and west sides of each tower shaft, connected with embedded fixings every 30m to form a supporting wall. They were used to lift and install the formworks and brackets for the transverse beams. The 2t single-cage elevators were incorporated into the south of eastern and western shafts of the pylons, and were used as the main elevators for workers and equipment. 75m^3/h and 50 m^3/h concrete mixers and two trailer pumps are used for concrete No 27 and No 28 pylons respectively. The construction platform outside pylons were erected using scaffolding to form a closed working platform as shown in Figure 4.

Fig. 4 Main pylon Construction

Fig. 5 Balanced cantilever construction

The concrete for the transverse beams were cast in time using formwork supported by steel pipe brackets. Seven rows and two arrays of 8.0m long, 630mmØ x 6mm thick steel pipes were welded to the embedded fixings on the top of pile caps. These were adopted as the upright columns of the lower transverse beam. While for the upper transverse beam, 42.0m long, 630mmØ x 6mm thick steel pipes of five rows and two arrays were used and welded to the embedded fixings on the top of the lower transverse beam. Four 400mmØ x 5mm thick steel pipes with a spacing of 10.0m were used and welded together. Single steel strand was separately stretched, supplemented with a vacuum mud jack to guarantee the quality. A symmetric stretching method was adopted, and double control criteria of elongation values and tension forces carried out in applying pre-stress.

Steel pipe brackets are used in the 0-3# segments of the main girder. Platforms supported by underwater pile piers with bowl-buckle supporting frames were adopted in much of the rigging zone of side span and end transverse beam near pier number 27. And in those elements located near pier number 28, strip foundations of concrete with bowl-buckle supporting frames were used. The construction procedure of the main girder was: casting the concrete of 0-3# segments on brackets firstly, stretching longitudinal and transverse strands. Following this the balanced cantilever construction method was adopted with a frontal support stiff –traveler to support the construction of 6-m long standard segments. The stiff-traveler weighed 224t, the standard segment 450t and the heaviest segment 540t. The ratio of dead load to live load was 0.415. The design of the traveler accommodated room for stretching devices and transition devices. It can withstand reversal and loosen clips prevent the slip of cables under low stress. A steel pipe tied arch system with large stiffness, lightweight and easy down movement was adopted in internal molds and brackets to facilitate fast movement of the traveler. The actual longitudinal and transverse deformation of the traveler was respectively 1/2000 and 1/1000. The traveler is shown in Figure 5. In hot weather, construction formwork was removed early, watering adopted early, while during winter season construction measures such as heated water, laying insulated formwork ensure the appropriate temperature of concrete of the main girder during construction. Stretching strands twice and so on were adopted to reduce concrete cracking. The concrete of main girder was cast symmetrically from the front to the back, from the center to the side. It was cast in layers of 50cm thick from the bottom slab of the box girder to the web plate and then to the roof. The grout holes and vibrating mouths were properly distributed to guarantee the symmetric construction of the two sides of the main girder in the same direction of the pylon.

After the baiting of cables, they were installed, pre-stressed and anchored one by one. Following this integral pre-stressing and general maintenance was carried out and grease filled into anchorages points. The construction of cables included fabrication of parallel steel cable system, rigging of cables matched with the casting of main girder, pre-stressing, anchorage, adjustment of cable force, transition of cable force, installation of waterproof cover, grout filling in the anchorage device and grease filling in the protection cover.

7 Conclusions

The construction contract of Wuhekou cable-stayed bridge was signed on March 12th, 2003 and on December 20th, 2005 the bridge was opened. The completed bridge is shown in Figure 6.

Fig.6 The Completed Wuhekou Cable-stayed Bridge

The main girder construction of the bridge was adopted the balanced cantilever method with a frontal supported traveler. A system of lifting, anchoring,, positioning and adjusting of travelers was developed. The technique made the casting of girder and erection of cables integrated, simple and convenient.

Because of the pre-cast girder with long span and wide deck, parallel strand cables with epoxy coating, main girder with double side boxes, fully pre-stressed components and tiny longitudinal stringers were applied. The strength detection of concrete and static test of the whole bridge demonstrated that under the action of load and environmental factors, the integral longitudinal and transverse flexural rigidity and torsion stiffness of main girder, pylons were rational. Large-scale pre-stressed concrete cable-stayed bridge contains many special details where the structural elements are complicated and stress variation is great, although it needs to continue investigation that large-scale pre-stressed concrete cable-stayed bridge with long span and wide deck can find a wide application.

Wind-resistant design of cable-stayed-suspension hybrid bridges

X.J. Zhang[1] and David A. Stern[2]

[1]Department of Civil Engineering, Zhejiang University of Technology, Hangzhou 310014, China; email: xjzhang@zjut.edu.cn
[2] Department of Civil & Environmental Engineering, University of California, Los Angeles 159310, USA; email: davidstern9@gmail.com

Abstract

The cable-stayed-suspension hybrid bridge is a cooperative system of the cable-stayed bridge and suspension bridge. It takes the advantages of each and also makes up for some deficiencies of the two bridge structures, and therefore has better spanning capacity. By analyzing a cable-stayed-suspension hybrid bridge with main span of 1400 m, effects of design parameters including the cable sag to span ratio, the suspension to span ratio, the side span length, the layout of stay cable planes and the subsidiary piers in side spans etc on the wind stability are investigated. Utilizing 3D nonlinear aerostatic and aerodynamic analysis, and based on the wind stability, the optimal values of these design parameters are determined.

1 Introduction

The cable-stayed-suspension hybrid bridge is a new cable-supported bridge developed from the traditional cable-stayed bridge and suspension bridge. The design benefits from the main advantages of each system, and also makes up for the deficiencies in the structural behavior, construction, economy and the wind stability of traditional suspension and cable-stayed bridges. Therefore this hybrid becomes an attractive design alternative for long and particularly super long-span bridges.

The idea of using cables to support bridge spans was conceived by Roebling, Dischinger, Steinmann and Gimsing etc. (Gimsing, 1997). This concept was successfully employed in the rehabilitations of some existing suspension bridges, and also frequently proposed in the design of many strait-crossing bridges(Gimsing, 1997). In 1997, the first modern cable-stayed-suspension hybrid bridge in the world was built in China with a main span of 288 meters (Meng, et al. 1999). In the 21st century, many long and particularly super long-span bridges are planned for the sea-crossing engineering projects. Many of them are projected to be built under natural conditions unsuitable for cable-stayed or suspension bridges, such as soft soil foundation, violent typhoons, and deep-water foundations. However, due to its advantages, the cable-stayed-suspension hybrid bridge becomes a competitive design alternative favorable for these bridges. Just like other cable-supported bridges, the cable-stayed-suspension hybrid bridge is also a structural system of great flexibility, but vulnerable to wind action. Therefore wind stability becomes a very important factor to be considered in the bridge design.

This paper provides a guide for wind-resistant design for cable-stayed-suspension bridges. It is based on an analysis of a bridge with main span of 1400 m, and examines the effects of design parameters including the cable sag to span ratio, the suspension to span ratio, the side span length, the layout of stay cable planes and the subsidiary piers in side spans etc. on the wind stability, and based on the wind stability, the optimal values of these design parameters are determined.

2 Description of the sample bridge

The example cable-stayed-suspension hybrid bridge, as shown in Fig.1, consists of a main span of 1400 m and two side spans of 319 m. This design was proposed for construction in the east channel of Lingding Strait in China (Xiao, 2000). The central span consists of the cable-stayed portion of 788 m and the suspension portion of 612 m. The distance of the two main cables is 34 m, the cable sag to suspension portion ratio is 1/10, and the spacing of hangers is 18 m. The stay cables are anchored to the girder at 18 m intervals in the central span and 14 m in the side spans. The deck is a steel streamlined box steel girder, 36.8 m wide and 3.85 m high. The tower is a door-shaped frame with 3 transverse beams, and its height from the deck level is about 194 m.

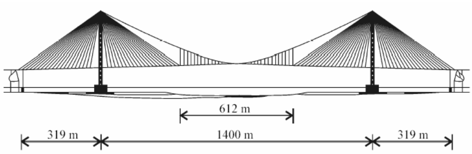


Fig.1 Elevation of the example cable-stayed-suspension hybrid bridge

3 Aerostatic stability analysis

Under the wind attack angle of 0°, parametric analysis on the aerostatic stability is conducted by three-dimensional nonlinear aerostatic analysis (Zhang, et al. 2002). In the analysis, the drag, lift and twist moment components of the aerostatic load are considered for the deck. Due to the similarity of the deck's aerodynamic shapes in the example bridge and the Runyang Bridge, the aerostatic coefficients obtained from the wind tunnel test of the Runyang Bridge (Chen and Song, 2000) are used herein. For cables, hangers and towers, only the drag component is considered, and the corresponding drag coefficient is 0.7 for cables and hangers and 2.0 for the towers.

3.1 The cable sag to span ratio

It is to be noted that the cable sag to span ratio herein is defined with respect to the suspension portion. As the cable sag increases, the lateral and torsional displacements are both increased dramatically, whereas the vertical displacement is decreased. Conversely, as the cable sag decreases, the lateral and torsional displacements are both decreased, whereas the vertical displacement is increased. Therefore, the disadvantages of a sag that is too big or too small is confirmed analytically for the aerostatic stability of cable-stayed-suspension hybrid bridges, and adjusting the cable sag to span ratio to approximately 1/10 reduces deck displacement and improves aerostatic stability for cable-stayed-suspension hybrid bridges.

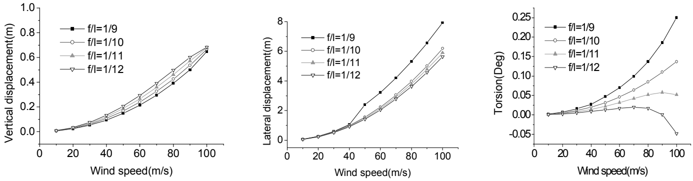


Fig.2 Effect of the sag to span ratio on the deck's displacements at midspan

3.2 The suspension to span ratio

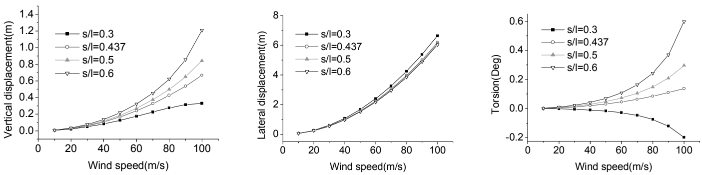


Fig.3 Effect of the suspension to span ratio on the deck's displacements at midspan

In general, the suspension length has little influence on the lateral displacement, but significant influence on both the vertical and torsional displacements. As the suspension portion increases, the cable-stayed-suspension hybrid bridge behaves as a suspension bridge, with decreased structural stiffness. The vertical and torsional displacements are therefore increased greatly. Inversely, as the suspension portion decreases, the lateral and torsional displacements are both increased. Therefore, a suspension portion that is too short or too long is not favorable for cable-stayed-suspension hybrid bridges, and viewed from the aspect of aerostatic stability, a suspension to span ration ranging from 0.4 to 0.5 is optimal.

3.3 Side span length

Fig.4 shows the evolution of the deck's displacements at midspan with wind speed under the side spans of 394m(equal to the cable-stayed portion in central span) and 314m(shorter than the cable-stayed portion in central span) respectively.

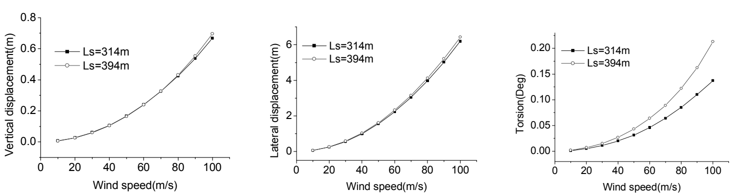


Fig.4 Effect of side span length on the deck's displacements at midspan

As the side span increases, the deck displacements are all increased, particularly the torsional displacement. The fact can be attributed to the decrease of structural stiffness in the case of long side spans. Therefore considering the aerostatic stability, a short side span is favorable for cable-stayed-suspension hybrid bridges.

3.4 Layout of the stay cable planes

For the example bridge, the stay cable planes are inclined inward. To investigate the effect of the layout of stay cable planes on the wind stability, two cases are assumed: one where the stay cable planes are vertical, and another where the stay cable planes are inclined outward.

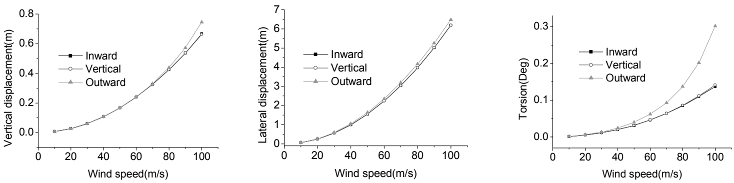


Fig.5 Effect of layout of the stay cable planes on the deck's displacements at midspan

As shown in Fig.5, the layout of stay cable planes has very little influence on the vertical and lateral displacement, but has significant effect on the torsional displacement. Structural displacements are basically identical for the cases of vertical and inward-inclined stay cable planes, however in the case of outward-inclined stay cable planes, the vertical and lateral displacements are both slightly increased, while the torsional displacement is significantly increased. Therefore considering the aerostatic stability, vertical and inward-inclined stay cable planes are favorable for cable-stayed-suspension hybrid bridges.

3.5 Subsidiary piers in side spans

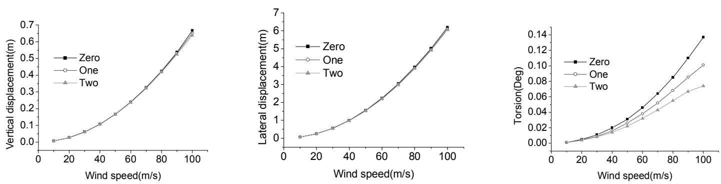


Fig.6 Effect of the subsidiary piers in side spans on deck displacements at midspan

The vertical and lateral displacements are marginally influenced by the subsidiary piers, though the torsional displacement is reduced in the case where the subsidiary piers are installed. As a result, the improvement of aerostatic stability by the subsidiary piers in side spans is very limited.

4 Aerodynamic stability analysis

Under the wind attack angle of 0°, effects of design parameters on the aerodynamic stability of the bridge are investigated by three-dimensional nonlinear aerodynamic analysis (Zhang, et al. 2002). In the analysis, the aerodynamic coefficients obtained from the wind tunnel test of the Runyang Bridge (Chen and Song 2000) are used herein, and the structural damping ratio is taken as 0.5%.

4.1 The sag to span ratio

Table 1 Effect of the sag to span ratio on the critical wind speed

The sag to span ratio	1/12	1/11	1/10	1/9
Critical wind speed(m/s)	80.1	85.5	92.7	103.7

The critical wind speed is increased greatly as the cable sag increases. To effect an increase of cable sag, the tower's height and the inclination angles of the stay cables are both increased. This improves the supporting efficiency of the stay cables and provides further structural stiffness to the bridge, and leads to the increase of modal frequencies particularly the torsional frequencies. Viewed from the aspect of aerodynamic stability, the cable sag should not be too small for cable-stayed-suspension hybrid bridges.

4.2 The suspension to span ratio

Table 2 Effect of the suspension to span ratio on the critical wind speed

The suspension to span ratio	0.3	0.437	0.5	0.6
Critical wind speed(m/s)	99.8	92.7	88.0	79.8

As shown in Table 2, a short suspension portion is aerodynamically favorable for cable-stayed-suspension hybrid bridges. The fact can be attributed to the reduction of structural stiffness and further the modal frequencies of the bridge in the case of long suspension portion.

4.3 Side span length

Table 3 Effect of the side span length on the critical wind speed

Side span length(m)	394	314
Critical wind speed(m/s)	90.8	92.7

As shown in Table 3, short side spans are confirmed analytically to be aerodynamically favorable for cable-stayed-suspension hybrid bridges. The fact can be attributed to the decrease of modal frequencies in the case of short side spans.

4.4 Layout of the stay cable planes

Table 4 Effect of the layout of stay cable planes on the critical wind speed

Layout of the stay cable planes	Inward	Vertical	Outward
Critical wind speed(m/s)	92.7	92.9	104.8

The outward-inclined stay cable planes are confirmed analytically to be most aerodynamically stable. This phenomenon also occurs for cable-stayed bridges. The fact can be attributed to the increase of modal frequencies in the case of outward-inclined stay cable planes. However, it is contrary to the case of aerostatic stability. Generally, aerodynamic stability of cable-supported bridges is more critical than the aerostatic stability, and therefore considering the wind stability, the outward-inclined stay cable planes are favorable for cable-stayed-suspension hybrid bridges.

4.5 Subsidiary piers in side spans

Table 5 Effect of the subsidiary piers in side spans on the critical wind speed

Number of the subsidiary piers in side spans	0	1	2
Critical wind speed(m/s)	92.7	96.5	97.7

Table 5 shows that the subsidiary piers help to improve the aerodynamic stability of the bridge. The fact can be attributed to the increase of structural frequencies and also the modal shape getting more complicated in the case of the subsidiary piers installed in side spans, which leads to a multimode-coupled aerodynamic response.

5 Conclusions

In this paper, by analyzing the cable-stayed-suspension hybrid bridge with main span of 1400 m, effects of design parameters including the sag to span ratio, the suspension to span ratio, the side span length, the layout of stay cable planes and the subsidiary piers in side spans etc. on the wind stability are investigated by utilizing 3D nonlinear aerostatic and aerodynamic analysis. The results show that the cable-stayed-suspension hybrid bridge has good wind stability as the sag to span ratio is approximately 0.1, the suspension to span ratio is between 0.4 and 0.5, and for the short side span, outward inclined stay cable planes and subsidiary piers are employed.

References

Chen, A.R., Song, J.Z., 2000. *Wind-resistant Research on the Runyang Bridge*. Research Report, Tongji University.

Gimsing, N.J., 1997. *Cable-supported bridges: Concept & Design*. John Wiley & Sons Ltd, England.

Meng, Y., Liu, D., Sun, S.H., 1999. “Study on the design of long-span cable-stayed-suspension hybrid bridges”. *Journal of Chongqing Jiaotong Institute*, 18(4), 8-12.

Xiao, R.C., 2000. “Research on the design of cable-stayed-suspension hybrid bridges”. *China Civil Engineering Journal*, 33(5), 46-51.

Zhang, X.J., Sun, B.N., Xiang, H.F., 2002. “Nonlinear Aerostatic and Aerodynamic Analysis of Long-span Cable-stayed Bridges Considering Wind-structure Interactions”. *Journal of Wind Engineering and Industrial Aerodynamics*, 90(9), 1065-1080.

Theoretical Study on Cable Tension Detection Considering Support Vibration

WEI Jinbo,[1] LI Guoqiang[2], and Mohammad Reza Razavi[3]

[1]A507 Department of Building Engineering, Tongji University, 200092 Shanghai, China; PH (86)15921534301; e-mail: weijinbo0117@126.com
[2]A520 Department of Building Engineering, Tongji University, 200092 Shanghai, China; PH (86)-21-65983313; e-mail: gqli@mail.tongji.edu.cn
[3]Mineral Engineering Department, New Mexico Tech, 801 Leroy Place, Socorro, NM, 87801, USA; PH +1 (575) 835-6447; e-mail: Mehrdad@nmt.edu

Abstract: This study presents a systematic approach to determine the cable tension when vibration of the cable supports is taken into account. A dynamic model by including support vibrations was developed for cable structures. Boundary conditions were modeled with equivalent stiffness and mass. The displacement function was obtained from the model and then energy method was used to obtain stiffness and mass matrices to find the relationship between the cable tension and frequency. We studied several cases for ignoring support conditions and very large or very small equivalent stiffness values. In these cases the calculated cable tension was less than the actual tension and the accurate results were obtained when vibration of the supports were included in the model. Therefore, in dynamic analysis of cable structures, vibration of the supports should be included in the model.

Keywords: support vibration, cable tension, frequency, energy method

1. Introduction

Recent improvements in developing advanced materials, modeling, and construction technology has increased application of the cable structures in industry. Cables are flexible tension only members, which only a relatively small stress or strain may tend to large displacements (Zhang Kaiying 2005). True determination of the internal tensile force of a cable in a cable structure is very critical in both construction and maintenance of the structure. In construction the cable tension affects the overall shape of the structure as well as the distribution of the cable internal forces. On the other hand the cable tension is sensitive to the structure deformation, so in maintaining the structure it is used as a measure of the safety and performance of the structure. In addition, a catastrophe may happen when a cable structure collapses due to erosion or fatigue of the cables. Therefore, determination of the cable tension plays a very important role in safety of the cable structures.

Currently two different methods are used to measure the cable tension; static and vibration methods. In static method the tension is measured by using a load cell or a hydraulic jack, while in vibration method natural frequencies are used to measure the cable tension, indirectly (Wu Yu 2006, Zhang Yuxin 2006, and Cheung Y K et al. 2006). The vibration method is widely used in practice because it is simple and fast.

The objective of this paper is to introduce a new technique to find the relationship between the cable tension and the natural frequencies of the cable. This technique resolves the deficiencies of the earlier approaches by taking vibrations of the supports into account. Furthermore, three different cases are considered to discuss the effect of the support vibration on the cable tension.

2. Theory

The fundamental assumptions to develop a mathematical model for dynamic analysis of a cable by considering vibration of the supports are:

1. Density of the cable, ρ, is known.
2. Cable can vibrate only in a vertical plane.
3. Cable is too flexible to carry any bending moment.
4. Direction of the cable tension at any point of the cable is parallel to the tangent line to the cable at that point.
5. Diameter of the cable compared to the length of the cable is very small.

The interaction between the support and the other members is modeled by using equivalent mass and stiffness. Figure 1 shows a simple diagram of the mathematical model.

Because of the vibration of the supports, the vertical displacement, v, at supports is a function of time, t. Suppose the vertical displacements at the supports due to vibration are:

$$v\big|_{x=0} = A\sin w_0 t\,,\quad v\big|_{x=l} = B\sin w_l t$$

where A and B are constants; w_0 and w_l are circular frequencies of the supports at the start ($x=0$) and at the end of the cable ($x=l$), respectively.

Vertical displacement, v, is unknown, and the differential equation for a vibrating string is used to find v as follows:

$$\frac{\partial^2 v}{\partial t^2} = a^2 \frac{\partial^2 v}{\partial x^2} \tag{1}$$

where $a^2=T/\rho$, T is tension and ρ is the density of the cable.

Substitution of boundary conditions gives:

$$v(x,t) = u(x,t) + \frac{x}{l} B\sin w_l t + \frac{l-x}{l} A\sin w_0 t \tag{2}$$

where $u(x,t)$ is the displacement of the vibrating string at location x and time t,

when there is no support vibration.
Substitution of equation (2) in equation (1) gives:

$$\frac{\partial^2 u}{\partial t^2} - a^2 \frac{\partial^2 u}{\partial x^2} = w_l^2 B \frac{x}{l} \sin w_l t + w_0^2 A \frac{l-x}{l} \sin w_0 t$$

(3)

Vibration of the supports has been included in equation (3) and this equation is equivalent to the fixed-ends cable under the effect of the vertical force $\left(w_l^2 B \frac{x}{l} \sin w_l t + w_0^2 A \frac{l-x}{l} \sin w_0 t \right)\rho$. This force also can be considered as the inertia force caused by the vibration of support.

Solving equation (3) yields:

$$u(x,t) = \sum_{n=1}^{\infty} \frac{a_n}{w_n (w_l^2 - w_n^2)} (w_l \sin w_n t - w_n \sin w_l t) \sin \frac{n\pi}{l} x + \sum_{n=1}^{\infty} \frac{b_n}{w_n (w_0^2 - w_n^2)} (w_0 \sin w_n t - w_n \sin w_0 t) \sin \frac{n\pi}{l} x$$

(4)

where $w_n = \frac{an\pi}{l}$, $a_n = \frac{2}{l} \int_0^t \frac{w_l^2 B}{l} x \sin \frac{n\pi x}{l} dx$, $b_n = \frac{2}{l} \int_0^t \frac{w_0^2 A}{l} (l-x) \sin \frac{n\pi x}{l} dx$

Now *v(x,t)* is obtained by substituting equation (4) in equation (2) as follows:

$$v(x,t) = \sum_{n=1}^{\infty} \frac{a_n}{w_n (w_l^2 - w_n^2)} (w_l \sin w_n t - w_n \sin w_l t) \sin \frac{n\pi}{l} x + \sum_{n=1}^{\infty} \frac{b_n}{w_n (w_0^2 - w_n^2)} (w_0 \sin w_n t - w_n \sin w_0 t) \sin \frac{n\pi}{l} x + \frac{x}{l} B \sin w_l t + \frac{l-x}{l} A \sin w_0 t$$

(5)

According to the structure dynamics theory frequencies f_b and f_t are determined by $f_b = \frac{1}{2\pi}\sqrt{\frac{k_b}{m_b}}$, $f_t = \frac{1}{2\pi}\sqrt{\frac{k_t}{m_t}}$. Boundary conditions are also defined by:

$k_b v\big|_{x=0} = T \frac{\partial v}{\partial x}\big|_{x=0} + m_b \frac{\partial^2 v}{\partial x^2}\big|_{x=0}$, $k_t v\big|_{x=l} = T \frac{\partial v}{\partial x}\big|_{x=l} + m_t \frac{\partial^2 v}{\partial x^2}\big|_{x=l}$. To identify equivalent mass (m_b and m_t) and stiffness (k_b *and* k_t), the solution is approximated with the solution of fixed-ends case, which the solution is already available for this case.

The displacement function of the cable, equation (5), can be split into two parts, $v_1(x,t)$ including all the *sin* terms, which the term number, *n*, appears in their arguments, and $v_2(x,t)$ including the rest of the terms:

$$v(x,t) = v_2(x,t) + v_1(x,t) = v_2(x,t) + \sum_{n=1}^{\infty} A_n \sin\frac{n\pi x}{l}$$

According to the principle of stationary potential energy $\delta\Pi=0$, and now energy method can be applied to $v_2(x,t)$ to establish a relationship between the cable tension and frequency.

The stiffness matrix [*K*] and the mass matrix [*M*] are obtained from the matrix form of the "*fit in the right seat*" principle (Guo Xiangrong et al. 2000) as follows:

$$[K] = \begin{bmatrix} \frac{T}{l}+k_b & -\frac{T}{l} & & & \\ -\frac{T}{l} & \frac{T}{l}+k_t & & & \\ & & \frac{\pi^2 T}{2l} & & \\ & & & \frac{4\pi^2 T}{2l} & \\ & & & & \ddots \end{bmatrix}$$

$$[M] = \rho\begin{bmatrix} \frac{l}{3}+\frac{m_b}{\rho} & \frac{l}{6} & \frac{l}{\pi} & \frac{l}{2\pi} & \frac{l}{3\pi} & \cdots \\ \frac{l}{6} & \frac{l}{3}+\frac{m_t}{\rho} & \frac{l}{\pi} & \frac{-l}{2\pi} & \frac{l}{3\pi} & \cdots \\ \frac{l}{\pi} & \frac{l}{\pi} & \frac{l}{2} & & & \\ \frac{l}{2\pi} & \frac{-l}{2\pi} & & \frac{l}{2} & & \\ \frac{l}{3\pi} & \frac{l}{3\pi} & & & \frac{l}{2} & \\ \cdots & \cdots & & & & \ddots \end{bmatrix}$$

Free vibration characteristic equation is stated by:

$$([K]-\omega^2[M])\{\phi\} = \{0\} \tag{6}$$

where ω is circular frequency, and $\{\varphi\}$ is the dominant mode of vibration.
The relationship between frequency and cable tension can be obtained from form equation (6).

3. Effect of the Support Vibration on Cable Tension

Case 1. When the equivalent stiffness values (k_b and k_t) go into infinity, the cable tension, T, is determined by:

$$T = \frac{4\rho l^2 f_n^2}{n^2} \tag{7}$$

The cable tension from equation (7) is the same as the cable tension for the case with no support vibration since the displacements at the ends are negligible, so the change in frequency can be neglected.

Case 2. Suppose that equivalent mass is 0.15 *kg*, and equivalent stiffness values at the both ends vary from 0.45×10^1 *N/m* to 0.45×10^8 *N/m*. Table 1 shows the ratio of the fundamental frequencies determined from equation (6) for this case. In table 1 f_l and f are the fundamental frequencies for the case with support vibration and for the case with no support vibration, respectively.

As it is seen in table 1 the cable tension is much smaller than the actual tension for small equivalent stiffness values. Thus, for small equivalent stiffness values the effect of support vibration must be taken into account otherwise large errors should be expected. On the other hand for the large equivalent stiffness values more than 0.45×10^5 *N/m* the effect of the support vibration can be neglected.

Case 3. Suppose that the equivalent stiffness value is 0.45×10^4 *N/m*, and the equivalent mass values vary from 0.15 *kg* to 0.15×10^3 *kg*. For this case table 2 shows the ratio of the fundamental frequencies determined from equation (6). The definitions of f_l and f in table 2 are the same as the definitions of f_l and f in table 1. Table 2 shows that the frequency ratio (f_l/f) is less than 1 and it decreases as the equivalent mass increases. In this case the cable tension determined from equation (7) is less than the actual value, which means the equivalent mass should not be ignored when the equivalent mass is more than 1.5 *kg*.

4. Conclusion

Based on the structural characteristics of a flexible cable, a model for dynamic analysis of the cable by considering vibration of the supports was developed. The vibration equations of a flexible cable by considering the vibration of the supports are deduced according to the principle of stationary potential energy. In addition, the stiffness and mass matrices are derived by the "*fit in the right seat*" principal. The analysis results for the flexible cable tension show that for small equivalent stiffness values (less than 0.45×10^5 *N/m*) or large equivalent mass (more than 1.5 *kg*) the effect of support vibration on natural frequencies of the cable should be considered to avoid large errors. The results of this study may be extended to solve the similar problems of prestressed flexible cables. The features of developed model in this study show that it is a practical model for dynamic analysis of flexible cables.

References

Cheung Y K, Chan H C, Cai CW(2006). “Natural vibration analysis of rectangular networks”. J. Space Structures, 3 (3), 139-152.

Guo Xiangrong, Chen Huai (2000). “The effect of elastic support on the natural frequency of flexible cable of cable stayed bridge.” J. Zhengzhou Technological University, 21(1), 34-36.

Wu Yue (2006). “Study on wind-induced vibration of membrane-cable structures with the consideration of aeroelastic effect”. J. Progress in Steel Building Structures, 8(2), 30-37.

Zhang Kaiying(2005). “Theoretical and experimental study on detection of cable tension”. J. Progress in Steel Building Structures, 9(12), 25-28.

Zhang Yuxin(2006). “Cable tension detection of beam string structures”. Doctoral Dissertation, Tongji University.

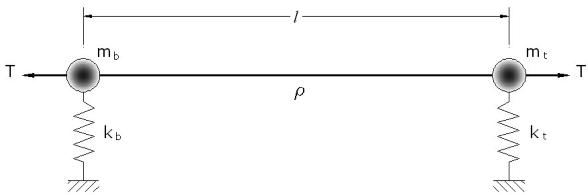


Figure 1. The physical model for cable

Table 1. Fundamental frequencies for the cable

K_t , K_b	0.45e1	0.45e2	0.45e3	0.45e4	0.45e5	0.45e6	0.45e7	0.45e8
f_l/f	0.091	0.279	0.690	0.952	0.997	0.998	0.999	1.000

Table 2. Fundamental frequencies for the cable

m_b	0.15e0	0.15e1	0.15e2	0.15e3
f_l/f	1.000	0.942	0.897	0.742

SUBSURFACE UTILITY ENGINEERING
A Technology-Driven Process that Results in
Increased Safety, Fewer Design Changes, and Lower Costs

By Nicholas M. Zembillas. Senior Vice President/Principal,
Utility Division, TBE Group, Inc. U.S.A
&
Co- President. Suyi Tongda Science and Technology Ltd.

ABSTRACT:
A lack of reliable information on the location of underground utilities can result in costly conflicts, damages, delays, utility service disruptions, redesigns, claims and even injuries and lost lives during construction activities. While the location of subsurface utilities might be found on plans and records, experience has often shown that the utility locations are not exactly as recorded or the records do not fully account for the buried utility systems.

An engineering process known as Subsurface Utility Engineering (SUE) has proven to be a welcome solution to providing this much-needed underground utility information. Combining geophysics, surveying, civil engineering, and nondestructive excavation technologies, SUE can provide accurate mapping of existing underground utilities in three dimensions during the design phase to avoid unnecessary relocations, eliminate unexpected conflicts with utilities, and enhance safety during construction. The use of SUE services has become a routine requirement on highway and bridge design projects, and is strongly advocated by the Federal Highway Administration (FHWA) and state departments of transportation. A savings of \$4.62 for every \$1.00 spent on SUE was quantified by the FHWA from a total of 71 projects. These projects had a combined construction value in excess of \$1 billion. Qualitative savings were non-measurable, but it is clear that those savings are also significant and may be many times more valuable than the quantifiable savings.

In addition to highway designs, SUE is gaining strong endorsement by other industries and market sectors, involved with design of construction projects involving congested underground utilities.

As China's economy grows at a rate leading many other countries, there are more and more utility conflicts occurring daily all over China. Current China best practices provide approximate utility location markings just before construction or full the development of city-wide GIS databases. Both are intended to provide approximate location information and lack the precise and accurate depiction of underground location and characteristic's needed for engineering and design of projects. There is no clear evidence that a defined process such as SUE exists in China for the engineers, designers, contractors, project owners and utility owners. Such situation happened in North America pre 1987 and thus the introduction of SUE. SUE has successfully decreased utility conflicts, saved project costs and improved safety by providing accurate utility information for engineering and design. Now, locating and SUE both exist in North American, locating provide low quality level service for construction needs but SUE provide full range of high quality level engineering

services for engineering needs. As we can see, SUE can benefit the development of China economy.

SUBSURFACE UTILITY ENGINEERING: THE PROCESS

SUE is an engineering process that takes advantage of recent technologies to achieve its objective of accurately identifying, characterizing, and mapping underground utilities. The three major activities of designating, locating, and data management can be conducted individually to meet the specific needs of a given project, but are most advantageously used in combination to create a complete three-dimensional mapping of a utility system.

DESIGNATING:

The process of using a surface geophysical method or methods to interpret the presence of a subsurface utility and to mark its approximate horizontal position (its *designation*) on the ground surface. (Note: Utility owners and contractors sometimes call this process "locating.")

LOCATING:

The process of exposing and recording the precise vertical and horizontal location of a utility, through th use of state of the art vacuum excavation. It is non destructive and extremely time efficient than other conventional digging methods.

DATA MANAGEMENT:

The data management activity involves the incorporation of surveyed utility information obtained by designating and locating into an appropriate system for future use. The range of data management activities includes updating information on existing utility drawings, incorporation and depiction on design plans, and the production of completely new utility maps using computer-aided design and drafting (CADD) programs.

QUALITY LEVELS: American Society of Civil Engineers, Standard Guideline for the Collection and Depiction of Existing Subsurface Utility Data

The question of "how much information is really needed on subsurface utilities" to adequately design and/or construct a project is ultimately a question as to the degree of risk that the project owner, engineer, and contractor are willing to accept. Within the SUE process, this issue is addressed by an American Society of Civil Engineers (ASCE) defined set of four quality levels that represent different combinations of traditional records research, site surveys, and the three primary SUE activities mentioned above. The use of such quality levels allows project owners to certify on project plans that a certain quality level of information has been provided. A large project may include all quality levels of information, in which case the highest level of accuracy may be needed only at those points along a utility's path where conflicts may occur. Lower levels of quality may be adequate in those areas where little to no conflict is anticipated. The four ASCE quality levels are as follows:

QUALITY LEVEL D:

Information derived from existing records or oral recollections.

QUALITY LEVEL C:
Information obtained by surveying and plotting visible aboveground utility features and by using professional judgment in correlating this information to Quality Level D information. An aboveground feature, which is visible at the surface, includes such items as valves, manholes, telephone pedestals, etc.

QUALITY LEVEL B:
Information obtained through the application of appropriate surface geophysical methods to determine the existence and approximate horizontal position of subsurface utilities. Quality Level B data should be reproducible by surface geophysics at any point of their depiction. This information is surveyed to applicable tolerances defined by the project and depicted on plan documents.

QUALITY LEVEL A:
Precise horizontal and vertical location of utilities obtained by the actual exposure (or verification of previously exposed and surveyed utilities) and subsequent measurement of subsurface utilities, usually at a specific point. Minimally intrusive excavation equipment is typically used to minimize the potential for utility damage. A precise horizontal and vertical location, as well as other utility attributes, is shown on plan documents. Accuracy is typically set to 15-mm vertical and to applicable horizontal survey and mapping accuracy as defined or expected by the project owner.

NOT JUST FOR HIGHWAYS:
The term "Subsurface Utility Engineering" was coined about fifteen years ago at the 1989 FHWA National Highway Utility Conference. The FHWA quickly accepted this term, adopted a prime advocacy position, and continues to promote the use of the SUE process in all highway and bridge design projects. Today, through the efforts of the FHWA in documenting and promoting the significant return on any dollars invested in the SUE process, more than 40 State highway agencies are using SUE on their projects.

In addition to highway design, SUE is gaining strong endorsement by the military for use in design of construction projects involving congested underground utilities. Typical applications include construction of underground utilities like sanitary force mains and petroleum oil and lubricant (POL) lines, upgrades to utilities on wharves, airfield runway/taxiway repairs, new site development, and utility system management.

Incorporating the SUE process into the engineering design phase of any construction project that requires excavation around subsurface utilities makes good sense. Typical applications related to military programs include, but are not limited to, the following:

- **Construction or renovation of roadways and runways –** directly correlated with the FHWA's highly successful application of the SUE process at the design phase to avoid costly utility conflicts, damage, delays, service disruptions, redesigns, claims, and injuries.

- **Waterfront construction and repair –** waterfront utilities are typically placed underground in order to maintain overhead clearance for crane operations. As a result, waterfront-berthing areas are very congested with underground utilities such as steam, oily waste, sanitary waste, compressed air, communications, fuel, electrical, and potable water.
- **Repair, Replacement or Removal of Petroleum, Oil and Lubricant (POL) transmission systems –** the need often exists to designate and visually verify the location of existing underground utilities during the replacement or installation of POL lines.
- **Location of underground storage tanks (USTs) –** just as with utility lines, the SUE process is effective in precisely locating underground tanks, even to the point of exactly locating the dimensions of the tank in those sensitive areas where highly controlled excavation is required.
- **Upgrade of water and wastewater treatment facilities –** treatment plants is typically filled with underground piping, compressed air, electrical service and monitoring and control wiring. Upgrades to existing treatment plan facilities frequently involves construction work where existing underground utilities must be identified and protected.
- **Clearance of areas in support of new utility corridors –** whether the new utility systems are replacements for existing underground mains or overall facility upgrades, there is a need to identify existing utilities at conflict points in order to avoid damages and delays.
- **Facility wide updating of "as builts" or other types of utility mapping –** SUE is a viable option for those installations that want to manage their underground infrastructure under the expectation of an extended operational life.

The same technologies used in the SUE process can also find more creative applications in environmental restoration programs. For example, the surface geophysical and nondestructive vacuum excavation technologies represent an alternate means to identify, expose, and possibly characterize buried objects such as underground storage tanks (USTS) or objects in landfills that my be a source of environmental release or explosion if disrupted by more aggressive excavation equipment. In cases where contamination is detected in subsurface holding tanks or catch basins, utility-locating technologies can be used to trace the various pipelines entering the tank/basin back to possible sources of the observed contamination. Vacuum excavation also provides an alternative for the highly controlled removal of soil around buried utilities or other sensitive structures such as security systems.

Whereas cost savings resulting from the avoidance of utility conflicts has been the primary driver for the use of the SUE process in highway projects, many of the aforementioned applications at military installations may find their primary value in reduced construction time and in increased safety of workers and the public. Recognizing the level of importance placed on personnel safety and the degree of upfront planning and review typically carried out to assure a safe working environment, it would be of great advantage to know with certainty the location of potentially impacted utilities at the planning stage, prior to contract bidding and field mobilization.

For maximum effectiveness, the SUE process should be incorporated early in the development of every project that may have an impact on underground utility facilities, particularly in built-up areas. When subsurface utilities are discovered during the construction phase, the costs of conflict resolution and the potential for catastrophic damages are at their highest. That is why the collection and systematic depiction of reliable data for existing subsurface utilities is critical if engineers are to make informed decisions and support risk management protocols regarding a project's impact on these utilities.

ALLOCATION OF RISK:

The fact that readily available information on utility locations is often incomplete and inaccurate has been a "given" among project owners, engineers, and contractors for decades. One can easily relate to the following scenario that has been played out on many military construction projects in response to this informational uncertainty and the associated risks:

- The project owner would initially assign responsibility for utility mapping to the design engineer by including general language in the scope of work. Typically, the standard of practice for the design engineer would be to compile utility information from the utility owner, local public works files, oral recollections of "old timers", and other readily available sources of such data, and then to correlate this information with the site survey for the project. This information corresponds to ASCE Quality Level "D" and "C".
- Recognizing the likelihood that utilities would either be missed or erroneously located through this process, the design engineer would include a disclaimer on the plans. In essence, the disclaimer would pass the responsibility for underground utilities on to the construction contractor. Either the project owner or the contractor would rarely challenge such disclaimers, and thus the design engineer would incur minimal risk for any errors related to utility depictions on the design plans.
- Facing a bid competition, and recognizing that project owners are typically held responsible for delays and costs associated with "unknown" or "differing site conditions," contractors would have no option other than to assume that the utility depictions on the plans were complete and accurate.
- At the start of construction, the "one-call" statutes would often result in utility owners coming into the field and marking expected utility locations during construction. These locations may or may not have coincided with the design plans.
- As a result, the project owner would often face a situation in which: (1) the initial construction bids would likely have a contingency built in to account for the informational uncertainty surrounding underground utilities and the likelihood that claims would have to be developed and negotiated; (2) the contractor would secure a preferable position since any additional work would be negotiated and performed outside of a bid situation; (3) the financial risk

would be magnified by the potential costs of construction downtime, schedule delays, redesign, and utility relocation; and (4) a worst-case scenario could develop that may involve utility damage, service outages, consequential damages, and injury to works and the public.

Over the last fifteen years, four events began to change this scenario. First, a convergence of new equipment and data-processing technologies now allows for the cost-effective collection, depiction, and management of existing utility information. These technologies encompass surface geophysics, surveying techniques, computer-aided design and drafting and geographic information systems, and minimally intrusive excavation techniques. Second, competition in the marketplace allowed project owners to shift more responsibility and liability to the design engineers and contractors. Third, there has been an overall growth in the litigious nature of society and the associated increase in risk to all parties involved in design and construction projects. And fourth, the birth of Subsurface Utility Engineering as a distinct service occurred in the 1980's. The subsurface utility engineer became the individual with the appropriate expertise and tools to allow construction activities to be planned away from high risk utilities when possible, to characterize the nature of utility conflicts before work begins in the field, to coordinate utility relocations and easements, and to develop utility "as builts" that are complete and accurate.

SUMMARY:

The technology is now available to achieve a complete and precise three-dimensional mapping of subsurface utilities prior to or at the design phase of a project. If Quality Level A information is collected through the full use of the SUE process, the project owner, design engineer, and contractor can proceed with confidence that utilities have been identified and categorized as to their horizontal location, depth, size, composition, and condition. The use of the SUE process will continue to grow as project owners request higher quality levels of utility information to reduce and better manage their risks. The ASCE has recently published related standards that will better define the allocation of risk among the parties. In the end, the use of Subsurface Utility Engineering will tend to shift the risk of bad information to the party most capable of handling that risk – the subsurface utility engineer.

Related cost analyses have shown that the use of the SUE process provides a significant return on investment. As a very rough rule of thumb, the cost of incorporating the SUE process is about 10 percent of the total preliminary engineering cost, or about 1 percent of the total project cost. These costs are small when compared to the overall savings on projects where the SUE process is used. The FHWA has recently documented a savings for \$4.62 for every \$1 invested in SUE. The return is higher in areas of densely congested underground utilities.

In this era of partnerships, SUE represents a new way of doing business, in which the past adversarial relationship among project owners, design engineers, utility companies and contractors has been replaced by a cooperative effort to reach an appropriate balance between the risk of informational uncertainty and the associated cost of reducing that risk.

Heat Transfer Model and Numerical Simulation for Microwave Hot In-Place Recycling of Asphalt Pavements

Song-qing Zhu[1], Jin-fei Shi[1], , and Tong-sheng Sun[1] edited by Thomas J. McGean[2]
[1]College of Mechanical Engineering, Southeast University, Nanjing, China; email:zsqztclw@hotmail.com [2] Melfa, VA 23410, USA; email:tjmpe@verizon.net

Abstract: To assess the temperature field when microwave heating is used to recycle asphalt mixtures, a two-dimensional mathematic model considering the heat absorbed and conducted by microwave internal heat-generation was developed based on a Fourier heat transfer model. The microwave internal heat-generation was modeled by approximating the radial field by a two dimensional planar surface with uniform dielectric properties throughout. The control volume based finite difference method (CV-BDM) was used to establish a discrete form of the energy conservation equations, and a numerical simulation was conducted. Results showed that non-linear heat transfer. For confirmation, an experiment was conducted using a microwave power system at 2.45GHz and recording the temperature variation as a function of surface location and heating time within samples of asphalt mixture. The data confirmed that this microwave heating method could both heat the mixture in a reasonable time and improve temperature uniformity.

Introduction

Asphalt has been widely used in constructing high-quality roads, bridges and airport pavements. However, due to heavy traffic loads and environmental factors, functional damage to the surface of asphalt pavements such as formation of cracks and potholes is inevitable over time. This requires prompt repair and maintenance. Microwave recycling, which has the advantages of deep penetration, small thermal inertia and ease of control (Zhu et al., 2006), has currently become an important hot in-place recycling (HIPR) technology, and has attracted more and more attention.

Recent research on microwave recycling has included experiments on ways to repair asphalt pavements, the sensitivity of various parameters on the microwave recycling process, and measurement of the dielectric properties of asphalt mixtures. (Bosisio,etal.,1974;Al-Ohaly,et al.,1988;Shoenberger,et al.,1995). In addition, further research has identified several factors critical to the success of in-place microwave heating when used to repair pavement. These key factors include post-heating temperature distribution within the asphalt mixture and the effective control of heating temperature and heating depth. Hopstock(2003)and Zanko et al.(2004) from Minnesota performed the first theoretical analysis of microwave heating of asphalt mixtures. However, in their studies only one-dimensional heat transfer was taken into account. The microwave radiator in HIPR is of the horn type with a pyramidal shaped radiating cone. It was noted in the experiments that the uneven distribution of electrical field intensity on the surface of roads resulted in a corresponding uneven distribution of temperature on the asphalt mixture surface. This led to the formation of hot-spots and cold-spots on the asphalt surface. In other words, the temperature of the mixture in hot-spots has met the recycling criteria but in cold-spots the mixture has not completely melted, leading to decreased recycling efficacy and repair quality. Apparently, one-dimensional analysis has significant limitations.

In our analysis we have considered two dimensional heat conduction using microwave as an intrinsic heat source which is a closer approximation to the actual conditions found in microwave heat recycling. Theoretically, the physical parameters of all materials may change as a function of temperature. Here we have ignored any temperature-induced changes in the physical parameters of asphalt mixtures in order to simplify the analysis.

1. Heat transfer model within microwave recycling

The microwave radiating heater is positioned above the asphalt pavement as shown in Fig.1. $D_1 \times D_2$is the surface size of the horn, h is the height of the horn above the pavement surface, and H is the thickness of the pavement. A coordinate system is established as shown in Fig.2 with the z-axis in the direction of microwave propagation and the *x-y* plane on the pavement surface. Microwave energy, emitted by a magnetron-tube antenna and focused by the horn cavity, heats the pavement thereby hot-recycling the asphalt.

The assumptions used to construct the mathematic model are as follows:1)For propagation in the *z*-direction, the thickness of the recycling layer is small compared with the microwave attenuation depth within asphalt mixtures, so a uniform temperature distribution can be assumed;2)Radiation and convection heat exchange can be ignored;3)The microwave energy is completely absorbed by the asphalt mixture with no loss due to energy reflected off the pavement surface.;4)The horn surface is adiabatic; that is no heat is lost through the horn to the atmosphere.

1.1 .Governing equation of heat transfer

According to the above assumptions, a two dimensional Fourier model with unsteady state temperature distribution within the asphalt-mixture was constructed. Thermal properties such as thermal conductivity and specific heat were assumed constant throughout the pavement. The energy equation within the asphalt mixture samples can be written as

$$k_{eff}\left[\frac{\partial^2 T(x,y,t)}{\partial x^2}+\frac{\partial^2 T(x,y,t)}{\partial y^2}\right]+\phi(x,y)=\rho C_p\frac{\partial T(x,y,t)}{\partial t} \tag{1}$$

where T is the temperature(°*C*), ϕ is the strength of the internal heat source(W/m^3), t is the time in seconds, k_{eff}, ρ and C_p are the thermal conductivity($W/m/°C$), the material density(kg/m^3) and the specific of asphalt mixtures ($J/kg/°C$), respectively. As noted, all have been assumed to be constants in the current analysis.

1.2. Microwave internal heat-generation

An asphalt mixture is a lossy medium for microwave energy and this energy dissipation generates internal heat which increases the pavement temperature. The sinusoidal time variation of the microwave field is negligible when compared with the temperature variations within the asphalt mixture, since the microwave frequencies are very high (2.45GHz). The root mean square (RMS) dissipation power in an arbitrarily small volume of asphalt mixture can be expressed as

$$P_i = 0.5561\times10^{-6} f\varepsilon_r^{'} \tan\delta \cdot \widetilde{E}_i^2 V_i = Kf\varepsilon_r^{'} \tan\delta \cdot \widetilde{E}_i^2 V_i \tag{2}$$

where P_i is the dissipation power in ith mixture sample volume(W), f is the microwave frequency(Hz), $\varepsilon_r^{'}$ is the relative dielectric permittivity(F/m), $\tan\delta$ is the dielectric loss tangent, $\widetilde{E}_i$ and V_i are the averaged value of instantaneous electric field amplitude (V/m) and volume(m^3)of the ith sample, respectively.

As the ith sample is a micro-unit with $dxdydz$ magnitudes, the strength of the intrinsic heat source can be written as

$$\phi_i(x,y) = P_i / V_i = Kf\varepsilon_r^{'} \tan\delta \cdot \widetilde{E}_i^2 \tag{3}$$

In reality, when radiation heating is processed using a pyramidal horn antenna, the average distance between the opening of the horn and the surface of the asphalt pavement h<10(λ/2π), and so can be considered as near field radiation. So the surface field can be approximately considered as radiation field, when h=0 the surface field is named radiation especially. A surface coordinate system can be established as shown in Fig.2, where the center of the surface is the coordinate origin, the electric field is in the x-axis, the magnetic field is in the y-axis and the propagation direction is in the z-axis. The center of the pyramidal horn surface is set as phase zero. The surface field can be expressed as

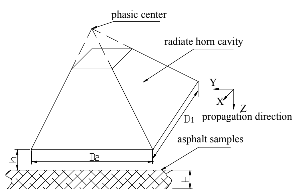


Fig. 1 Schematic of microwave heating recycling

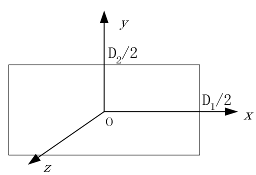


Fig.2 Coordinate of pyramidal horn surface

$$\widetilde{E}_i = E_0 \cos(\frac{x(i)}{D_1}\pi) e^{-j\frac{\pi}{\lambda}(\frac{x(i)^2}{R_1}+\frac{y(j)^2}{R_2})} \tag{4}$$

In the equation, R_1 is the horn length in the *H*-field, R_2 is the horn length in the *E*-field, and $x(i)$ and $y(j)$ are x-direction and y-direction node positions respectively.

1.3. Initial and boundary conditions

The initial condition of the asphalt samples can be specified as

$$T = T(x, y, t) = T_0 \quad \text{at} \quad t = 0 \tag{5}$$

As a result of symmetry, shown as Fig.2, the thermal boundary conditions of the asphalt sample walls may be approximately assumed as

$$T = T_\infty \quad \text{at} \quad x = D_1 / 2\,; \quad \partial T / \partial x = 0 \quad \text{at} \quad x = 0 \tag{6}$$

$$T = T_\infty \quad \text{at} \quad y = D_2 / 2\,; \quad \partial T / \partial y = 0 \quad \text{at} \quad y = 0 \tag{7}$$

where T_0 and T_∞ are the surroundings and initial temperatures respectively, and $T_0 = T_\infty$.

2. Numerical simulation

2.1. Numerical method and mesh generation

In view of the symmetry of the four quadrants as shown in Fig.3, a mesh may be established using one-fourth of the pavement surface with the $D_1 \times D_2$ dimensions taken as $150 \times 120mm$. A uniform mesh generation with $\Delta x = \Delta y = 5mm$ requires 15 nodes in the *x*-direction and 12 nodes in the *y*-direction. Let i *and* j be the x and y labels for each node, and $x(i)$ and $y(j)$ be the *x* and *y* coordinates for each node (see (4)) respectively. Heat flux within the control volume is assumed uniform. For the 2D model, taking the thickness of 1 in *z*-direction, $\Delta x \Delta y$ is the volume of the control volume. The control volume based finite difference method (CV-BDM) was used to establish the discrete scheme of the conservation equations. The discrete 2D equation is specified as

$$a_P \theta_P = a_E \theta_E + a_W \theta_W + a_N \theta_N + a_S \theta_S + b \tag{8}$$

2.2. The calculation results and discussions

Original calculation parameters used were as follows: An AC-13I asphalt sample was selected with asphalt aggregate ratio of 5.3%, $\rho = 2435kg/m^3$, $C_p = 0.8896J/g/°C$, $\kappa_{eff} = 1.560W/m/°C$. Other parameters were $T_0 = T_\infty = 18°C$, $\varepsilon_r' = 5.8$ and $\tan\delta = 0.0344$. A 2M210F magnetron-tube with a working frequency of 2.455GHz and effective output power of 650W was used. Therefore $E_0 = \sqrt{480\pi P' / (D_1 D_2)} = 7.38 \times 10^3\, Vm^{-1}$ (Gu, &Shen ,1980).

The temperature distribution fields within the sample are shown in Fig.4 after heating for 600 seconds and 900 seconds. .The highest temperature is at the coordinate origin, namely the sample's center. The temperature gradient along the *x*-direction is smaller than in the *y*-direction and the temperature is lower at the edge of the sample, so the heat loss rate must be decreased in order to improve the uniformity of temperature distribution. The increase of temperature is obviously nonlinear; it rises slowly from $18°C$ to $86.2°C$ in the first 10 minutes but then it increases rapidly to $143.4°C$ within the next five minutes. In other words the higher the temperature, the faster the rate at which the temperature increases. Therefore, the heating time must be carefully monitored to avoid pitch coking.

3. **Experiment system of microwave recycling**

3.1. Experimental device

The experimental device is shown in Fig.5. For the safety of operators during the experiment, a metal web hood was made to shield the radiation. A grid template was made for measuring the temperature of the sample. The surface of the sample was divided into an 8×10 grid.

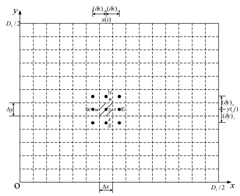


Fig.3 Mesh grid for 2D CV-BDM

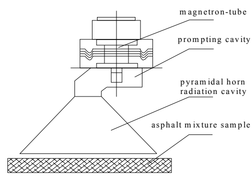


Fig.5 Microwave heating experiment device

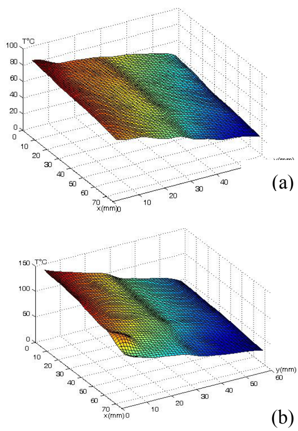


Fig.4 Simulation temperature distributions within the samples: (a) t=600s; (b) t=900s

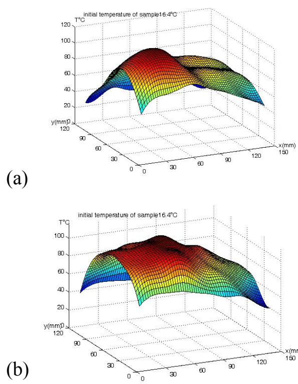


Fig.6 Experimental temperature distributions within samples: (a) continuous heating; (b) intermittently heating

3.2. Results and discussion

The temperature was obtained using an infrared thermometer while heating the sample continuously for 15 minutes and also while intermittently heating the sample by heating for five minutes at one minute intervals, and the recorded temperature fields were plotted as shown as Fig.6. The results indicate that the highest temperature is near the center of the surface. The highest temperature was about 15℃ lower than the numerical simulation due to reflection of energy off the pavement surface and energy lost from the horn surface to the atmosphere. The results show that heating intermittently produces a more uniform temperature distribution than heating continuously. In addition, the temperature distribution agrees reasonably well with the numerical simulation. Therefore, it is expected that use of this analytical model can help to improve HIPR technology.

4. Conclusions

Based on the Fourier equations for heat transfer, a two-dimensional heat transfer model was built and microwave internal heat-generation was studied according to the law of near-field radiation. The results, in good agreement with the numerical simulation, indicate that pavement quality can be improved by ameliorating heating techniques. These methods provide a practical guide for microwave hot-in place recycling of asphalt pavements.

References

Al-Ohaly A. A., and Terrel R. L. (1988). “Effect of microwave heating on adhesion and moisture damage of asphalt mixtures.” Transportation Research Record, 1171, 27-36.

Bosisio R. G., Spooner J., and Granger J. (1974). “Asphalt road maintenance with a mobile microwave power unit.” J. Microwave Power, 9 (4) 381-386.

Hopstock D. M. (2003). “Microwave-absorbing road construction and repair material” Final Report to NRRI on Idea Evaluation Subcontract , Minnesota.

Ruilong Gu, Minyi Shen.(1980). “Microwave technology and antenna.” Defense Industry Press, Beijing. (in Chinese)

Shoenberger J.E., Rollings R.S., and Graham R. T. (1995). “Properties of microwave recycled asphalt cement binders.” *Proceedings of the Conference on Physical Properties of Asphalt Cement Binders*. Dallas, TX, ASTM Special Technical Publication, 1241,199-213.

Zanko L. M., and Hopstock D. M. (2004). “Minnesota taconite as a microwave-absorbing road aggregate material for deicing and pothole patching Applications.” *Minnesota Department of Transportation Research Services Section*, 3, 1-16.

Zhu Songqing, Shi Junfei, and Wang Hongxiang (2006). “Modeling and experiment of microwave heating for hot in-place recycling of asphalt pavements.” J. Southeast University (Natural Science Edition), 36(3), 393-396 (in Chinese)

Study on the Modification Mechanism of Asphalt with Waste Packaging Polymer

Changqing Fang[1], Tiehu Li[2], Shisheng Zhou[3], Xin Wang [4], Lance Heiliger[5]

[1]Institute of Printing and Packaging Engineering, Xi'an University of Technology, Xi'an 710048, China; email:fcqxaut@163.com;
Department of Materials Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China; email:fcqxaut@163.com
[2]Department of Materials Science and Engineering, Northwestern Polytechnical University, Xi'an 710072, China; email:litiehu@nwpu.edu.cn
[3]Institute of Printing and Packaging Engineering, Xi'an University of Technology, Xi'an 710048, China; email:zhoushisheng@xaut.edu.cn;
[4]Institute of Printing and Packaging Engineering, Xi'an University of Technology, Xi'an 710048, China; email:wangxin0207@163.com;
[5]Member ASCE, Fayetteville, Arkansas

Abstract: Instead of traditional polymer modifiers in asphalt, this experiment tested the addition of the common plastic bag (primarily LLDPE). Shredded plastic bags were added by percent weight for samples of 2%, 4%, and 6%. Softening point, penetration, ductility, low temperature-cracking performance and the micro-structure of the modified asphalt were studied. Analysis of the mixing process and structure of the modified asphalt indicated that swelling and molecular level crack-branching were the main reasons for the modified effect.

Key words: Waste-PE in packing (WPE); modified asphalt binder; swelling; Shear yield; Crack hinging

1. Introduction

"White pollution", the ubiquitous plastic packaging in the modern society, is now discarded without thought or care. The end result is a serious environmental problem, and a huge waste of natural resources (Shan Shan Chung and Chi Sun Poon 1996; Azni Idris et al. 2004; Paul S et al. 1999).

Asphalt is a commonly used binder material. When mixed with aggregate it constructs a strong and stable mixture used for constructing durable and high quality roadways. However, ordinary asphalt binder can no longer be used in modern construction due to increases in traffic volume, speed, and oversize loads. These conditions exacerbate the asphalt binder shortcomings such as high paraffin content, poor adhesion, low ductility, high temperature sensitivity.

Improving the performance of asphalt has become an urgent need of modern highway construction (Akmal N and Usmani AM 2001; Lewandowski LH 1994; Fawcett AH et al. 1999). One solution is the modification of the asphalt binder. In this paper, performance enhancements from recycled waste polymer are studied, and the modification mechanism is analyzed.

2. Experiment Details

2.1 Raw materials

In this experiment a common asphalt binder (X-110）from Xi'an Petrochemical Plant was selected. New binder was used in this study, and the virgin asphalt contained no impurities. Typical characteristics are shown in Table 1. Recycled plastic bags (main ingredient LLDPE) were used in this experiment. Those characteristics are depicted in Table 2. To ensure purity the recycled plastic was washed and dried at 60°C until the water content was less than 1% by weight.

2.2 Asphalt modification procedure

First the asphalt was melted at 180°C. In order to mix the WPE with asphalt, the cleaned plastic was shredded to approximately 1.5×2.5 cm^2. Shredded plastic bags were added by percent weight for samples of 2%, 4%, and 6%, and stirred for 30 min at 180°C. The binder was sheared at this temperature for an additional hour. After that the sample was kept at a temperature about 120°C for 30 minutes while it increased in volume. Finally the modified binder was subjecting to a high-speed shear for 1 hour. At the end of the process, the WPE was uniformly dispersed in the asphalt.

2.3 Testing and instrumentation

The consistency of the modified asphalt was measured with the EL46-5380/01 penetrometer manufactured by the YUEHUA Company (China). The softening point was assessed by using an RKA 2 automatic instrument (Germany). Ductility was tested using the ELE46-2615/01 manufactured by the YUEHUA Company (China). The microstructures of the modified asphalt were observed by The OLYMPUS Company's CX40-RFL32J fluorescent microscope. Fourier Transform Infrared Spectroscopy (FTIR) utilized a spectrometer manufactured by the Equinox Companies, type 55. Low-temperature cracking properties were tested using the LDDS-04 developed by the Xi'an Institute of Leak Proofing.

3. Results and discussion

3.1 Test Results

Figure 1 displays the FTIR curves of the original and modified asphalt binder. It shows that the original asphalt (0wt% of WPE) has obvious absorption bands at 2925 cm^{-1}, 2856 cm^{-1} and 1460 cm^{-1}, which are attributed to the vibration of C-H bonds of methyl and the methylene groups. The position of the main absorption band of the modified asphalts (2%, 4% and 6%) is the same as the original asphalt with the exception that the peak at 2925 cm^{-1} and 2856 cm^{-1} increases slightly with the increase of WPE. Such a slight increase is due to the interaction of the methyl and methylene groups with the asphalt binder, but suggests that the combination is primarily due to physical forces and not a chemical process.

Figure 2 shows the softening point curve of the original and modified samples. Figure 3 and Figure 4 are the ductility and penetration results, respectively. Note that as the WPE content increases, the softening point becomes higher. Compare 6wt% of WPE enhances the asphalt's softening point to 76°C from the original 47°C. With higher percentages of WPE, the ductility and penetration decreased significantly. In short, WPE packaging waste can improve the high-temperature performance and viscoelasticity of asphalt.

3.2 Theory

The improvement of WPE modified asphalt performance can be explained as follows. First, WPE has a wide viscoelastic domain of -80°C~110°C (asphalt is about -15~55°C). Second, flexible linear macromolecules in a long chain have a larger capacity to adapt to external forces, characterized by good flexibility, elongation, impact resistance properties, and a large molecular weight (up to 300,000). WPE possesses a multi-branching tree structure consisting of long chain linear molecules with alkyl and methyl branched-chains. The long high speed shear mixing process is necessary to uniformly distribute the packaging waste polymer in the asphalt binder. The mixture of WPE and asphalt is a thermodynamically incompatible mixture. WPE and asphalt will not disperse at a molecular-level, so there will be clear delineations between the materials. However, when heated, asphalt becomes a complex solvent consisting of a large number of aromatic solvents, polycyclic aromatic hydrocarbon molecules and wax molecules. These individual components interact with WPE molecules making them swell and link together, forming a network structure (Figure 5). The expanded and linked network absorbs more energy (Lee N 1995; Sulaiman, SJ. and Stock AF. 1996). As a result of the WPE characteristics and individual reactions with asphalt, the viscoelasticity and softening point increase and ductility is decreased.

3.3 Shear yield mechanism

Figure 6 shows the results of the low-temperature cracking experiment. It can be seen that the original cracking temperature is -25°C. After modification with 6% by weight of WPE, it has been decreased by 10oC to -35°C and yield stress increased by 75% from 3.2MPa to 5.6MPa. These results show that the modified asphalt possesses better low temperature performance.

It can be seen from Figure 2 that the penetration of modified asphalt is lower than that of the original asphalt. The reason is that stress is concentrated on internal or external defects triggering a minor cavity. As the stress increases the oriented molecular chains break, leading to micro cracks. With the introduction of small particles of WPE the cracking pattern is altered. When this happens, the energy will be redirected, resulting in greatly increased viscoelastic and plastic characteristics and delaying failure.

3.4 "Crack hinging"

Raw asphalt at low temperatures is known for strong molecular interaction and brittleness, so low temperatures need less energy to start cracking. Figure 6 indicates that the low-temperature cracking resistance is improved after binder modification. That low temperature cracking of PE modified asphalt requires higher energy cannot be explained with traditional cracking theories. Instead it can be explained by a “crack hinging”. The crack will take the path of least resistance, so it avoids the polymer molecules associated with the WPE. A longer cracking path “hinging” around the polymer molecules absorbs more energy. (Figure 7 is a sketch of the “crack hinging”) Hence, the low temperature cracking performance is enhanced through the WPE modification.

4. Conclusions

Modifying raw asphalt with WPE improve the overall performance of an asphalt binder. Viscosity and the high temperature performance are improved through the swelling WPE molecules. The stress-strain test showed that the modified binder has improved viscoelasticity and plasticity. This is particularly shown by the improved resistance to low temperature cracking. An explanation for the improved low temperature cracking resistance is explained by “crack hinging”. In conclusion, the addition of WPE enhances asphalt binder, and it helps eliminate “white pollution”.

Acknowledgements

The authors acknowledge the financial support provided by Shaanxi Programs for Science and Technology Development (Fund No.2006K06-G14) and the Xi’an Programs for Industry (Fund No. GG06070).

Reference

Akmal N, Usmani AM.(2001). “Asphalt, Number One Thermoplastic Polymer”, Polymer News, 24:136-40.

Azni Idris, Bulent Inane, Mohd Nassir Hassan. (2004). “Overview of waste disposal and landfills/dumps in Asian countries”, Material Cycle and Waste Management in Asia, (6):104-110.

Fawcett AH, McNally T, McNally GM. (1999). “Modification of a Bitumen with Various Polymers for Use in Built-Up Roofing Membranes”, Polymer, 40:6337-6349.

Lee N. (1995). “Low Temperature Nature of PE-Modifier Binders and Asphalt Concrete Mix”, Asphalt Paving Technology, 64:207-238.

Lewandowski LH.(1994). “Polymer Modification of Asphalt Binders”,Rubber Chemistry Technology, 67(3): 477-480.

Paul S. Phillips, Adam D. Read, Anna E. Green, et al. (1999). “UR waste minimization clubs: a contribution to sustainable waste management”, Resources, Conservation and Recycling,(27):217-247.

Sulaiman, SJ, Stock AF. (1996). “Use of Fracture Mechanics for the Evaluation of Asphalt Mixes”,Asphalt Paving Technology, 65:500-533.

Shan Shan Chung and Chi Sun Poon. (1996). “The Attitudinal Differences in Source Separation and Waste Reduction between the General Public and the

Housewives in Hong Kong", Journal of Environmental Management, (48):215-227.

Table1 Characteristics of raw asphalt

virgin asphalt	Penetration (25°C,100g)/0.1 mm	Ductility (25□)/cm	Softening point(°C)	Solubility %
X-110#	42	112	54	99.7

Table 2 Recycled WPE characteristics

Main component	Purity	Melt Index	Softening Temperature (°C)	Density (g/cm^{-3})	Thickness (mm)
LLDPE	92%	7.4	125	0.93	0.10

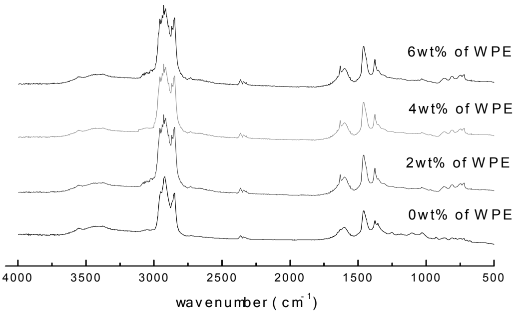


Figure 1 - FT-IR spectra of raw asphalt and WPE modified asphalts

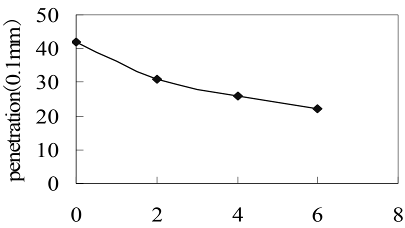


Figure 2 - Asphalt Penetration

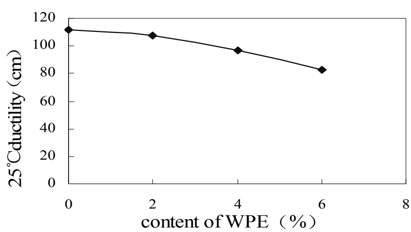


Figure 3 – Asphalt Ductility

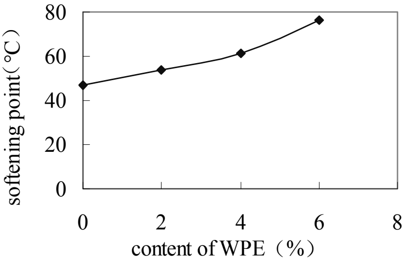
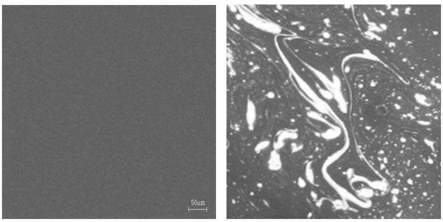


Figure 4 – Softening Point



A B

Fig.5 – Sample Micrographs (A: 0wt% of PE, B: 6wt% of WPE)

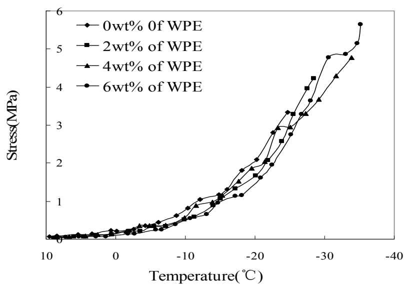


Fig.6 – Stress-Temperature relationship

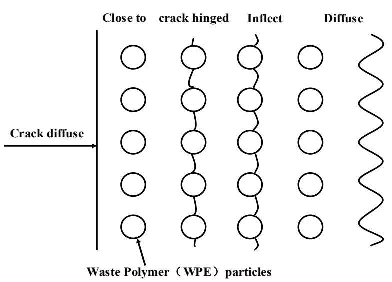


Figure 7 – Sketch of crack hinging action

Dense-graded Asphalt Treated Base Design Method Investigation

Qian Hua

BeiJing Road Engineering Quality Monitoring Station, BeiJing, 100076,

Abstract: Asphalt treated base (ATB) materials have shown good road performance. There are some differences in the mix design methods used for ATB and Asphalt Concrete (AC) and this paper investigates four different design methods for ATB. Each design method is analyzed by using volume and mechanical indexes, and pavement performance. Based on this analysis, the proper design method of ATB is proposed.

1. ATB Design Method

This paper describes the comparison of experimental data with specification criteria recommend in the SUPERPAVE method, the Bailey method, and a mechanical method. The analysis includes determining the effect of different design methods on asphalt pavement performance.

1.1 Test Results

1.1.1 Graduation Design Results

The materials used for the evaluation were ESSO 70# asphalt bitumen, coarse aggregate, fine aggregate, and a filler of Anshan stone from Shan xi. The gradation of the ATB aggregates are provided in Tables 1-1, 1-2, and 1-3. The nominal maximum aggregate sizes for the different design methods were 26.5 mm, 31.5 mm, and 37.5 mm. The aggregate gradation curve diagram is not included in this paper.

1.1.2 ATB Volume Index Comparison of Different Design Methods

The volume index of the ATB are provided in Tables 1-4, 1-5, and 1-6 for the nominal maximum aggregate sizes of 26.5 mm, 31.5 mm, 37.5 mm and the different design methods. The number in table1-4 correspondence to the nominal maximum aggregate sizes of 26.5 mm, 31.5 mm, 37.5 mm and the different design methods.

1.2 Currently Recommended Specification Method

The Technical Specifications for the Construction of Highway Asphalt Pavements (JTG F40-2004) requires that the Marshall Mix design method be used for asphalt pavements. Traditionally, using this volume index method, the materials are selected, the aggregate gradation is defined, and the optimum asphalt cement content is determined. Compared with the old specification, it considered the effect that aggregate absorbs some asphalt. The asphalt content was determined by including two items: aggregate absorption and effective asphalt. It used the effective specific gravity to calculate the volume index.

The average value curve of the ATB-40, ATB-30, and ATB-25 was selected for the gradations used to for the evaluations.

Table1-1. Mixture Gradation of 26.5mm Nominal Maximum Aggregate Size

gradation	Mass passed percent (%)													
	37.	31.	26.	19	16	13.	9.5	4.7	2.3	1.1	0.6	0.	0.1	0.07
ATB-2	100	100	95	70	58	52	42	30	23.	17.	13	9.	6.5	4
SUP-25	100	100	89.	81.	74.	63.	51.	35	23	15.	11.	7.	5.7	3.5
BLF-25	100	100	93	87	78.	62.	46.	33	23.	15.	11.	7.	5.4	3.9

Table1-2. Mixture Gradation of 31.5mm Nominal Maximum Aggregate Size

gradation	Mass passed percent (%)													
	37.5	31.5	26.5	19	16	13.2	9.5	4.75	2.36	1.18	0.6	0.3	0.15	0.075
ATB-30	100	95	80	62.5	55	49.5	41	30	23.5	17.5	13	9.5	6.5	4
SUP-30	100	95.9	84.1	71	65.4	56.2	47.1	32.5	21.5	14.6	10.3	6.9	5	3
BLF-30	98.9	95.5	85.7	74.3	65.7	50.2	35	27.2	18	12.5	9.2	6.6	5.1	3.8

Table1-3. Mixture Gradation of 37.5mm Nominal Maximum Aggregate Size

graduation	Mass passed percent (%)													
	37.5	31.5	26.5	19	16	13.2	9.5	4.75	2.36	1.18	0.6	0.3	0.15	0.075
ATB-40	95	83.5	75	60	53	47	40	30	23.5	17.5	13	9.5	6.5	4
SUP-40	100	100	93.4	79	63.2	57.7	49.4	41.4	29	19.5	13.4	9.6	6.6	4.8
BLF-40	98.4	93.8	80.2	65.1	57.7	45.3	33.4	26.4	17.3	12.1	9	6.5	5.1	3.8

1.3 SUPERPAVE Method

The design equivalent single axle load (ESAL) for the design lane traffic was assumed to be 1.0×10^7 ESAL to 3.0×10^7 ESALs with a minimum wearing surface of 100 mm, which means the ATB lies under the road surface a minimum of 100 mm. After calculating the upper and lower gradation limits, the mixture gradation for SUP-25, SUP-30, and SUP-40 was determined and is presented above. The performance grade (PG) for the bitumen was PG 64-16 and based on traffic level, the gyratory compaction levels were selected. The initial number of gyrations was 8, the design number of compaction gyrations was 100, and the maximum compaction number of gyrations was 160. The optimal asphalt content was selected based on the traffic level and nominal maximum size.

Table1-4. ATB Volume Index of 26.5/31.5/37.5mm Nominal Maximum Aggregate Size

Test index	ATB-25	SUP-25		BLF-25	
	Marshall method	SUPERPAVE method	Marshall method	mechanical method	Marshall method
Asphalt aggregate ratio (%)	3.6/3.3/3.0	3.4/3.2/2.9	3.7/3.4/3.1	3.6/3.3/3.0	3.5/3.3/3.1
Saturated surface-dry bulk specific gravity	2.456/2.455/2.468	2.461/2.462/2.475	2.460/2.448/2.470	2.465/2.461/2.471	2.461/2.470/2.476
VV (%)	4.7/5.0/5.1	4.0/4.0/4.0	4.1/5.2/4.9	3.9/4.7/4.6	4.7/4.6/4.7
VMA (%)	12.1/11.9/11.2	12.1/11.8/11.6	12.0/11.9/11.0	11.6/11.9/10.9	12.1/11.7/11.1
VFA (%)	61.6/57.9/54.6	67.3/66.6/61.8	64.9/56.8/55.0	63.2/55.6/70.3	59.9/58.4/56.3
Stability (KN)	23.7/24.4/27.4	-	21.3/20.2/20.6	-	21.2/22.3/20.9
Flow value (0.1mm)	3.5/3.4/4.3	-	3.9/5.0/5.4	-	3.5/3.8/4.5
Ratio of filler to bitumen	1.18/1.30/1.44	1.12/1.03/1.16	1.02/0.96/1.07	1.16/1.29/1.34	1.20/1.29/1.39
DA (μm)	8.29/7.50/6.75	8.96/9.21/8.28	9.89/9.89/8.96	8.98/8.57/8.32	8.69/8.57/8.01

1.4 Bailey Method

Two different coarse aggregates with sizes of 20 - 40 mm and 10 - 20mm; three fine aggregate with sizes of 5 – 10 mm, 3 – 5 mm, and 0 – 3 mm; and a filler material obtained from lime stone were selected for the Bailey Method analysis. The gradations for BLF-35, BLF-30, and BLF-40 are presented in 1-1 – 1-3.

To ensure good aggregate to aggregate interlock in the mix, we select suit aggregate density , The loose density is the low limit of aggregate interlock within the mix in theoretical , it is the cut-off point of coarse grading and fine grading ; dry-tamping density is the upper limit , it will be over 1.1 times the loose density .

Marshall samples were then made and volume index tests conducted. The aggregate structure of stone mastic asphalt (SMA) was used as the standard, $VCAmix<VCA_{DRC,}$ to determine the structure formed (see Tables 1-4).

1.5 Mechanical Method

The Mechanical Method is based on surface theory where the strength of a mixture is developed through its components. The surface theory indicates that the strength of an asphalt mixture is determined by two types of phenomena; one type is the aggregate to aggregate interlock or the strength of the mineral skeleton, which is represented by frictional resistance (frictional angle φ) on the surface of the granular materials; the other part is the binding strength that is provided through the asphalt bitumen, represented by coacervation forces that provide a stretching resistance and is expressed by cohesion force C.

Table1-5 Technique Index of 31.5mm Nominal Maximum Aggregate Size

Design density (multiple)	Asphalt content (%)	Max possible specific gravity	Saturated surface-dry bulk specific gravity	VV(%)	VCAmix (%)	VCA_{DRC} (%)
1.00	3.3	2.578	2.447	5.1	43.8	40.9
1.02	3.3	2.579	2.451	5.0	41.1	40.9
1.04	3.3	2.579	2.449	5.0	40.7	40.8
1.06	3.3	2.580	2.458	4.8	40.2	40.8
1.08	3.3	2.580	2.455	4.8	39.9	40.8

The ATB used in the evaluation satisfied the performance requirements for permanent deformation resistance and fatigue resistance. But the maximum binding power and maximum friction power of the mixture have different optimum asphalt contents. A suitable asphalt content must be selected to obtain a high-quality mixture to ensure the binding power and friction power work together. In the condition of maximum compaction, the asphalt content has little effect on the friction resistance power. Therefore, the maximum binding power optimal asphalt content can be used ensuring good deformation and fatigue resistance.

Once the gradation has been selected, the following steps can be used to design the mixture using the mechanical method:

(1) Compact the mixture to the maximum compaction level, measure the density, and percentage of voids of mixture sample at multiple asphalt contents;

(2) Conduct compression and splitting strength tests, determine the values for R, and r at the different asphalt contents;

(3) Construct the graph showing the relation of R×r, R, r versus asphalt content, and select the optimum asphalt content;

(4) Determine the percentage of voids, VMA, at the optimum asphalt content.

The aggregate gradation was determined using the Bailey design method. The height of column sample was 95.3 mm, the diameter was 150 mm, and it was shaped using the Marshall break-compaction device. The numerical value of break-compaction was 112 times. The compression strength test was used to measure R and the splitting strength test was used to measure r. Based on conversion relationship between compression strength and indirect tensile, the indirect tensile strength r can be scaled so that the relationship of R×r, and r versus asphalt content can be developed. The graph is not included in this paper.

The optimum asphalt content was determined based on the maximum value of R×r , the percentage of voids (VMA), experience, and project details. The optimum asphalt content for the BLF-25 gradation was 3.6%, the optimum asphalt content for the BLF-30 gradation was 3.3%, and the optimum asphalt content for the BLF-25 gradation was 3.0%.

2. Comparison of Different Design Methods

A comparison of ATB properties obtained through the different design methods is shown in table 1-6. Through the comparison of physical index and mechanical index properties using the different design methods, it can be seen that using the aggregate graduation designed by Bailey method and selecting the optimum asphalt content using the Marshall test, an ATB that exhibits good performance can be designed.

Table 1-6 Qualification Comparison Table

Property	Recommended Specification Method	SUPERPAVE Method	Mechanical Method
Ratio of coarse	Higher	lower	-
Ratio of medium	lower	Higher	-
Ratio of fine	Higher	medium	-
asphalt use level	medium	lower	Higher
percentage of voids	Higher	lower	medium
VFA	lower	Higher	medium
Ratio of filler to bitumen	medium	lower	Higher
Effective thickness of bitumen film	medium	lower	Higher

The methods can be summarized as:

(1) The large scale Marshall compaction test method is the recommended compaction method in specifications for ATB design. The equipment used in this method is simple, easy to operate, and calculations using this method are easy. But sample compaction using this method does not accurately simulate field compaction. Additionally, the stability index and flow value has little meaning because of the large aggregate size of ATB, especially the nominal maximum size of aggregate greater than 26.5mm. More research is required to determine an index that can be used to represent the mixture performance.

(2) In the SUPERPAVE design method, sample compaction is accomplished using the SUPERPAVE Gyrator Compactor (SGC). The SGC provides a more realistic representation of field compaction and the compaction effort is more repeatable. However, the equipment more complex, the operation of the equipment is more difficult requiring more operator training, and the equipment is more expensive. The data is significantly affected by the SGC compaction angle,

which is difficult to verify. Additional research is required concerning the rational of setting a central point and forbidden region.

(3) In the Bailey design method, it is important to select design density that forms an aggregate skeleton to satisfy the three initial design parameters. This will determine the asphalt content and samples can be compacted. The sample density is determined and then VCAmix is calculated and compared with VCA_{DRC}. The aggregate gradation is adjusted to ensure maximum density for the aggregate skeleton structure.

Bailey design method equipment is simple and easy to use and it provides a dense mix gradation. Additional research is required to define the optimum design density and the span of graduation design parameters: CA, FAc, and FAf .

(4) The Mechanical method is based on surface theory in which the strength of the mixture is used to design the ATB. Sample preparation can be accomplished by using the SGC to form large cylinder samples (about φ150 mm×100 mm) or by using the large striking device to form a large cylinder sample (about φ152.4 mm×95.3 mm).

The Mechanical method currently used in ATB design is in the development phase. It does not have a method to select the optimum aggregate gradation; it only provides a method to determine the optimum asphalt content. The Mechanical method can be used along with the Bailey design method. With this process, the aggregate gradation is selected using the Bailey design method and the optimum asphalt content is selected using the Mechanical method. The Mechanical method is simple to use but research is required to determine the applicability of the method to different aggregate materials and gradations.

3. Conclusions

Through laboratory analysis and evaluation of project samples, it has been shown that each ATB design method has advantages and disadvantages. Based on this study, it is recommended that the Bailey design method be used to select the aggregate gradation and the Marshall method be used to select the optimum asphalt content. It is believed that this design method can be used to design high performing ATB. However, additional research is required to determine if the laboratory analysis can be correlated to field performance.

References

[1] Technical Specifications for Construction of Highway Asphalt Pavements . JTG F40-2004.

[2] HAO Pei-wen, XU Jin-zhi . Key technologies of aggregate blending by Bailey method . Journal of Chang'an university (natural science edition) , vol.24 No.6

[3] Yu xin,Wu Jian-hao. Discussion About Application of Bailey Method. Highway. Aug.2003-No8.

Application of Finite Element Analysis to Access the Rutting Potential in Asphalt Pavements

Xu-Qiu Teng, Ph.D.[1]; Xiao-Zhong Li[2]; and Katie Chou, Ph.D., M.ASCE[3]
1 Assistant Professor, Dept. of Civil Engineering, Lanzhou Jiaotong Univ., 88 West Anning Rd., Lanzhou 730070, China; E-mail: Lucyphd2003@126.com
2 Assistant Professor, Dept. of Highway Engineering, Lanzhou Jiaotong Univ., 88 West Anning Rd., Lanzhou 730070, China; E-mail: Lxztxq@126.com
3 Associate, CH2M HILL Inc., 15010 Conference Center Dr., Chantilly, VA 20151, U.S.A.; E-mail: kchou@ch2m.com

Abstract: The finite element analysis program, ANSYS, is utilized in this research to numerically assess the rutting depth of an asphalt pavement under selected loading conditions. A generalized Maxwell model is used to simulate the viscoelastic behavior of the hot mix asphalt. The viscoelastic parameters of asphalt materials are obtained using a Frequency Sweep Test. Excluding the hot mix asphalt, a Drucker-Prager model is utilized to simulate material behaviors of underlying pavement courses. Twenty-five selected cases are analyzed to better understand the effects of rutting for asphalt pavements. Factors considered in this study include the magnitude of wheel loading, the wander distance, the speed of wheel loading, the number of load repetitions, and the modulus of subgrade. The analysis results indicate that the magnitude of loading and the modulus of subgrade have the greatest impact on the rutting development followed by the factor of wander distance. The speed of wheel loading and the number of load repetitions are found to be insignificant in quantifying the rutting depth.

Keywords: Rutting; Finite Element Analysis; Asphalt Pavements; ANSYS; Maxwell Model; Drucker-Prager Model.

Introduction

Due to a substantial traffic growth in China, the pavement distress of rutting has become a critical issue of newly constructed asphalt pavements. The rutting of an asphalt wearing course not only impacts its surface smoothness but also raises a safety concern of the general public. The accumulated rutting depth may lead to the development of new distresses and exacerbate the condition of existing ones. With numerous benefits of utilizing a flexible base with an asphalt pavement, this asphalt pavement has been widely applied in China. Some researchers and practitioners are worried that a thicker asphalt layer may result in a deeper rutting depth of the asphalt pavement.

Since the application of an asphalt pavement with a flexible base has a relatively short history in China, there is little field data to better understand the development of rutting. As a result, a sophisticate finite element analysis (FEA) program, ANSYS, is chosen to mathematically analyze a rutting depth of a flexible pavement.

Finite Element Model Construction

The FEA program, ANSYS, is utilized to numerically estimate a critical rutting depth of a flexible pavement under various conditions. The versatile finite element model of the flexible pavement is constructed using assumptions below.

Assumptions

1. The material behavior of hot mix asphalt is assumed to be viscoelasticity. Additionally, a Drucker-Prager model (D-P model) is assumed for materials of underlying pavement courses.
2. The total wheel load is a constant during the analysis with a uniformly distributed contact tire pressure.
3. The traffic induced pavement responses are zero at model boundaries.
4. The interface between pavement courses is fully bonded and continual.

Model Size and Boundary Conditions

A layered elastic pavement analysis program, BISAR, is used to verify that a proper model extension of the flexible pavement is established. The size of the asphalt pavement model is increased until stable pavement responses are obtained. The numerical predictions of stresses and strains using ANSYS are compared with the outputs of BISAR for cases of layered elasticity. In this research, the asphalt pavement model consists of 18cm asphalt wearing course, 30cm asphalt stabilized base course, and 30cm treated subbase course. Due to a plane of symmetry, only half of the pavement model needs to be simulated. A minimum model size of the asphalt pavement is 5m × 5m × 7.5m with the plane of symmetry along the Y axis.

Boundary conditions are crucial in the FEA. A constraint of zero horizontal displacements is imposed along the plane of symmetry and borders of horizontal model extents. The constraint of zero vertical displacements is imposed at the bottom of the subgrade.

Material Models

The viscoelastic parameters of asphalt materials are obtained using the Frequency Sweep Test. A generalized Maxwell model with four (n) Maxwell elements and one spring element as shown in Figure 1 is used to simulate the viscoelasticity of asphalt materials. Since an effective temperature ranges between 33 °C and 37 °C in China (Zhang and Li, 1995), the effective temperature of 35 °C is employed in this study.

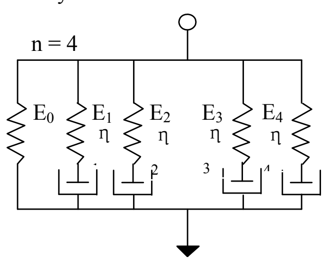


Figure 1. A generalized Maxwell model.

An asphalt concrete mix (AC-20), and a large stone asphalt mix (LSAM-40) are used in surface and base courses, respectively. The viscoelastic parameters of surface and base are shown in Table 1 where E is the Young's Modulus and η is the viscosity.

Table 1. Viscoelastic Parameters

E (MPa)	AC-20	LSAM-40
E_0	114	563
E_1	167	1,107
E_2	96,790	775,322
E_3	126,589	404
E_4	84,159	1,340,368
η (MPa.s)	AC-20	LSAM-40
η_1	69,284	22,010
η_2	446	4,068
η_3	1,028	794,015
η_4	165	2,581

Additionally, the elastic modulus of subbase materials is 350MPa with a cohesion of 0.06MPa and an internal friction angle of 25°. To better understand the correlation between the rutting potential and the subgrade strength, five moduli of subgrade (40MPa to 120MPa with a 20MPa increment) are examined. The selected cases for FEA will be discussed in a later section. A cohesion of 0.03MPa and an internal friction angle of 25° are used in subgrade materials.

Wander Distance and Equivalent Loading Time

Field test results indicate that the relationship between a loading time and a wander distance is a normal distribution curve (Hua, 2000). This normal distribution curve has a mean (μ) at the centerline of a travel lane, a standard deviation (s), and a maximum wander distance (D). In this research, a wander distance represents a distance between the centerline of a travel lane and the center of a wheel load. Five wander distances (0.14m to 0.42m with a 0.07m increment) are studied to prove if the wander distance plays a major role in the development of rutting.

The time versus the wander distance curve follows Eq. 1 where s = D/3. There is a 99.74 percent probability that a wheel pass (X) occurs within the wander distance on the either side of the travel lane centerline (Sheng, 1997).

$$P\{\mu - 3s \prec X \le \mu + 3s\} = \phi(3) - \phi(-3) = 0.9974 \quad (1)$$

The ratios of loading time are plotted against three wander distances for the dual wheel loadings as shown in Figure 2. The dual spacing between wheels is approximately 0.16m and the tire width (W) is 0.1566m.

A detailed correlation between a loading time and an element length is extremely complicated. To simplify the analysis procedure, a rational method of the principle of equal area technique is applied to estimate an equivalent loading time for a given element. Therefore, the equivalent loading time is a constant for an individual element. Figure 3 illustrates the conversion between the original loading curve and the equivalent loading curve.

The loading time of position x = χ is calculated using Eq. 2 (Hua, 2000) where the functions, F(x) and T, represent the normal accumulated distribution and the loading time, respectively.

$$t_{(\chi)} = \left[F\left(\chi + \frac{W}{2}\right) - F\left(\chi - \frac{W}{2}\right) \right] \times T \quad (2)$$

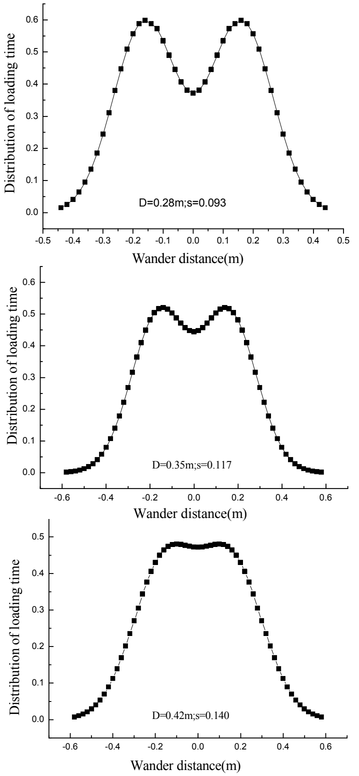


Figure 2. Loading time versus wander distance (dual-wheel loading).

Selected Cases for Analysis

Five key factors are considered in this study including the magnitude of wheel loading, the wander distance, the speed of wheel loading, the number of load repetitions, and the modulus of subgrade. It is too time consuming to perform a full spectrum of parametric studies for five chosen factors. Instead, an orthogonal layout is applied to arrange the calculation sequence.

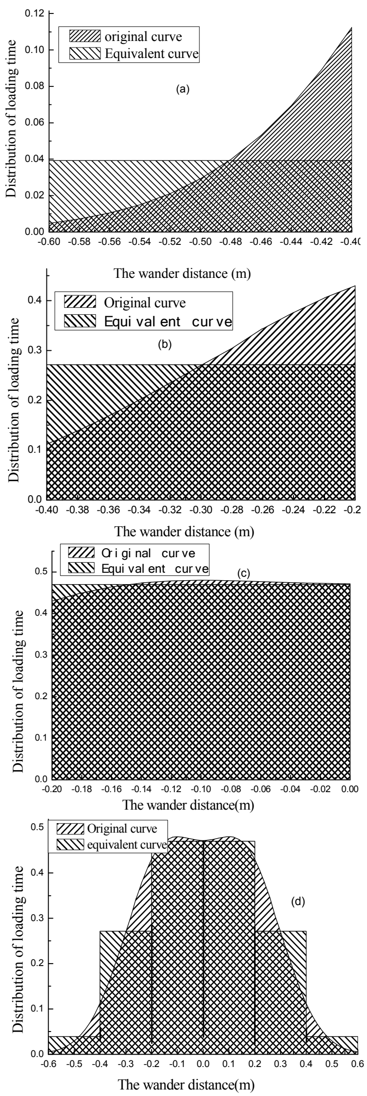


Figure 3. Illustration of the equivalent loading time.

Twenty five selected cases as listed in Table 2 are analyzed using ANSYS to numerically evaluate the critical rutting depths.

Analysis Results

A squariance of total variation is composed of a squariance of factor variation (S_j) and a stochastic error (S_E). The parameter, F_j, conforms to the F-distribution of probability statistics and is calculated using Eq. 3. The F_j indicates if a significant effect of factor J is observed for results of variance analysis. Based on the analysis results, the significance levels of five factors are determined as shown in Table 3.

$$F_j = \left(S_j / f_j\right) / \left(S_E / f_E\right) \tag{3}$$

Note that parameters, f_j and f_E, represent the degree of freedom of the J column and the error column.

As shown in Table 3, the subgrade modulus and the load magnitude have the greatest impact on the development of rutting. The wander distance also plays a significant role in contributing the rutting depth. Yet, the wheel loading speed and the number of load repetitions are found to be insignificant in developing the rutting.

To effectively reduce the rutting depth of an asphalt pavement, both the load magnitude and the subgrade strength need to be considered. The subgrade needs to be properly compacted during construction to provide an adequate support for the above pavement structures. The heavy trucks need to be monitored and controlled to prevent a severe growth of rutting caused by excessive wheel loads.

Conclusion

The rutting depth is numerically computed by using a commercial FEA program, ANSYS. Five factors are studied to evaluate their contributions of rutting development for asphalt pavements. The analysis results show that the load magnitude and the subgrade modulus have the greatest impact on the rutting development follows by the wander distance. To minimize the rutting depth that may lead to new distresses and safety issues of asphalt pavements, the subgrade needs to be properly prepared during construction. Subsequently, the heavy trucks need to be controlled during the service life of flexible pavements to reduce the wheel load induced rutting.

References

Guo, D.-Z., and Ren, R.-B. (2001). “Layered mechanics of viscoelastic material.” MS thesis, Harbin Institute of Technology Press.

Zhang, D. and Li, J. (1995). “Prediction of rutting in asphalt pavement.” *Journal of China Highway and Transport*, 8(1).

Sheng, J. (1997). “Probability theory and mathematical statistics.” MS thesis, Higher Education Press.

Hua, J.-F. (2000). “Finite element modeling and analysis of accelerated pavement testing devices and rutting phenomenon.” PhD dissertation, Purdue University.

Wang, R.-X. (2000). “Mathematical statistics.” MS thesis, Xi'an Communication University Press.

Table 2. Results of Numerical Analysis

Parameters / Test number	Load magnitude, A (KN)	Wander distance, B (m)	Wheel speed, C (Km/h)	Load repetitions, M	Empty column, E	Subgrade modulus, F (MPa)	Rut depth (mm)
1	75	0.14	40	1,000	1	40	1.387
2	75	0.21	60	10,000	2	60	1.579
3	75	0.28	80	100,000	3	80	1.295
4	75	0.35	100	1,000,000	4	100	1.420
5	75	0.42	120	10,000,000	5	120	1.257
6	100	0.14	60	100,000	4	120	0.971
7	100	0.21	80	1,000,000	5	40	2.851
8	100	0.28	100	10,000,000	1	60	2.148
9	100	0.35	120	1,000	2	80	1.874
10	100	0.42	40	10,000	3	100	1.947
11	125	0.14	80	10,000,000	2	100	1.349
12	125	0.21	100	1,000	3	120	1.399
13	125	0.28	120	10,000	4	40	3.615
14	125	0.35	40	100,000	5	60	3.439
15	125	0.42	60	1,000,000	1	80	2.826
16	150	0.14	100	10,000	5	80	1.842
17	150	0.21	120	100,000	1	100	2.303
18	150	0.28	40	1,000,000	2	120	2.047
19	150	0.35	60	10,000,000	3	40	5.535
20	150	0.42	80	1,000	4	60	4.058
21	175	0.14	120	1,000,000	3	60	2.674
22	175	0.21	40	10,000,000	4	80	3.139
23	175	0.28	60	1,000	5	100	2.350
24	175	0.35	80	10,000	1	120	3.039
25	175	0.42	100	100,000	2	40	6.478

Table 3. Results of Variance Analysis with $F_{0.05}(4,4) = 6.390$

Parameters / Variation source	S_j	f_j	$\overline{S_j} = S_j / f_j$	F_j	Significance level
Load magnitude, A	15.1788	4	3.7947	30.4	Highly significant
Wander distance, B	9.0572	4	2.2643	18.2	Significant
Wheel speed, C	0.4165	4	0.1041	0.84	Insignificant
Load repetitions, M	1.5058	4	0.3765	3.02	Insignificant
Subgrade modulus, F	16.5317	4	4.1329	33.1	Highly significant
Error(E)	0.4988	4	0.1247		

Rut Prediction for Semi-rigid Asphalt Pavements

K. Su[1], L.J. Sun[2] and Y. Hachiya[3]

[1] Ph.D., Airport research Center, Port and Airport Research Institute, 3-1-1, Nagase, Yokosuka, 239-0826, Japan; PH (+81-46)844-5641; Email:xahusukai@hotmail.com;

[2] Prof., Department of Transportation Engineering, TongJi University, Shanghai, 200- 0092, China; PH (+86-21) 65980561; email: ljsun@mail.tongji.edu.cn:

[3] Ph.D., Airport research Center, Port and Airport Research Institute, 3-1-1, Nagase, Yokosuka, 239-0826, Japan; PH (+81-46)844-5400; Email: hachiya@ipc.pari.go.jp;

Abstract

A new rutting prediction model for semi-rigid asphalt pavements is put forward based on the shear deformation mechanism. The most notable feature of this model is the efficient consideration of shear factors, namely, shear-strength is employed to evaluate the anti-deformation capacity of asphalt mixture and shear stress is used to represent the resistance of pavement structure to rutting. Four asphalt mixtures are used to carry out wheel tracking tests at different temperatures and contact pressures, and then the model is determined by the optimization analysis method. Meanwhile, the vehicle speed, which has a pronounced influence on rutting, is successfully introduced into the model by means of Boltzmann's linear superposition principle. Finally, the model is calibrated using accelerated pavement test results.

Introduction

Accurate estimates of rutting are an essential input for the efficient pavement management systems. Rutting in asphalt pavement includes densification and shear flow of hot-mix asphalt, whereas the majority of severe rutting is caused by the shear flow within the asphalt mixtures (Eisenmann et al. 1987 and Myers et al. 2002). This is especially true for semi-rigid asphalt pavements as a result of their stiffer stabilized base: in this case, asphalt layer is responsible for most of the shear deformation. In the past, efforts have been made by a number of pavement researchers to apply the shear concept to the mix design of asphalt concrete (McLeod 1951 and Monismith et al. 2006). However, researches in this direction have not achieved a widespread accepted result by now, probably due to the complexity of the test methods, such as triaxial tests and repeated simple shear test at constant height.

The paper presents a new mechanistic-empirical approach to predict rutting in semi-rigid asphalt pavements. A simple and practical shear test method is employed to take into account the shear properties of asphalt concrete: Static Uniaxial Penetration Test (SUPT) (Sun et al. 2006). This work aims to lay the foundation for putting forward appropriate design procedures to control rutting in the future.

Framework of Rut Prediction Model

Much research (Eisenmann et al. 1987, Sousa et al. 1994 and Myers et al. 2002) has indicated that shear deformation is mainly associated with pavement structure characteristic, pavement materials properties, temperature and traffic (vehicle speed, load magnitude and repetitions). Traditionally, predictions of rutting progress have been based on the most widely used power equation as shown in Equation (1) (SHRP 1993), which empirically accounts for the rut depth as a function of temperature and loading repetitions.

$$RD = \alpha \times N^{\beta} \times T^{\theta} \tag{1}$$

where *RD* is the rut depth after the loading repetitions of *N* at the temperature of *T*, α, β and θ are the coefficients of the equation.

However, recent studies indicate that the exponent of loading repetitions generally varies with exposure to different loads or if different materials are used (Su 2006). The value of β in Equation (1) mainly depends on the magnitude of stress and the properties of asphalt mixture. To reflect this, Equation (1) is modified to give Equation (2) by introducing a function for shear stress and shear strength instead of the constant β value above.

$$RD = \alpha \times N^{m} \times T^{\theta} \tag{2}$$

where $m = (\tau/\tau_0)^{\mu}$, τ is shear stress in the pavement, which can be calculated by the finite element method, and τ_0 accounts for the shear strength of asphalt concrete measured by SUPT. Note that the ratio of τ to τ_0 can minimize the error resulting from using elastic theory for computing τ_0 and τ. In this model, shear stress is used to differentiate the resistance of pavement structures to rutting, and shear strength is only used to evaluate the shear strength of asphalt mixtures, therefore, 20°C and 60°C were respectively designated as their representative temperature.

Because of the viscous property of asphalt concrete, lower speed usually results in greater rut depth, as found on gradients and at intersections. However, it is very difficult to precisely capture the relationship between deformation and speed due to its complexity. Therefore, the Boltzmann’s linear superposition principle (Guo 2001), that is, that the total rut depth can be directly calculated by summing all the deformation increments at different times, is coupled with another empirical equation stating that the loading duration is the inverse of vehicle speed (Pell and Taylor 1974) to take into account the influence of vehicle speed on rutting. In other words, the rut depth caused by one application of a load operation with the speed of 10 cm/s is equivalent to that caused by ten similar loadings operation at a speed of 100 cm/s. The resulting complete prediction model is illustrated in Equation (3), where V_{ref} is the reference speed and N_v is the number of load repetitions at the speed of interest V.

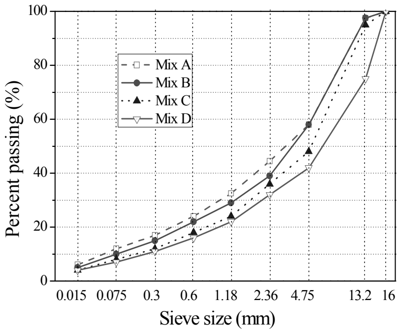
$$RD=\alpha\times(\frac{V_{ref}}{V}\cdot N_V)^m\times(T)^\theta \tag{3}$$

In reality, the shear stress, temperature and properties of asphalt concrete vary through the depth in asphalt pavement. To take this into account, a method of accumulating all sub-layer deformations as the total deformation was employed. The thickness of each sub-layer is defined as 10mm, and the mid-point of each sub-layer is designed as the computation point. Thus, one can obtain the total rut depth by simply summing all deformation increment through the entire thickness based on the actual temperature, shear stress and shear strength in each sub-layer.

Laboratory Experimental Program

Wheel Tracking Test. A wheel tracking test with a solid rubber-faced tire was used to provide the data of the rut evolution with the repetitions of loading. The slab specimen used in this test was 300mm wide, 300mm long and had a thickness of 50mm. Tests were carried out at three temperatures, 20°C, 40°C and 60°C at a constant tracking speed of 0.16m/s. Four asphalt mixtures were used, of which aggregate gradations were as shown in **Figure 1**. Mix A was tested at three pressure levels of 0.56MPa, 0.72MPa and 1.10 MPa, whereas mix B, C and D were tested at the standard pressure of 0.72MPa. The wheel tracking test results are shown in **Figure 2**. Just as mentioned above, pronounced differences in deformation development trend were observed for different asphalt mixtures or under various load levels.

Static Uniaxial Penetration Test. Basically, the shear resistance of asphalt concrete as an intrinsic property of the material and should be constant under given conditions. However, variations in confining pressure mean that triaxial test fails to give a unique value for shear strength. Considering this deficiency and also the complexity of triaxial test, a static unixial penetration test (**Figure. 3**) was



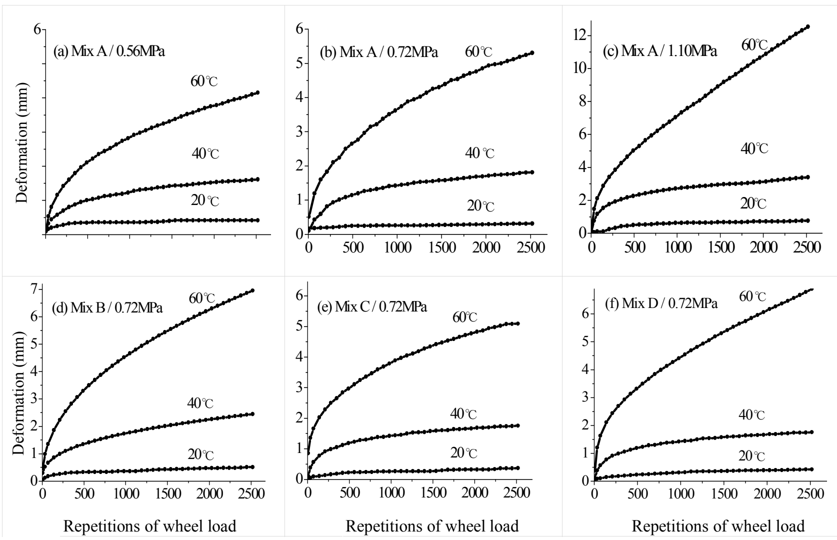


Figure 2. Results of wheel tracking tests

employed, named SUPT. This test can directly determine the shear strength of asphalt mixture as verified by Bi (Bi 2004), who showed that the distribution of shear stress in a SUPT specimen agreed perfectly with that in a real pavement structure, and also the damage mechanism in SUPT was mainly dependent on shear forces. Chen has further demonstrated the superior repeatability of SUPT and consistency with triaxial test at 0.3MPa confining pressure (Chen 2006).

In this test, the specimen size was 100mm in diameter by 100mm in height. Such specimens can be easily molded by the Superpave gyratory compactor. The loading head (a steel rod with 28.5 mm diameter) penetrates the SUPT specimen at a strain -controlled loading rate of 1mm/min. The shear strength denoted τ_0 is defined as:

$$\tau_0 = k \cdot P_{max} \tag{4}$$

where P_{max} is the penetration strength provided by the SUPT specimen, and k is named as the shear strength coefficients with a suggested value of 0.339 based on the ratio of maximum shear stress to the vertical circular pressure of 1MPa applied on a semi-infinitive space when the Poisson ratio is 0.35.

Prior to testing, each specimen was held in an environmental chamber for over 5 hours at 60°C. Shear strength tests of mix A, B, C and D were conducted, and the shear strength values were 1.063, 0.861, 1.152 and 0.984 MPa, respectively.

Finite Element Analysis of Shear Stress. The shear stress in the slab specimen used for wheel-tracking test was calculated by a three-dimensional finite element method as shown in **Table 1**. Note that the measured resilient modulus of asphalt concrete was used to compute the shear stress, but the results indicated that resilient modulus had negligible effect on the value of shear stress in this finite element model.

Determination of Prediction Model by Optimization Analysis

The model was determined by fitting the wheel-tracking test results using the Solver optimization technique in Excel by minimizing the sum of errors arising from the error (SSE) between the measured and the predicted rutting. In this case, every input data point corresponds to a rut depth, a temperature and a shear stress as well as a shear strength. This model, as expressed in Equation (5), gives an excellent fit as evidenced by a higher value of determined coefficient up to 0.838.

$$RD = \sum_{i=1}^{n} 10^{-5.542} \cdot T_i^{2.524} \cdot \left\{ \frac{0.58}{V} \cdot N_V \right\}^{0.752 \cdot \left(\frac{\tau_i}{\tau_{0i}} \right)^{0.468}} \tag{5}$$

To shift the model from the laboratory to actual pavement, a full-scale road test was performed to provide confirmatory evidence for the laboratory tests. This calibration not only considers the downward rut depth but also upheaval on the two neighboring sides of rut. Here, Equation (5) becomes:

$$RD = \sum_{i=1}^{n} 10^{-5.72} \cdot T_i^{2.512} \cdot \left\{ \frac{0.58}{V} \cdot N_V \right\}^{0.743 \cdot \left(\frac{\tau_i}{\tau_{0i}} \right)^{0.472}} \tag{6}$$

Table 1. Shear stress in the wheel tracking slab (MPa)

Tire pressure	0～1 cm	1～2 cm	2～3 cm	3～4 cm	4～5 cm
0.56 MPa	0.1993	0.1502	0.0999	0.0735	0.0623
0.72 MPa	0.2395	0.2184	0.1498	0.1114	0.0948
1.10 MPa	0.3414	0.3676	0.2662	0.2017	0.1729

Conclusions

In this study, a method is developed for the prediction of rutting in semi-rigid asphalt pavements, and the most significant enhancement in this new approach is that the concept of shear is successfully introduced. The new approach takes into account traffic and environmental characteristic as well as the pavement structure and the properties of the materials. The fit of this model when both rutting and local upheavals as well as temperature gradient are taken into account is improved after a calibration using the results obtained in a full-scale road test.

Though this new method looks very promising, considerable efforts are still needed to assess the new rut model before it can be applied to the actual pavements.

References

Eisenmann J. and Hilmer A. (1987). "Influence of wheel load and inflation pressure on the rutting effect at asphalt pavement experiments and theoretical investigations." Proceedings of the sixth international conference on the structural design of asphalt pavements, 392-403, Ann Arbor.

Guo D. Z and Ren R. B. (2001). "Visco-elastic mechanics in multi-layered pavements." Haer Bin Technology University Press, China.

Myers L. A., Drakos C. and Roque R. (2002). "The combined effects of tire contact stresses and environment on surface rutting and cracking performance." Proceedings of the ninth international conference on asphalt pavements, Copenhagen, Denmark.

Sun L. J. and Bi Y. (2006). "Top-down cracking analysis and control for asphalt pavements." Proceedings of the tenth international conference on the structural design of asphalt pavements, Quebec, Canada.

McLeod, N. W. (1950). "A rational approach to the design of bituminous paving mixtures." Journal of the Association of Asphalt Paving Technologists, Vol.19. 82–87.

Pell P. S. and Taylor F. (1974). "Fatigue characteristic of bituminous materials." Journal of the Association of Asphalt Paving Technologists, Vol.43.

Su K. (2007). "Mechanism and prediction model of rutting in asphalt pavements." Ph.D. Dissertation, School of Transportation Engineering, Tongji University, Shanghai, China.

Sousa J. B., Craus J. and Monimith C. L. (1994). "Summary report on permanent deformation on asphalt concrete." SHRP-A-318, Strategic Highway Research Program, National Research Council, Washington D.C., 40-60

Chen X. W., Huang B. S. and Xu Z. H. (2006). "Uniaxial penetration testing for shear resistance of hot mix asphalt mixtures." 2006 TRB annual meeting CD-ROM, Washington D.C.

Moisture damage of asphalt mixes modified with SEAM pellets

Sheng Xiaojun[1], Ali Selim[2], Sun Lijun[1]
1 Key Laboratory of Road and Traffic Engineering of the Ministry of Education ,Tongji University ,Shanghai,201804 ,China
2 South Dakota State University, Brookings , SD, 57007 USA

Abstract: In China, the use of Sulfur Extended Asphalt Modifier (SEAM) has been recently introduced to the pavement community. This asphalt modifier does not only conserve energy, reduce asphalt content by 16%-40%, but also has good performance in terms of high strength and good rutting resistance. Preliminary experience with asphalt mixtures containing SEAM showed vulnerability to moisture damage. To improve upon resistance to moisture damage, large number of experiments was conducted, such as retained Marshall Stability, Freeze-thaw, and split test on four different types of aggregate, and five different types of anti-stripping agents. The influences of gradation, air void and asphalt content on reducing moisture damage due to the inclusion of SEAM were also analyzed. The improvement to different aggregate caused by different anti-stripping agent differs significantly. Using coarser gradation yielded higher resistance to moisture damage of mixtures containing SEAM.

Key words: SEAM, moisture damage, Anti-stripping agent, Gradation, Asphalt content

Introduction: As a new type of asphalt modifier and additive, SEAM does not only reduces mixing and compaction temperatures of asphalt mixture, conserves energy, reduces asphalt content by 16%-40%, but also has good performance in terms of high strength and good rutting resistance[1,2]. SEAM was introduced into China in 2002, and was successfully used in both Ji-Gu and, Yu-Jin Highways in Tianjin, China. Since its introduction, some problems were encountered, especially the low resistance to moisture damage which limited their applications. Hence, the urgent need to improve the resistance to moisture damage and achieve optimal performance was the main tasks for this research work. A large number of experiments were conducted, such as Retained Marshall Stability Test, Freeze-thaw, and Split test. These tests were performed on four different types of aggregate; five different types of anti-stripping agents to determine what combination of material and anti-stripping agent type/dosage yielded the best mixture that resisted moisture damage the most. The influences of gradation, air void and asphalt content to the SEAM's resistance to moisture damage were also analyzed.

Test materials

asphalt: Shell AH-70, Penetration 68, Softening point 47.6˚C, Ductility(15 °C)> 150.

SEAM: Dark brown, granular, density 1.991 g/cm^3.

Anti-stripping agents: **solid** AR-78,

liquid PAVEBONELIFE, PA-1, MORLIFE-2200 and TJ -066.

Table 1 Aggregate Test [4]

Aggregate Source	crushed stone value	wearied stone value	flat and elongated particle
Sichuan granite	23.7%,	17.9%,	19.1%;
Changshu,limestone	20.6%,	18.7%,	31.3%;
Wuxi limestone	20.4%,	17.1%,	18.8%;
Henan limestone	19.7%,	17.5%,	16.2%.

Asphalt and coarse aggregate adhesion test

To assess the adhesion characteristics between the binder and the aggregate, the boiling water test was conducted. Test results revealed that base binder had as low value as 2 for the Sichuan granite and as high as 5 for the Wuxi limestone. After adding the SEAM all four aggregates achieved 5 which is the highest value expected. The proportions of the base binder to the STEAM were 60:40. Later analysis revealed that this test alone does not represent how the entire mixture behaves when it is subjected to other performance tests.

Using Anti-stripping agents to enhance resistance to moisture damage of SEAM asphalt mixtures

Experiments were conducted on the base asphalt mixture, SEAM asphalt mixture, and SEAM asphalt mixture with different anti-stripping agent at different doses, to determine the main cause of moisture damage and find out means for improvement. Due to space limitation, only Sichuan test data will be addressed in this paper.

Table 2 shows the parameters obtained from the Marshall Mix Design. The Marshall mix design method was used with mixing temperature of 130°C and mixing time of 3 minutes.

To examine the susceptibility of various asphalt mixtures to moisture damage, the Marshall immersion test was conducted. Table 3 reveals the results of that test.

Table 2 Marshall results on Base Binder and SEAM [5]

Mixture types	Optimum AC (AWB) (%)	Bulk Density g/cm^3	Theo. Density g/cm^3	VV %	VFA %	VMA %	MS KN	FL 0.01 mm
Binder	5.0	2.382	2.492	4.4	70.1	14.8	13.7	30.5
SEAM	5.97	2.391	2.526	5.3	65.2	15.3	13.3	34

Table 3 Immersion Marshall Test Results [5]

Mixture types	Anti-stripping dosage（%）	0.5h Stability（KN）	48h Stability（KN）	Retained Stability（%）
Base Asphalt	-	13.7	12.4	90.5
SEAM Asphalt	-	13.3	10.5	78.6
SEAM Asphalt（TJ-066）	2	12.5	7.0	56.0
	4	13.9	12.8	92.1
	6	14.0	12.2	87.1
	8	11.5	9.5	82.6
SEAM Asphalt（AR-78）	2	11.5	12.5	108.7
	4	12.2	13.2	108.2
	6	15.0	13.5	90.0
	8	17.2	14.2	82.6
SEAM Asphalt（PAVEBONELIFE）	4	15.6	14.2	91.0
	6	16.1	15.3	95.0
	8	13.6	16.5	121.3
SEAM Asphalt（MORLIFE-2200）	4	14.5	8.6	59.3
	6	14.0	11.4	81.4
	8	13.3	10.5	78.9
SEAM Asphalt（PA-1）	4	11.5	10.4	90.4
	6	12.1	11.1	91.7
	8	11.4	13.0	114.0

Table 3 shows that when SEAM was added, residual stability declined from 90.5 % to 78.6 %, Because of that decline, it was clear that inclusion of anti-stripping agent was inevitable. After adding various anti-stripping agents at various dosage, retained Marshall Stability generally improved or exceeded the demand level. To further explore how mixtures containing both sulphur and anti-stripping agents behave under traffic loads, the freeze-thaw test was conducted on several specimens. Table 4 shows test results.

Table 4 Freeze-Thaw Split Test Result [5]

Mixture types	Anti-stripping dosage（‰）	Splitting strength before frost (KN）	Splitting freeze-thaw cycle strength (KN）	Splitting tensile strength （MPa）	TSR （%）
Base Asphalt	-	10.54	-	1.01	54.5
		-	5.87	0.55	
SEAM Asphalt	-	11.89	-	1.15	24.9
		-	2.99	0.29	
SEAM Asphalt （TJ-066）	2	15.30	-	1.46	29.3
		-	4.43	0.43	
	4	14.16	-	1.36	44.9
		-	6.29	0.61	
	6	14.75	-	1.42	61.6
		-	9.16	0.87	
	8	14.72	-	1.43	51.5
		-	7.54	0.74	
SEAM Asphalt （AR-78）	2	15.52	-	1.48	64.4
		-	9.82	0.96	
	4	17.95	-	1.73	56.6
		-	10.10	0.92	
	6	15.65	-	1.56	71.5
		-	11.29	1.11	
	8	16.30	-	1.59	64.5
		-	10.60	1.02	
SEAM Asphalt PAVEBONELIFE）	4	18.77	-	1.83	56.5
		-	10.63	1.04	
	6	16.21	-	1.62	82.1
		-	13.55	1.33	
	8	16.13	-	1.57	78.9
		-	12.56	1.24	
SEAM Asphalt MORLIFE-2200）	4	15.92	-	1.53	53.8
		-	8.50	0.82	
	6	14.44	-	1.42	65.5
		-	9.62	0.93	
	8	14.64	-	1.41	64.5
		-	9.37	0.91	
SEAM Asphalt （PA-1）	4	18.34	-	1.77	61.6
		-	11.17	1.09	
	6	18.22	-	1.80	67.2
		-	12.24	1.21	
	8	17.80	-	1.73	61.4
		-	10.89	1.06	

Main conclusion extracted from this test is that the anti-stripping PAVEBONELIFE and AR-78 from at a dosage of 6% works the best. Similar conclusions were reached when liquid anti-stripping agents were incorporated into asphalt mixtures [3]. After doing the same tests to the other 3 aggregates from different origin, it was concluded that using different types of anti-stripping agent improved resistance to moisture damage of all aggregate, however, different anti-stripping agents have different influence on different aggregates. Anti-stripping agent performance ranking is given below in table 5.

Table 5 Anti-stripping agent optimization

Origin	PAVEBONELIFE	AR-78	PA-1	MORLIFE-2200	TJ-066
Sichuan	1	2	3	4	5
Wuxi	1	2	3	4	5
Henan	2	5	1	3	4
Changshu	2	3	1	4	5

Note: 1 represents the best, 5 represents the worst.

Effect of other factors on enhancing resistance to moisture damage.

a- Gradation

Gradations 1 and 2 yielded TSR values of 48.4% and 90.3% respectively concluding that gradation 2 is better.:

b- Air Void

Small VV specimens are better than large VV specimens in anti-water damage properties. Therefore, lower air void improves resistance to moisture damage. Raising mixing temperature appropriately can reduce air void.

c- Asphalt content

Test results showed that the Sichuan stone Moisture damage resistance is better than the other three. The main reason lies in the larger asphalt content. Therefore raising asphalt content can reduce air void and increase binder film resulting in better moisture resistance.

Conclusions

When SEAM is added to an asphalt mixture, the adhesion levels generally rise to five. The single-particle adhesion test results can not really show the ability of the mixture to resist moisture damage. Test showed that resistance to moisture damage generally declined after adding SEAM. The reason is that the reduction of effective asphalt binder leads to decreased strength. Major conclusion of this research work includes:

1. Anti-stripping agents do significantly improve resistance to moisture damage. Different anti-stripping agents have different effects on different

aggregates. For each aggregate, both the right type of anti-stripping agent and its dose must be determined before it is specified for any job.
2. Improving gradation can significantly improve SEAM mixture performance. Results show that too much fine aggregate is not conducive to resisting moisture damage. Coarse gradation performs better and should be considered.
3. Reducing Air Void, within limits, can increase resistance to moisture damage.
4. Increasing asphalt content to reduce air void and improve the effective film thickness of asphalt can definitely improve resistance to moisture damage.

References:

[1] LU Zheng-lan; SUN Li-jun; ZHOU Chao-hui. “Performance and mixture design of high quality asphalt modified by seam”, Acta pertolei sinica(Petrleum Processing section),v21， n5,October,2005,p73-78.
[2] De Filippis (Univ ‘La Sapienza’);Giavarini C. Santarelli, M. L. “Sulphur-extended Asphalt: Reaction Kinetics of H_2S Evolution”, Fuel,v77, n5, April 1998, p459-463.
[3] Selim, A. et. al. “Liquid Anti-stripping Agents can be Advantageous over Hydrated-Lime in Resisting Moisture-Induced Damage in Hot Mix Asphalt” Proceedings of the Fifth International Conference of Bearing Capacity of Roads and Airfields, Trondheim, Norway. July 1998. Vol. II p 1173-1182.
[4] Ministry of Communications – The People’s Republic of China. Test Method of Aggregate for Highway Engineering, Beijing. People’s Communications Publishing House, 2000.
[5] Ministry of Communications – The People’s Republic of China. Standard Test Method of Bitumen and Bituminous Mixtures for Highway Engineering, Beijing. People’s Communications Publishing House, 2000.

Inclusion of Moisture Effect in Fatigue Test for Asphalt Pavements

Qing Lu[1], P.E., and John T. Harvey[2], P.E., Member, ASCE

ABSTRACT: Damage of asphalt mixes due to moisture is a complex phenomenon. Conventional moisture-related tests do not well simulate field conditions, in which traffic loading is an essential component. Flexural beam test has been widely used in current pavement design to evaluate material properties of asphalt mixes, but has not been well explored to study the moisture effect. This paper developed a typical fatigue-based test procedure for comparative evaluation of moisture sensitivity of different mixes. It is a controlled-strain flexural beam fatigue test performed at 20°C, 10 Hz, and 200με on specimens pre-saturated under 635 mm-Hg vacuum for 30 minutes and preconditioned at 60°C for one day. Test results showed that the fatigue based test procedure can distinguish mixes with different moisture sensitivities, and give a ranking of mixes consistent with prior field experience. An extension of the test procedure for use in the pavement design was also discussed, but further work is needed to implement the test procedure.

INTRODUCTION

Moisture damage, which can be understood as the deterioration of asphalt mixes by loss of adhesion between asphalt and aggregate (named "stripping") and/or loss of cohesion within the binder due to water, is a complex phenomenon affected by a variety of factors. Moisture damage in asphalt pavements may develop at various rates. Usually it is perceived as a premature failure, in which asphalt mixes severely disintegrate and stripping is manifest. This type of failure primarily results from poor compatibility between asphalt and aggregate, and often occurs within a few years after construction. Moisture may also act in a slow way, in which stripping is not obvious, but mix performance, such as resistance to fatigue cracking and rutting, is gradually reduced by moisture.

Most current tests for evaluating moisture sensitivity of asphalt mixes, such as the tensile strength ratio (TSR) test and the Hamburg wheel tracking device (HWTD) test, are empirical in nature. The conditioning procedures in these tests generally do not well simulate the actual field conditions. Results from these tests are used to screen mixes

[1]University of California Pavement Research Center, 1353 S. 46th St., Bldg. 452-T, University of California, Berkeley Richmond, CA, 94804-4603; PH (510) 665-3596; FAX (510) 665-3562; e-mail: qlu@ucdavis.edu

[2]Department of Civil and Environmental Eng., University of California at Davis, Engineering III, Room 3153, One Shields Avenue, Davis, CA, 95616; PH (530) 754-6409; FAX (530) 752-1228; e-mail: jtharvey@ucdavis.edu

and to prevent premature failure of asphalt pavements due to water, but cannot be used in performance models to predict the pavement life. There is a need to develop a test procedure that can better simulate the field conditions and can potentially be integrated into the pavement design procedure to predict pavement performance life. Pavement performance based tests, such as fatigue test or simple shear test, hold such promise.

This study focuses on development of a procedure that incorporates moisture effect in the fatigue test, and comparison of the test results from this procedure and two other tests. An extension of the test procedure for use in pavement design is also discussed.

DETERMINATION OF TYPICAL TEST PROCEDURE

Based upon a comparative study conducted in the SHRP-A-003 project (Tayebali et al. 1994), the flexural beam test was selected for fatigue evaluation. It is a four-point bending test in which the middle one-third part of the beam is theoretically subjected to pure bending without any shear deformation. Two loading modes are common in the test: controlled-strain and controlled-stress, in which a fixed sinusoidal wave of deformation or load is applied to the beam. In this study, controlled-strain mode was used because it is relatively simple to operate and it better simulates the field conditions where the deformation of AC layers is partly constrained by the underlying structures. This is more rational for thin AC overlays on old pavements, which is a major rehabilitation practice on current U.S. highways.

To evaluate the moisture sensitivity of asphalt mixes by fatigue response variables, specimens were conditioned by both moisture and repeated loading. The key issue in the development of the test procedure is how to determine an appropriate conditioning procedure. For field pavements, traffic loading and environmental factors change with time in wide ranges. Moisture damage thus develops at various rates under different conditions of moisture content, temperature, and traffic loading. Moisture effect on the fatigue response then should be evaluated under different loading and environmental conditions. This would require a large number of fatigue tests covering the typical loading characteristics (load magnitude, frequency) and environmental characteristics (moisture content, temperature), which is beyond the scope of this study. As an alternative, moisture damage was mainly evaluated at typical worst case scenarios.

The laboratory fatigue test is essentially an accelerated performance test in which the wheel loads applied on pavements in 15-30 years are condensed into the repeated loading applied on specimens in a few hours or days. While moisture damage is presumably partly due to the interaction of traffic loading and moisture, it also develops in the non-loading condition. Moisture damage unrelated to loading may develop slowly for months or years in pavements, which mostly will not occur in the short test period at typical fatigue test temperatures. Therefore, a preconditioning process before the fatigue test was needed to introduce the moisture effect unrelated to loading.

The subsequent work focused on determining typical test parameters and the preconditioning procedure.

Determination of Test Parameters

Three test parameters- test temperature, strain level, and loading frequency- were determined in consistent with the conventional beam fatigue test. The test temperature was set at 20°C. Two strain levels (200$\mu\varepsilon$ and 400$\mu\varepsilon$) were selected for mixes with two different binders (A: AR-4000 and P: PBA-6a) respectively. For a typical pavement structure, 200με is usually around the average value of the actual strain level at the bottom of asphalt concrete layer containing the AR-4000 binder. Mixes containing the PBA-6a binder have stiffness much lower than that of mixes containing the AR-4000 binder. Given the same pavement structure and wheel load, the strain in the mixes containing the PBA-6a binder would be higher than the strain in the mixes containing the AR-4000 binder. To allow for this difference, the strain level selected for mixes containing the PBA-6a binder was increased to 400με. A preliminary study showed that this change of strain level did not change the moisture effect on the fatigue response of the mixes. As in the conventional beam fatigue test, a test frequency of 10 Hz was selected for the sinusoidal loading wave. This frequency simulates in-pavement stress pulses corresponding to vehicle speeds in the 24 to 48 km/h range, and is sufficiently large enough to permit rapid testing while still representing the load pulses generated by rapid moving traffic (Tayebali et al. 1994).

Determination of Preconditioning Parameters

The primary objective of the preconditioning process is to introduce certain amount of moisture damage in the specimen. Three parameters were determined for the preconditioning process: moisture content (or saturation level), conditioning temperature, and conditioning duration. A sensitivity study was performed to identify the relative importance of these parameters. The determination of each parameter was then discussed subsequently.

The sensitivity study included four mixes, consisting of two aggregates (A, B), two treatments (nil, hydrated lime), and one binder (AR-4000). Aggregate B has better compatibility with aggregate A. Two levels were selected for each of the three conditioning parameters. The results revealed that among the three conditioning parameters, the conditioning temperature has most important effect on the moisture resistance of asphalt mixes. High temperature significantly promotes moisture damage in mixes, especially in untreated mixes. On the other hand, the level of moisture content does not significantly affect the extent of moisture damage. The conditioning duration has an intermediate effect. The sensitivity study also revealed that the ranking of the four mixes is consistent when evaluated by initial stiffness, fatigue life, or visual inspection of stripping percentage.

Based upon the above findings, a saturation level of about 50-80 percent was selected as the high moisture level in the specimens. A separate study revealed that the application of 635 mm-Hg vacuum for 30 minutes resulted in a saturation of about 60 percent and did not significantly affect the mix strength.

The preconditioning temperature has significant effect on test results, and needs to be selected carefully. Initially 25°C was selected because it is more common in the pavements. Most mixes conditioned at this temperature for one day, however, showed an extended fatigue life due to moisture instead of reduced, as revealed in an early study (Lu and Harvey 2006a). While in another long-term study, it was found that moisture has a time effect, and fatigue life is usually reduced by moisture after a long-term conditioning at the mild temperature (Lu and Harvey 2006b). Field survey has revealed that moisture exists in pavements for a long period, therefore the one-day conditioning at 25°C tends to be insufficient to introduce the amount of moisture damage that will occur in the field. It was decided to choose 60°C as the preconditioning temperature.

To keep the test duration short, it was decided to condition specimens for one day.

As a summary, the preconditioning procedure was determined as follows: saturate the specimen at 635 mm-Hg vacuum for 30 minutes and then place it in a 60°C water bath for 24 hours. After preconditioning, the specimen was cooled to 20°C and wrapped with Parafilm, a moisture-resistant, thermoplastic flexible plastic sheet, to retain its internal moisture.

COMPARISON OF RESULTS FROM DIFFERENT TESTS

The test procedure determined in the previous section was compared with two common tests, the TSR test and the HWTD test. For the TSR test, the procedure specified in the Caltrans version CTM 371-03 was followed. The HWTD test was conducted in accordance with AASHTO T 324. Results of all three tests are summarized in Table 1.

The fatigue responses of mixes containing two different binders are quite distinct from each other. Mixes containing the AR-4000 binder showed a continuous decrease of stiffness until specimens cracked. Mixes containing the PBA-6a binder initially showed a quick reduction of stiffness, but then the stiffness deterioration became trivial after about one million repetitions, which would need a very long time to reach 50 percent stiffness reduction. The fatigue test was therefore terminated at three million repetitions. The fatigue lives of the PBA-6a mixes are all very large based on extrapolation. Considering the uncertainty introduced by the extrapolation, no inference was made based on these data. A direct examination of the stiffness deterioration curves revealed that except for mix APN, moisture had little influence on the stiffness deterioration process of the PBA-6a mixes, no matter what the preconditioning temperature was. For the mix APN, moisture shifted downward the stiffness deterioration curves, and to a larger extent when the preconditioning temperature was 60°C. For the AR-4000 mixes preconditioned at 60°C, the fatigue life ratios (FLR) shown in Table 1 indicate that the fatigue lives of the two untreated mixes (AAN and BAN) were all reduced by moisture, with BAN less affected than AAN. On the other hand, the fatigue lives of the two treated mixes (AAM and BAM) were all extended by moisture. Based upon the fatigue response, the relative ranking of the mixes are as follows: mixes containing the PBA-6a binder are less affected by moisture than mixes containing the AR-4000 binder; mixes containing aggregate B are less affected than mixes containing aggregate A; mixes treated with hydrated lime are less affected than untreated mixes.

The TSR test results are consistent with the fatigue based test results and the field experience, while the HWTD test does not distinguish mixes containing different aggregates. As a summary, the test procedure determined in this study distinguishes mixes with different moisture sensitivities, gives a ranking of mixes consistent with prior field experience.

TABLE 1 Normalized Fatigue Test Results and TSR, HWTD Test Results for Comparison

Mix Type[a]	Preconditioning Temperature 25°C		Preconditioning Temperature 60°C		Tensile Strength Ratio (%)	Rut Depth after 20,000 Passes (mm)
	ISR	FLR	ISR	FLR		
AAN	0.88	0.30	0.63	0.28	29	6.62
AAM	0.91	1.60	0.83	2.13	85	6.82
BAN	0.93	1.31	0.78	0.66	52	7.84
BAM	0.91	1.74	0.86	1.54	91	6.94
APN	0.85	-	0.77	-	47	42.3
APM	1.00	-	1.16	-	86	10.17
BPN	0.94	-	1.06	-	85	56.40
BPM	1.02	-	1.05	-	100	13.73

[a]First letter represents aggregate (A or B); second letter represents binder (A – AR-4000, P – PBA-6a); third letter represents treatment (N – nil, M – hydrated lime).

INCORPORATION OF MOISTURE EFFECT IN PAVEMENT DESIGN

The use of performance based test to evaluate moisture effect enables us to explicitly incorporate moisture effect in the pavement design, which is impossible in the traditional test case. This section provides a simple example showing the possible application of the test results.

In the traditional asphalt pavement design, fatigue life can be expressed by a function of maximum tensile strain and initial mix stiffness (Monismith et al. 1985):

$$N_f = \alpha (1/\varepsilon_t)^{\beta} (1/S_{mix})^{\gamma} \tag{1}$$

where ε_t = tensile strain, S_{mix} = initial stiffness, N_f = fatigue life, α, β, γ =experimentally determined parameters. The existence of moisture will affect all the variables and parameters on the right side of the above equation, and so also influence the fatigue life. Pavements in the field will experience various environmental conditions, including different moisture contents and temperatures. It is assumed that the pavement condition can be represented by "dry" and "wet" statuses, and the different fatigue responses in these two statuses can be characterized by the laboratory fatigue test in dry and wet conditions respectively. Moreover, fatigue damage is assumed to be cumulative and can be calculated by the linear-sum-of-cycle-ratios, or Miner's Law (Miner 1945). The two assumptions remain to be validated by field data, but they are used here for illustrative purpose. For a particular pavement structure, its fatigue life then can be calculated from the fatigue responses in two conditions and the percentages of load repetitions in two conditions. Specifically, we have

$$n_1 + n_2 = N \tag{2}$$

$$\frac{n_1}{N} = r_1 = 1 - r_2 \tag{3}$$

where N = number of actual allowable traffic load applications, r_1, r_2 = percentage of traffic load applications when the pavement is in condition "dry" or "wet", which can be estimated from traffic and weather data. The actual fatigue life, N, then can be solved as below:

$$N = N_1 N_2 / (r_1 N_2 + r_2 N_1) \qquad (4)$$

As an example, we consider a typical pavement structure consisting of three layers: 0.15 m asphalt concrete, 0.30 m aggregate base, and subgrade. The Possion's ratio is assumed to be 0.35, 0.40 and 0.45 for the three layers respectively, and the modulus of elasticity is assumed to be 240 MPa and 40 MPa for the aggregate base and subgrade respectively. Each of two mixes is used for the asphalt concrete layer: AAN (aggregate A/AR-4000 binder/no treatment) and AAM (aggregate A/AR-4000 binder/hydrated lime). Their initial stiffness and fatigue life at different strain levels in both dry and wet conditions were measured by the flexural beam fatigue test. The parameters for equation (1) are estimated by linear regression on the test data. The average initial stiffness of each mix is input into the linear layered-elastic program ELSYM5 to calculate the maximum principal strain at the bottom of the asphalt concrete layer. With this strain and the initial stiffness, the fatigue life of each mix in each condition is obtained from equation (1). Suppose the pavement structure is in an environment where the percentage of traffic load applications when the pavement is in "dry" condition is 60 percent, its fatigue life in that environment is then estimated by equation (4), and shown in Table 2. It can be seen from Table 2 that when only the dry condition is considered, which is the current design practice, the number of allowable traffic load applications of the untreated mix (AAN) is around nine million, over twice as large as that of the treated mix (AAM). When both dry and wet conditions are considered, however, the number of actual allowable traffic load applications of the treated mix is over twice as large as that of the untreated mix.

TABLE 2 Calculation of Fatigue Life with Moisture Effect Included

Mix	Condition	Initial Stiffness (MPa)	Maximum Principal Strain in AC Layer (micron)	Fatigue Life in one Condition	Percentage of Traffic in Each Condition	Field Fatigue in Composite Conditions
AAN	Dry	10,306	91	8,725,382	60%	1,518,423
	Wet2	6,372	123	678,181	40%	
AAM	Dry	10,820	88	3,400,084	60%	3,444,300
	Wet2	8,848	100	3,512,825	40%	

Although others factors affecting the fatigue life in the field, such as temperature variation, traffic wandering and crack propagation, have not been considered in the analysis, the example above clearly shows the significant effect of moisture on the key parameter (fatigue life) in pavement design and its explicit inclusion in the pavement design procedure, which should be superior to most current design practices that vaguely include the moisture effect in a general shift factor.

CONCLUSIONS

This paper focuses on the development of the fatigue based test procedure for evaluating moisture sensitivity of asphalt mixes. A typical test procedure was determined for comparative evaluation of different mixes, which is a controlled-strain flexural beam fatigue test performed at 20°C, 10 Hz and 200με on specimens pre-saturated under 635 mm-Hg vacuum for 30 minutes and preconditioned at 60°C for one day. An extension of the test procedure for use in the pavement design was also discussed. The major findings of this chapter are summarized as follows:

1. Conditioning temperature significantly affects the moisture resistance of asphalt mixes. High temperature significantly promotes moisture damage in mixes, especially in untreated mixes. On the other hand, moisture content and conditioning duration have less effect on the extent of moisture damage in the fatigue test.

2. The typical fatigue beam test procedure determined in this study can distinguish mixes with different moisture sensitivities, and give a ranking of mixes consistent with prior field experience.

3. The fatigue based test procedure can be applied in pavement design to explicitly include the moisture effect. However, a thorough study of the fatigue response at the typical spectra of conditioning and test parameters should be conducted, and extensive field performance data need to be collected for test result calibration before this procedure can be actually implemented.

ACKNOWLEDGMENT AND DISCLAIMER

This work was undertaken with funding from the California Partnered Pavement Research Program for the California Department of Transportation (Caltrans). This funding is greatly appreciated. The opinions and conclusions expressed in this paper are those of the authors and do not necessarily represent those of Caltrans.

REFERENCES

Lu, Q., and Harvey, J. T. (2006a). "Evaluation of Moisture Sensitivity of Hot Mix Asphalt by Flexural Beam Fatigue Test." Asphalt Concrete: Simulation, Modeling, and Experimental Characterization. Geotechnical Special Publication No. 146. American Society of Civil Engineers, pp. 124-133.

Lu, Q., and Harvey, J. T. (2006b). "Laboratory Evaluation of Long-term Effectiveness of Antistripping Additives." CD-ROM, TRB, National Research Council, Washington, D.C.

Miner, M. A. (1945). "Cumulative Damage in Fatigue." Transactions, American Society of Mechanical Engineers, Vol. 67, pp. A159-A164.

Monismith, C. L., Epps,J. A., and Finn, F. N. (1985). "Improved Asphalt Mix Design." Journal of Association of Asphalt Paving Technologists, Vol. 54, pp. 347-406.

Richards, I. J., and Gabrawy, T. (2003). "Fatigue Testing of Asphalt to Improve the Discrimination of Moisture Sensitivity in Aggregate and Bitumen Systems." In Proceedings of 21st ARRB Conference, Cairns, Australia.

Tayebali, A. A., Deacon, J. A., Coplantz, J. S., Harvey, J. T., and Monismith, C. L. (1994). "Fatigue Response of Asphalt-Aggregate Mixes." SHRP-A-404, Asphalt Research Program, Institute of Transportation Studies, University of California, Berkeley.

Determination of SMA Slab Properties Using a Newly Developed Roller Compactor (Turamesin)

Muniandy, R.[1], Jakarni, F.M.[1], Hassim, S.[1], Selim A.[2]

1 Department of Civil Engineering, Faculty of Engineering,
Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

2 Department of Civil & Environmental Engineering
South Dakota State University, Brookings, SD 57007, USA

Abstract: Compaction of asphalt mixtures in the laboratory is critical to meaningful results. Over the years several compactors were development to closely simulate field compaction. However, none of the devices or equipment was found to effectively compact asphalt mixtures. Therefore, a new automatic roller compactor (Turamesin) was developed which is able to compact any asphalt mixtures to the required slab thickness and density. The machine was developed by researchers from the Universiti Putra Malaysia under the Ministry of Science, Technology and Innovation. In this study Stone Mastic Asphalt (SMA) mixture was selected. The study was carried out in order to establish criteria for slab compaction and explore the potential of the newly developed Turamesin for use in SuperPave mix designs. A total of nine slabs were prepared with different combination of compactive efforts. Data was collected and analyzed to determine the appropriate values of the compactive efforts that results in optimum performance of the asphalt mixture. It was observed that to compact slabs of asphalt mixture to a design air void of 4%, a compactive effort of applied pressure of 8.0 kgf/cm², and 75 passes of roller were required.

Key words: Compaction, Field simulation, Slab, Air voids, Roller compactor

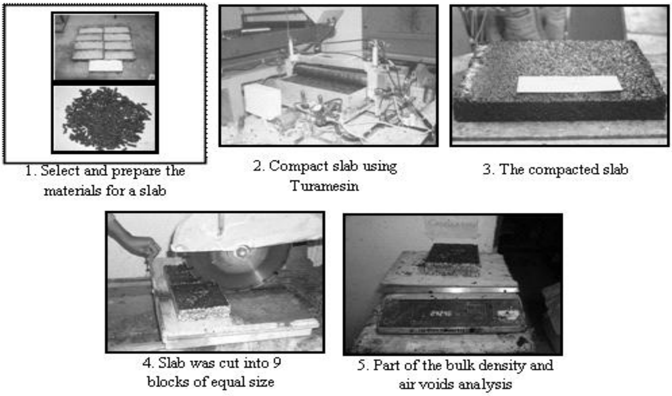
INTRODUCTION

Past studies have shown that if the field conditions for which mixtures are being designed are different from those for which the design method was developed, the mixtures may not be adequate for service even though it is designed according to that particular method [3, 5].Therefore, it is important that the laboratory compaction procedure of asphalt mixtures should be able to achieve mix properties such as density, air voids and resilient modulus as closely as possible to those of the materials placed in the field using standard compaction practices. Over the years, several laboratory compaction methods have been developed to simulate field compaction. These include impact, kneading vibratory and rolling compaction. Also other studies have been done on comparative evaluation of the various laboratory compaction methods based on their ability to simulate field compaction. Conclusions of different studies have indicated that rolling wheel compaction simulate properties that are closer to field compaction [4, 6, 9].

Although several compacting equipments were developed over the years, rolling wheel compactor is considered to be one the one that closely simulate field compaction and appeared to duplicate field compacted specimens quite well [9]. Unlike gyratory and kneading methods, rolling wheel compactor does not include any static leveling loading that might increase particle to particle contact by crushing aggregates together [5] However, the rolling wheel compactor is not widely used as a standard laboratory compaction equipment due to difficulties in controlling air voids in the finished specimens than other compaction method. In addition, the procedures for preparing the laboratory specimens is quite tedious and large quantities of aggregates-asphalt mixtures need to be prepared in order to construct several asphalt mix slabs. The currently available rolling wheel compaction devices are found to be expensive, bulky in size, and not easily portable. This has caused the researchers to look into a more simplified version of compactor such as gyratory compactor, although rolling wheel compactor is intuitively appealing for its obvious similarity to field compaction process[6]. Therefore, researchers at Universiti Putra Malaysia developed a new roller compactor called Turamesin to provide a solution to the problems cited earlier in this paper. Turamesin is used to compact asphalt mixtures using a steel wheel roller just like the heavy duty steel wheel roller used in actual projects. The compactor also provides a variable slope from 0° to 20° from horizontal plane for the purpose of skid resistance analysis and hydroplaning studies. Different levels of pressure can be applied up to approximately 10.0 kgf/cm^2 (143 psi) through a pneumatic system. The objective of this study was to establish a criteria for Turamesin in compacting SMA mix slabs. Two variable parameters of Turamesin were identified as being significantly important that affects the properties and performance of the compacted SMA slabs. These variables are applied pressure and the number of passes of the roller. A total of nine slabs were prepared in accordance with Method of Specimen Preparation of Asphalt Mix Slab Using Turamesin with different combinations of applied pressure and number of passes of the roller compactor. Table 1 provides the experimental matrix for the various combinations of the compactive efforts. Each compacted slab was then cut into nine blocks of equal sizes (166 mm × 150 mm × 70 mm) , and analysis of bulk density and air voids were performed. The results were then analyzed to develop a correlation between the compactive efforts and properties of the compacted slab. The overall test procedures involved are shown in Figure 1.

Table 1: Experimental Matrix for Combination of Compactive Efforts

Number of Passes	Applied Pressure (kgf/cm²)		
	4.0	6.0	8.0
20	Slab 1	Slab 4	Slab 7
40	Slab 2	Slab 5	Slab 8
60	Slab 3	Slab 6	Slab 9



1. Select and prepare the materials for a slab
2. Compact slab using Turamesin
3. The compacted slab
4. Slab was cut into 9 blocks of equal size
5. Part of the bulk density and air voids analysis

Figure 1: Test Procedures Involved

MATERIALS AND METHODS

In this study, 14 mm Stone Mastic Asphalt (SMA) mixtures with granite aggregates and modified rubberized asphalt was used and the determination of the optimum asphalt content was done through Marshall Mix design method in accordance with ASTM D1559-82. The properties of the aggregates and asphalt used in this study were tested in accordance with ASTM, AASHTO and BS standards. All the test results conformed to the specification requirements. The optimum asphalt content value from the Marshall Mix design analysis and the Maximum Density based on Rice Method (ASTM D2041) for the respective value were found to be 5.77% and 2.4200 g/cm^3 respectively. Once the optimum asphalt content value was obtained, the amount of materials required for each slab was calculated using the volume-density calculations with target air voids of 4%.

RESULTS AND DISCUSSION

The values of bulk density and air voids for the specified pressure and number of passes of the roller are given in Table 2. The number of passes - density relationship is shown in Figure 2. It was observed that bulk density and air voids increased as the applied pressure and/or number of passes of the roller compactor increased respectively. This increase was expected since increasing compactive effort and the number of passes forced the mixture particles closer together. All the tabulated data showed a positive quadratic curvilinear relationship at which the bulk density increased as the number of passes of the roller compactor increased for each specified pressure, but this increase tapered off beyond approximately a number of passes of 80. Beyond this point, the bulk density started to experience a slow rate of increase and expected to reach a steady-state condition, which can be defined as ultimate density. Continued rolling beyond this point is wasteful and can even be detrimental to pavement finish, in some cases.

Table 2: Results for Bulk Density and Air Voids Analysis

Applied Pressure (kgf/cm^2)	Number of Passes	Average Bulk Density (g/cm^3)	TMD (g/cm^3)	Air Voids (%)
4.0	20	2.2601	2.4200	6.61
4.0	40	2.2900	2.4200	5.37
4.0	60	2.299	2.4200	4.96
6.0	20	2.2802	2.4200	5.79
6.0	40	2.2999	2.4200	4.96
6.0	60	2.3100	2.4200	4.55
8.0	20	2.2999	2.4200	4.96
8.0	40	2.3101	2.4200	4.55
8.0	60	2.3200	2.4200	4.13

The bulk density and air voids are directly related to each other, thus a closely controlled density is required to ensure air voids stay within an acceptable range [1]. Research and past performance have shown that a final compacted void content of 4% is ideal for asphalt mixtures. Therefore, air voids obtained during mix design or from laboratory compacted specimen should be an estimate of the final compacted void content of 4%. For this reason, 4% was selected as the target air voids in order to establish a criteria for SMA slab compaction using Turamesin. Based on Figure 2, the desired air voids was achieved when 8.0 kgf/cm^2 of pressure was applied with a number of passes between 70 and 80. When the applied pressure was reduced to 6.0 kgf/cm^2, the same value of air voids may be achieved with higher numbers of passes of the roller compactor. Increasing the applied pressure above 8.0 kgf/cm^2 seems to be impossible due to the maximum limit of the pressure gauge of 10 kgf/cm^2. Therefore, the optimum value for applied pressure and number of passes that resulted the asphalt mix slab the closest properties as that of in-service pavement with 4% air voids, are suggested to be 8.0 kgf/cm^2 and 75 respectively.

Upon completion of this study, one of the most important observations that were made with regards to Turamesin is the compaction time. Turamesin seemed to be capable in fabricating slab with 75 roller passes in about 15 minutes, which can be considered very reasonable time compared to other roller compactors. The rolling speed of the roller compactor is about 10.5 rotations per minutes and one complete pass takes about 12 seconds. A slab compacted using Turamesin can produce up to 16 cylindrical core specimens of 100 mm diameter. When comparing the performance in terms of number of specimens produced over time, Turamesin seemed to be well ahead of the other types of compactors.

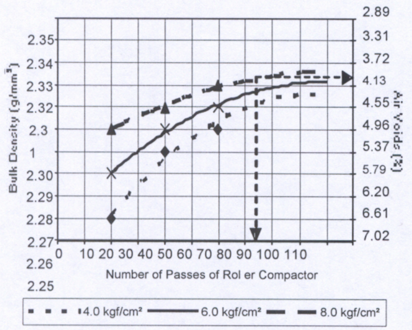


Fig. 2: Correlation between Compactive Efforts and Physical Properties of the Compacted Slabs

CONCLUSIONS

Based on the test results and analysis, it was found that for a slab to be compacted to the desired ultimate compaction with a 4% air void, the applied pressure of 8.0 kgf/cm^2 and 75 passes are required. The target air voids was selected based on researches and past performance that have shown that a final compacted void content of 4% is ideal for asphalt mixtures. Turamesin seemed to be capable in compacting slab within duration of 15 minutes for 75 passes, excluding mixing and coring time. However the time is expected to be reduced to about 8-10 minutes with the upgraded new model. A slab is capable of producing 16 cylindrical core specimens of 100 mm diameter.

Currently, the optimum asphalt content for slab preparation was determined through Marshall mix design method, using Marshall Impact Compactor for specimen fabrication. As Marshall Impact Compactor was found to be least effective in terms field compaction simulation, Turamesin is therefore recommended to be use for specimen preparation in determining the optimum asphalt content. In order to achieve the objective, it is recommended that the length of the mould of Turamesin be adjustable to cater for smaller slab dimension. This may be proven useful in SuperPave mix designs

REFERENCES

1. Brown, E.R., 1990. Density of Asphalt Concrete-How Much Is Needed? Research Report-NCAT Report No. 90-3, NCAT Auburn University
2. Button, J.W., D.W. Little, V. Jagadam and O.J. Pendleton, 1992. Correlation of Selected Laboratory Compaction Methods with Field Compaction. Texas Transportation Institute.
3. Consuerga, A., D.N. Little, H.V. Quintus and J. Burati, 1992. Comparative Evaluation of Laboratory Compaction Devices Based on Their Ability to Produce

Mixtures with Engineering Properties Similar to Those Produced in The Field. Transportation Research Record, 1228: 80-87.
4. Harvey, J., 1992. Effects of Laboratory Specimen Preparation on Aggregate-Asphalt Structure, Air Void Content Measurement and Repetitive Simple Shear Test Results. Transportation Research Record, 1454: 113-122.
5. Khan, Z.A., H.I. Wahab, I. Asi and R. Ramadhan, 1998. Comparative Study of Asphalt Concrete Laboratory Compaction Methods to Simulate Field Compaction. Construction and Building Materials, 12: 373-384.
6. Khosla, N.P. and S. Sadasivam, 2002. Evaluation of the Effects of Mixture Properties and Compaction Methods on the Predicted Performance of Superpave Mixtures. Final Report FHWA/NC/2002-030, NorthCarolina State University.
7. Muniandy, R. and M.F. Zaharudin, 2004. Evaluation of Turamesin for SMA Slab Compaction. Thesis, Universiti Putra Malaysia, Malaysia.
8. Roberts, F.L., P.S. Kandhal, E.R. Brown, D.Y. Lee and T.W. Kennedy, 1996. Hot Mix Asphalt Materials, Mixture, Design and Construction (2nd Ed.). NAPA education Foundation.
9. Sousa, J.B., J.A. Deacon and C.L. Monosmith, 1991. Effect of the Laboratory Compaction Methods on Permanent Deformation Characteristics of Asphalt Aggregate Mixtures. Proceedings Association of Asphalt Paving Technologist, 60: 533-585.

Monitoring and Evaluating Performance Requirements of Flexible Road Pavements

M. Losa[1], P. Leandri[2] and R. Bacci[2]

University of Pisa, Department of Civil Engineering, Via Diotisalvi, 2, 56126 Pisa, Italy;
[1] Associate Professor, PH +39 50 2217736; FAX +39 50 2217762; e-mail: losa@ing.unipi.it;
[2]PhD Students, PH +39 50 2217740; e-mail: pietro.leandri@ing.unipi.it; renato.bacci@ing.unipi.it

ABSTRACT: This paper presents key findings of a research project which is entrusted with setting up guidelines for monitoring and evaluation of pavement performance requirements with the aim of improving durability and both road safety and environmental sustainability.
The study has investigated some techniques for monitoring structural and surface characteristics of asphalt pavements, such as Ground Penetrating Radar, Falling Weight Deflectometer, Laser Profilometer, Skiddometer, and has addressed problems arising from the use of these techniques; the paper describes some measures needed to overwhelm them and for optimizing data collection procedures. In order to enhance the evaluation of asphalt pavements, the study has focused on data analysis and interpretation methodologies; it proposes refinements to the existing procedures for both data processing and interpretation.

INTRODUCTION

Monitoring of the physical condition of pavement assets is conducted both by the collection of field inspection data, in-service testing data, traffic history data, and by recording such data in a useful form.
Evaluation is the process of analyzing, interpreting, and assessing the collected data to determine its current condition state and rate of deterioration. The evaluation encompasses monitoring, but involves a judgment or determination of the meaning of the information collected.
The approach for pavement condition monitoring and evaluation depends upon the « definition » of pavement failure with respect to the user perception, the functions of safety, and the structural integrity and physical condition. The performance indicators based on these approaches of pavement evaluation dictate the calculation of pavement life estimates and M,R&R needs. The evaluation of pavements can involve one or more of the following: structural capacity, physical deterioration or distress, user –related factors (such as riding comfort), safety and appearance, and user-related costs and benefits associated with varying serviceability and with various rehabilitation measures.

DESCRIPTION OF THE EXPERIMENTAL PROGRAM

The "Leopoldo" research project (2004) was targeted towards the definition of guidelines for the design, construction and maintenance of regional road pavements, with the aim of improving road safety and environmental compatibility.
Ten road infrastructures were selected from within the highway network of the Region of Tuscany, from each of which an experimental stretch was selected for functional and structural evaluation.

FUNCTIONAL EVALUATION

The functional evaluation of a pavement is the effectiveness in fulfilling its intended function of good serviceability shown by « smooth ride » without bumps and high noise level. Major concerns are safety, sufficiency based on capacity and demand, serviceability or quality of service, noise pollution, and physical appearance. The use of objective measurements is preferred because of their repeatability and higher productivity. Where possible, the condition or performance indices should be based on objective measurements of selected performance indicators and on analysis of maintenance records.
In this project, the functional evaluation of pavements was carried out with regard to serviceability, safety and noise, but this work deals only with safety concerns.

Safety
Safety was evaluated by the skid resistance, determined by friction and texture measurements.
Skid resistance was determined by friction measurements performed both by the pendulum tester and by the fixed slip device Skiddometer BV 11; such measurements were integrated by macrotexture measurements, performed by both the sand patch method and a laser Profilometer (Losa et alii 2006, 2007.b).
The statistical analysis of data requires the evaluation of repeatability both for friction and macrotexture measurements; these values were evaluated at different measurement speeds and the results are reported in the table 1 for the Skiddometer BV11 and for the laser profilometer.

Tab. 1 – Repeatability of friction and macotexture measurements

SKIDDOMETER BV 11						LASER PROFILOMETER	
V = 20 km/h		V = 50 km/h		V = 80 km/h		V = 20 km/h	
σ	r	σ	r	σ	r	σ	r
0.016	**0.045**	0.021	**0.060**	0.027	**0.075**	0.032	**0.09**

The repeatability of measurements is an important issue to evaluate the statistical significance of measurements for the delineation of a section survey into homogeneous sub-sections; the minimum number of measurement points is determined as a function of the standard deviation of friction and macrotexture measurements in order to obtain a 95% confidence interval which is lower than or equal to the repeatability of the specific measurement considered, by using the following relationship:

$$n = \left(\frac{t_{(0.025,n-1)} \cdot \sigma}{a} \right)^2 \qquad (1)$$

where: t is the value at the 97.5^{th} percentile of the distribution function of the Student's T variable with n-1degrees of freedom;
σ is the standard deviation of the n measurement points;
a is the confidence interval equal to the repeatability of measurements.

The minimum number of measurement points is calculated taking into account that better results are obtained by averaging recorded data to characterize each 10 m long sub-section with a mean value of friction and macrotexture; this considering, by using the relation (1) and the values reported in table 1, only one mean value is enough to characterize a homogeneous sub-section from a statistical point of view, then its minimum length will be equal to 10 m.

The repeatability of measurements plays an important role in the analysis of repeated measurements on the same section too; in order to identify invalid readings, a paired test allows to evaluate that the difference between any two values of the parameter on the same section is lower than the repeatability of the measurement with a 95% reliability level.

Analysis of macrotexture data

In this study, a critical analysis of the procedures used in macrotexture profile analysis for the Mean Profile Depth (MPD) calculation and the Mean Texture Depth (MTD) estimation was carried out.

In order to identify a reliable procedure for the evaluation of pavement macrotexture, the effect of different data processing methods of pavement profiles was analyzed by using two types of filters: Butterworth's numerical filters and moving average filters. By means of a careful analysis of texture levels it was possible to identify which data processing method reproduces better the sand patch method for evaluation of texture depth and which procedure gives results more stable and significant both from a statistical point of view and to describe the involved phenomena.
The analysis showed that moving average filters are more suitable to reproduce the sand patch method; in particular better results are obtained by using a high-pass filter with a computation basis equal to 50 or 100 mm and a low-pass filter with a computation basis equal to 2.5 or 5 mm; by using numerical filters, we cannot reproduce reliably the sand patch method and the results obtained by using the two alternatives of the ISO Standard 13473-1 are not equivalent. The observations put forward in this study suggest that alternative 1 is preferable.

MTD values were compared to MPD values obtained from the means of values calculated for three different measurement directions (Parallel, perpendicular and at 45° with traffic direction). As can be seen in Figure 1, both for alternative 1 and for alternative 2 two similar relations are obtained, characterized by high coefficients of correlation, namely 0.95 and 0.94 respectively. This confirms that the best estimates of MTD, obtained through MPD measurements, are achieved by using MPD values calculated along more than one direction.

A difference can be noted as compared to the relation ($ETD = 0.8 \cdot MPD + 0.2$) contained in the ISO 13473-1. This difference is due exclusively to the fact that the ISO 13473-1 relation refers to profiles measured only in the direction parallel with forward moving traffic, rather than in several different directions as carried out in the present study. Thus in Figure 2, MTD values were set in relation to MPD values that pertain exclusively to the forwards direction. The following mean relation was obtained, where ETD and MPD are in mm:

$$ETD = 0.84 \cdot MPD_{longitudinal} + 0.2 \quad (2)$$

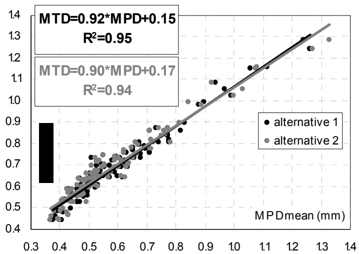


Figure 1 – MTD versus mean MPD from stationary Profilometer

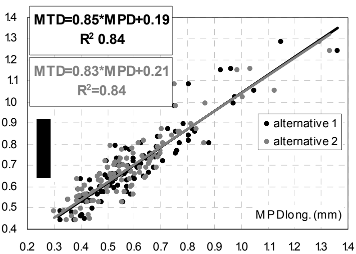


Figure 2 –MTD versus MPD in the forward traveling direction

In this case, the relation between ETD and MPD is analogous to the relation laid down by the ISO 13473-1, since the measurements refer to the same direction of movement. But for both alternatives the coefficients of correlation have decreased to 0.84 in comparison to the previous case, confirming the observation made above regarding the most suitable way of estimating mean texture depth through MPD measurements.

Analysis of friction data

Friction values obtained by both devices, texture and slip speed parameters were related by using the exponential form of the PIARC model (Wambold et alii 1995); in this way, the standard error associated to the harmonization between Skiddometer BV 11 and pendulum tester friction values were calculated with a fixed confidence level. Then such procedure was used in order to define threshold values for both skid resistance and Skiddometer BV 11 friction values.

STRUCTURAL EVALUATION

The structural evaluation is carried out to assess structural integrity and load carrying capacity. Nondestructive testing provides response measurements which are used to infer the information about the physical structure of the pavement behavioral evaluations. Deflection data can be used to establish homogeneous design sections.

Layer investigations were performed on each stretch using GPR (Ground Penetrating Radar) instrumentation; deflectometric tests were performed with Dynatest 8000 FWD equipment. Thickness measurements on the experimental sections were performed with a ground-coupled radar system (equipped with two antennae having a central frequency of 1600 MHz and 600 MHz, capable of functioning both in bistatic and monostatic mode). FWD tests were performed at 10 m center-to-center, 9 load repetitions were performed at each test point, applying 3 different load levels corresponding to 42, 50 and 70 kN, with a 300 mm ∅ segmented plate. Deflections d_i were recorded at the following radial distances i = 0, 200, 300, 450, 600, 900, 1200, 1500 and 1800 mm (Losa et alii 2007.a).

The main difficulties in performing structural analysis of pavements of this type reside in the following aspects: significant variability in layer thickness and mechanical characteristics of the materials; minimal thickness of the asphalt concrete layers, making backanalysis procedures unreliable; modeling of the underlying stone block roadbed and of the subgrade.
In these cases, it may be preferable to perform an estimate of the critical stresses and strains by more stable methods that are less sensitive to thickness variations.
In this project, a detailed structural analysis of some existing pavements was carried out, in order to characterize spatial variability of pavement materials and thickness as well as to identify both the deflection basin parameters and pavement variables that most directly influence the critical design parameters used in determination of pavement residual service life. Relationships between critical redesign parameters and FWD deflections for a 50 kN load level were investigated in order to assess structural capacity of pavements directly from the FWD deflection basin data and layer thicknesses of asphalt concrete (h_1) and subbase (h_2) without backcalculating the layer stiffness moduli.

On the basis of these results, the statistical model has the following form:
ε_{AC} ($\mu\varepsilon$) horizontal strain at the bottom of the asphalt concrete layer:

$$\log\varepsilon_{AC} = 0.387\cdot \log h_1 + 0.108\cdot \log h_2 - 0.242\cdot \log d_{900} + 0.080\cdot \log d_{1800} + 0.446\cdot \log SCI + 0.735 \cdot \log BDI - 0.869 \quad [3]$$

with R^2 = 0.972 and SEE = 0.039

ε_{SB} ($\mu\varepsilon$) vertical strain at top of the subbase layer:

$$\log\varepsilon_{SB} = 0.103\cdot \log h_1 + 0.185\cdot \log h_2 + 1.443\cdot \log d_0 - 1.264\cdot \log d_{300} + 0.883\cdot \log BDI + -0.367 \quad [4]$$

ε_{SB} ($\mu\varepsilon$) vertical strain at top of the subbase layer:

$$\log\varepsilon_{SG} = -1.060\cdot \log(h_1+h_2)+1.045\cdot \log d_0 + 0.178\cdot \log d_{900} - 0.183\cdot \log d_{1800}+2.663 \quad [5]$$

with R^2= 0.950 and SEE = 0.040

CONCLUSIONS

The analysis of experimental data has allowed to check and to improve some existing procedures for functional and structural evaluation of existing pavements.
As far as functional evaluation, skid resistance of pavements was evaluated by friction and texture measurements; a statistical analysis of data has allowed to calculate the repeatibility of such measurements, which is required to delineate a section survey into homogeneous sub-sections and to check the reliability of measurements; moreover, the coefficients of PIARC model have been evaluated for the equipments used in the experimental program in order to define skid resistance requirements in terms of the International Friction Index (IFI).
As far as structural evaluation, strain estimates provide results with dispersion lower than or equal to the dispersion obtained by evaluating the elastic moduli through the usual backcalculation procedures. Consequently, after estimating the elastic moduli and calculating the corresponding strains by a very accurate structural analysis, relations were formulated that were capable of determining the strains acting on the pavement directly from the measured deflection basin and from layer thicknesses. This gave a result characterized by a less dispersed estimate, without the need for time-consuming modulus backcalculation procedures.
The results are very useful for the particular type of investigated pavements because they allow to identify, for each layer, homogeneous sub-sections quickly on the basis of deflection parameters and to evaluate strains by a harmonized procedure.

ESSENTIAL REFERENCES

Leopoldo project web site <http:\\www.ing.unipi.it/leopoldo>

Losa, M., Bacci, R., Leandri P. (2006). “Skid Resistance of Pavement Surfaces from Fixed Slip Device Measurements”, Proceedings of SIIV National Congress, Cosenza, Italy.

Losa, M., Marvogli, M., Leandri P., Bacci, R. (2007.a). “Structural Analysis of Asphalt Pavements on Collector Rural Roads” – Proceedings of the 4th International Conference on Bituminous Mixtures and Pavements, Thessaloniki, Greece.

Losa, M., Leandri P., Bacci, R. (2007.b). “Measurements of Pavement Macrotexture with Stationary and Mobile Profilometers”, Proceedings of MAIREPAV5, Park City, Utah, USA.

Wambold, J. C., Antle, C. E., Henry, J. J ., Rado, Z. (1995). “International PIARC Experiment to Compare and Harmonize Texture and Skid Resistance Measurements” – PIARC, Paris.

Evaluation of Tennessee HMA Mixtures Using Simple Performance Tests

Xiang Shu[1], and Baoshan Huang[2]

Abstract

Hot mix asphalt (HMA) covers more than ninety percent of paved roadways in the US. Most HMA pavement distresses (such as rutting and fatigue cracking) can be characterized through laboratory performance tests. The NCHRP Project 9-19 has identified several tests as the simple performance tests (SPT) to characterize the performance of HMA mixtures. The paper presents a laboratory study carried out by the Tennessee Department of Transportation and the University of Tennessee. One type of HMA commonly used as surface mixture in the state of Tennessee, USA, was evaluated through the simple performance tests of dynamic modulus and flow number. The Asphalt Pavement Analyzer (APA) test was conducted as comparison to the simple performance testing. The dynamic modulus was tested at 10, 25, and 54 °C and three confining pressures. Three asphalt binders (PG 64-22, PG 70-22, and PG 76-22) were used for comparative purposes. The results from the present study indicated that the simple performance tests generally agreed with APA, i.e., mixtures produced with increased high temperature PG grade of asphalt cement exhibited higher dynamic modulus values, higher flow number values, lower rut depths, indicating potentially high resistance to rutting.

Introduction

The Superpave mix design and analysis method was developed under the Strategic Highway Research (SHRP) more than fifteen years ago and now have been partially implemented by many highway agencies in North America with successful results,

[1] Dept. of Civil & Envir. Engineering, Univ. of Tennessee, Knoxville, TN 37966, USA; PH (1-865)974-2608; FAX (1-865)974-2669; email: xshu@utk.edu
[2] Dept. of Civil & Envir. Engineering, Univ. of Tennessee, Knoxville, TN 37966, USA; PH (1-865)974-7713; FAX (1-865)974-2608; email: bhuang@utk.edu

especially the performance-grade (PG) binder specification and the volumetric mixture design procedure (Leahy et al., 1994). However, unlike the Marshall and Hveem mixture design methods, the Superpave mixture design procedure does not have a strength test to verify the suitability of the mixtures. In response to the need for a simple test to complement the Superpave mix design method, the National Cooperative Highway Research Program (NCHRP) Project 9-19 "Superpave Support and Performance Models Management" has identified three simple performance tests for hot-mix asphalt (HMA) mixtures (Witczak et al., 2002). The three tests are the dynamic modulus test, the flow number test, and the flow time test. From these three tests, dynamic modulus, flow number, and flow time can be obtained, respectively, and used to characterize the performance of HMA, mainly for rutting and fatigue cracking.

Dynamic modulus is one of the fundamental engineering properties widely used to characterize the viscoelastic behavior of asphalt mixtures and has also been selected as a basic material property input in the new AASHTO mechanistic-empirical (M-E) design guide (Pellinen, 2001; Andrei et al., 2004). With the adoption of dynamic modulus test as one of the simple performance tests, many researchers have recently examined the effects of sample preparation method, strain level, testing protocol, and testing history on the dynamic modulus values of asphalt mixtures (Robinette and Williams, 2006; Tran and Hall, 2006). The flow number test is very useful in evaluating the rutting resistance of HMA mixture by characterizing evaluating its creep characteristics.

This paper presents the findings of a study conducted by the Tennessee Department of Transportation and the University of Tennessee to investigate the simple performance characteristics of plant-produced asphalt mixtures.

Laboratory Experiments

Materials. Totally, there were three HMA mixtures used in this study (Table 1). All the mixtures were plant-produced field mixes from an array of locations within the state of Tennessee. Mixtures were collected from dump trucks in asphalt plants and transported to the University of Tennessee for testing. Once acquired, mixtures were only allowed to be reheated once for specimen compaction to avoid stiffening of the mixture. The mixtures meet the Tennessee Department of Transportation (TDOT) specifications for 411-D surface mixtures (TDOT, 2006).

Table 1. Laboratory Test Factorials

Mixtures			Mixture Performance Tests		
Mixture Type	Coarse Aggregate	Asphalt Cement	Dynamic Modulus	Flow Number	APA
D	Limestone	PG 64-22	9	3	3
		PG 70-22	9	3	3
		PG 76-22	9	3	3

Note: The number in the cell represents the number of tested samples.

Limestone was used as coarse aggregate for evaluation of HMA mixtures with a nominal maximum aggregate size (NMAS) of 12.5 mm for all mixtures. The fine aggregate used in the surface mixes consisted of natural sand and #10 screenings.

Three types of asphalt binder were used in the surface mixtures: PG64-22, PG70-22, and PG76-22. Asphalt contents of all HMA mixtures were verified to be approximately 5.3%.

Sample Preparation. To ensure the quality of each specimen prepared for testing, great care was given to maintain a consistent compaction process. The collected plant asphalt mixtures were reheated in a force draft oven at 160ºC for 2 hours before compaction. The Superpave Gyratory Compactor (SGC) was used to compact the mixtures into cylindrical specimens 150 mm in diameter and 170 mm in height. All specimens were compacted to 4 ± 1% air voids. Once the specimens were compacted and allowed to cool they were then cored into specimens 100 mm in diameter. A wet blade saw was then used to cut the ends of each specimen to a final height of 150 mm. Sawing operations were performed carefully to ensure the ends maintained absolute parallelism.

Prior to testing, all samples were checked for air voids in accordance with AASHTO T-269 " Percent Air Voids in Compacted Dense and Open Bituminous Paving Mixtures" to validate proper air void requirements.

Dynamic Modulus Test. The dynamic modulus test is a strain-controlled test performed as a 100 mm diameter, and 150 mm tall cored cylindrical specimen is subjected to a continuous haversine axial compressive load. In this study, the test was performed at loading frequencies ranging from 25 to 0.1 Hz and at three temperatures (10, 25, and 54 ºC) and three confining pressures (0, 103.5, and 207 kPa).

The dynamic modulus, $|E^*|$, is defined as the ratio of the amplitude of the sinusoidal stress, σ_0, applied to the specimen and the amplitude of the induced sinusoidal stain, ε_0:

$$|E^*| = \frac{\sigma_o}{\varepsilon_o} \tag{1}$$

Flow Number Test. The flow number test is a repeated-load test used to evaluate the creep characteristics of HMA as related to permanent deformation. The test was performed by applying a compressive load in haversine form with a loading time of 0.1 seconds and a rest duration of 0.9 seconds to a specimen at 54 °C and without confining pressures.

Asphalt Pavement Analyzer (APA) Rut Test. The APA is an empirical wheel loaded device producing pavement distress by continuously loading 6 identical 150 mm diameter and 75 mm tall specimens with an inflatable hose and roller. The APA rut test was conducted according to the AASHTO TP63-03.

Results and Discussion

Dynamic Modulus Test Results. Fig. 1 shows the typical dynamic modulus test results of HMA mixtures. It is observed that as the temperature decreased or the loading frequency increased, the dynamic modulus value increased for asphalt mixtures. This temperature and loading time-dependency of dynamic modulus is due to the viscoelastic constituent in asphalt mixture – asphalt cement. In general, the dynamic modulus test results from this study were highly repeatable with most coefficients of variation below 20% and several exceptions at higher temperatures.

Fig. 2 shows the effect of asphalt cement PG grade on dynamic modulus value of HMA mixtures. It can be seen that the dynamic modulus value increased as the asphalt cement PG grade progressed from PG 64-22 to PG 70-22 to PG 76-22. This indicated that asphalt cements with higher temperature PG grade are more resistant to permanent deformation. The verified asphalt contents for mixtures with all three binders were 5.3%.

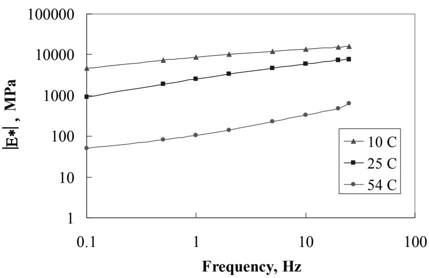


Fig. 1 Effects of temperature and loading frequency on |E*| (PG 64-22)

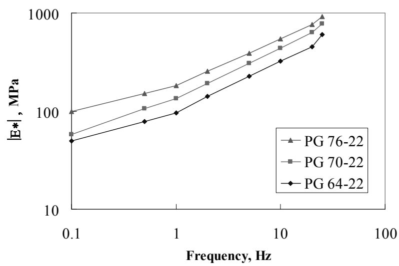
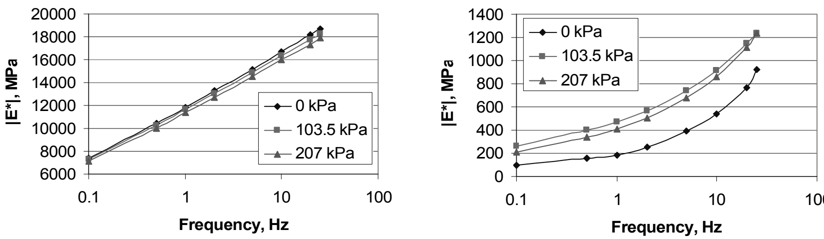


Fig. 2 Effect of asphalt binder (54 °C)

The effect of confining pressure on dynamic modulus value is shown in Fig. 3. It should be noted that the effect of confining pressure on dynamic modulus is influenced by many factors, such as asphalt cement (including PG grade and asphalt content), aggregate (including aggregate angularity, surface texture), and temperature, etc. Fig. 3 shows the effect of confining pressure on the dynamic modulus value of HMA mixtures containing asphalt binder PG 76-22. At low temperature (10 ºC), aggregate particles were bonded by asphalt cement tightly, the effect of confining pressure on dynamic modulus was observed to be minimal (Fig. 3a). With the increase in the temperature, asphalt cement became softer and the effect of aggregate on dynamic modulus became more and more significant. In the confined state, aggregate particles were restricted by the confining pressure, which resulted in the increased dynamic modulus value (Fig. 3b). If aggregate particles were less angular and more rounded, the effect of confining pressure was more significant in improving dynamic modulus values of mixtures. In other words, the effect of confining pressure was dependent on the role of aggregate structure in asphalt mixtures. If aggregate structure plays an important role in dynamic modulus and confining pressure helps to strengthen aggregate structure, the dynamic modulus value could be significantly improved by applying confining pressure.



(a) 10 °C (b) 54 °C

Fig. 3 Effect of confining pressure on |E*|

Flow Number Test Results. Fig. 4 presents the average flow number and failure strain results for asphalt mixtures. It is observed that when the high temperature of asphalt cement PG grade increased from 64 to 76, the flow number results increased significantly. This indicated that flow number results are sensitive to asphalt cement. Increasing the high temperature of asphalt cement grade could significantly improve the flow number value of mixture specimens, and hence the rutting resistance of HMA mixtures

APA Rut Test Results. Fig. 5 presents the average APA rut depth results for TDOT surface mixtures. It can be seen from Fig. 5 that the APA test was also sensitive to the change in asphalt cement. Use of increased high-temperature PG grade improved the rutting resistance of asphalt mixtures, if same aggregate is used in the mixtures.

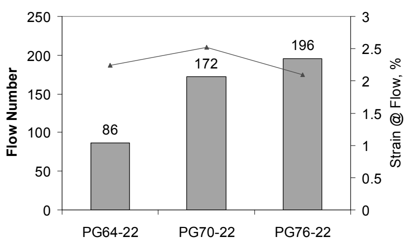


Fig. 4 Flow number test results

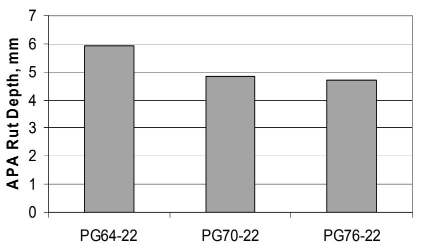


Fig. 5 APA rut test results

Summary and Conclusions

A laboratory study was conducted to investigate the simple performance characteristics of HMA mixtures used in the state of Tennessee, USA. Based on test results and discussion, the following conclusions can be summarized.

- The dynamic modulus test results were highly repeatable with several exceptions at higher temperatures for the mixtures used in this study.
- The effect of confining pressure on dynamic modulus was mobilized through aggregate structure. Any factors that can influence the role of aggregate in mixture rutting resistance could also affect the effect of confining pressure on dynamic modulus, such as temperature and aggregate characteristics.
- The simple performance tests and the APA test could produce consistent results, i.e., mixtures produced increased high temperature PG grade of asphalt cement exhibited higher dynamic modulus values, higher flow number values, lower rut depths, indicating potentially high resistance to rutting.
- Both dynamic modulus and flow number tests were sensitive to asphalt cement in this study.

References

Andrei, D., Witczak, M. W. , and Mirza, M. W. (2004). Development of a Revised Predictive Model for the Dynamic (Complex) Modulus of Asphalt Mixtures. In Guide for Mechanistic-Empirical Design of New and Reha-bilitated Pavement Structures, NCHRP Project 1-37A, Final Report, Appendix CC-4. Transportation Research Board of the National Academies, Washington, DC. www.trb.org/mepdg/guide.htm. Accessed July 15, 2007.

Leahy, R.B., Harrigan, E.T., and Von Quintus, H. (1994). Validation of Relationships Between Specification Properties and Performance. Report No. SHRP-A-409. Transportation Research Board, National Research Council, Washington, DC.

Pellinen, T. K. (2001). Investigation of the Use of Dynamic Modulus as an Indicator of Hot Mix Asphalt Performance. Ph.D. dissertation. Arizona State Univeristy, Tempe, Arizona.

Robinette, C. and Williams, R. C. (2006). "The Effects of the Testing History and Preparation Method on the Superpave Simple Performance Test," Journal of the Association of Asphalt Paving Technologists, Vol. 75, pp. 297–320.

TDOT (2006). Tennessee Standard Specification for Road and Bridge Construction. Tennessee Department of Transportation, Nashville, 2006 edition.

Tran, N. H. and Hall, K. D. (2006). "An Examination of Strain Levels Used in the Dynamic Modulus Testing," Journal of the Association of Asphalt Paving Technologists, Vol. 75, pp. 321–344.

Witczak, M. W., Kaloush, K., Pellinen, T., and El-Basyouny, M (2002). NCHRP Report 465: Simple Performance Test for Superpave Mix Design. TRB, National Research Council, Washington, DC.

Accelerated Pavement Testing on Thin Asphalt Pavements with Various Base and Subbase Layers

Zhong Wu[1], Zhongjie Zhang[2], Mark Morvant[3], Harold Paul[4] and Bill Temple[5]

[1]Louisiana Transportation Research Center, 4101 Gourrier Ave, Baton Rouge, LA 70808, E-mail: zhongwu@ltrc.lsu.edu, PH: (225) 767-9163, Fax: (225) 767-9108
[2]Louisiana Transportation Research Center, E-mail: DocZhang@dotd.louisiana.gov, PH: (225) 767-9162, Fax: (225) 767-9108
[3]Louisiana Transportation Research Center, E-mail: MarkMorvant@dotd.la.gov, PH: (225) 767-9124, Fax: (225) 767-9108.
[4]Louisiana Transportation Research Center, E-mail: HaroldPaul@dotd.la.gov, PH: (225) 767-9131, Fax: (225) 767-9108
[5]Louisiana Department of Transportation and Development, 1201 Capitol Access Road, Baton Rouge, Louisiana 70804-9245. E-mail: BillTemple@dotd.la.gov

Abstract

Six thin asphalt pavements were tested under accelerated pavement testing. Those pavements generally varied in the base and subbase materials, but had the same layer thicknesses. Pressure cells and Multi-Depth Deflectometers were used to monitor the vertical stresses and plastic deformations developed inside the pavement structures. Falling Weight Deflectometer (FWD) tests were performed during the accelerated loading. Test results indicated that a test section with a furnace slag stabilized Blended Calcium Sulfate (or BCS, an industry waste material) base performed significantly better than its counterpart section with a fly-ash stabilized BCS base, whereas both BCS sections appeared to out-perform the pavement sections with the foamed-asphalt stabilized bases and the crush limestone bases. In addition, a cement-treated soil subbase was found to have a better load carrying capacity than a lime-treated subbase. Further investigation illustrates that both the FWD data and instrumentation results could reveal the structural deterioration fairly well.

Introduction

The Louisiana Pavement Research Facility (PRF) is an outdoor, full-scale accelerated pavement testing (APT) research laboratory located in Port Allen, Louisiana. It has space for the construction of ten 65-m (215-ft) long by 4-m (13-ft) wide full-scale test pavement lanes. Traffic loading is provided by a machine called the Accelerated Loading Facility (ALF). The ALF wheel assembly models one half of a single axle with dual tires and the load is adjustable from 43.4 kN (9,750 lb) to 84.4 kN (18,950 lb) per load application. With a computer-controlled load trolley, the weight and movement of traffic is simulated repetitively in one direction at a speed of approximately 16.8 km/hr (10.5 mph). Therefore, the PRF provides an ideal APT facility to evaluate the structural performance of various base and subbase materials currently investigated by the Louisiana Department of Transportation and Development (Wu et al., 2005).

Objective

The objective of this study was to evaluate the structural behaviors of various base and subbase materials under accelerated testing based on non-destructive deflection measurements and instrumentation results.

APT Experimental Design

Six test sections were included in this study. Figure 1 presents a plan view and pavement structure cross sections of this APT study. A 50-mm hot-mix asphalt (HMA) surface layer were used in all six test sections. The sections generally varied in the base and subbase materials, but had the same layer thicknesses. As showen in Figure 1 test sections I and II had the slag stabilized BCS base and the fly-ash stabilized BCS base, respectively, and both sections sit on the top of a 10% lime-treated soil (LTS) subgrade layer. Sections III and VI were two foamed-asphalt (FA) stabilized base test sections, both having an 8% cement-treated soil (CTS) subgrade layer. Sections IV and V both had a same crushed limestone base course. Section IV had a 10% LTS subgrade layer, whereas, Section V used an 8% CTS subgrade layer. The detail thickness information can be referred to Figure 1.

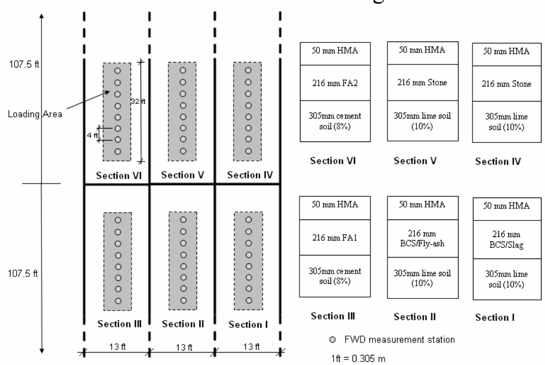


Figure 1. Test Section Layouts

The HMA mixture used was a 19-mm coarse-graded Superpave mixture, designed for intermediate volume roads in Louisiana. The crushed limestone bases in sections IV and V met the Louisiana's Class-II base course specification. The mix design of the two stabilized BCS materials was based on a laboratory study (Zhang and Tao, 2006), where a 10 percent Grade-120 ground granulated blast furnace slag and 15 percent Class-C fly-ash were added in the corresponding BCS bases in sections I and II, respectively. The FA bases on sections III and VI were designed based on the standard Wirtgen Cold Recycling Manual (Wirtgen, 1998). The FA base on section III (hereafter called as FA1) consisted of 48.6 percent reclaimed asphalt pavement (RAP), 48.6 percent recycled soil cement and 2.8 percent PG 58-22 asphalt binder, whereas, the FA base on Section VI (hereafter called as FA2) contained 97.5 percent RAP and 2.5 percent PG 58-22 binder. A silty-clay embankment soil was used in both LTS and CTS subbases as well as the subgrade. This soil consisted of 60.3 percent silt and 23.5 percent clay. More details of each material can be found elsewhere (Wu et al., 2005).

Discussion of Results

Accelerated Testing Results. An average rut depth of 12.5 mm at the surface and/or 50 percent of the trafficked area with cracking more than 5 m/m^2 was selected as the initial failure criterion. The APT results indicated that four sections - I, II, III and V failed due to pure surface rutting. No visible cracks were observed on those sections when the average rut depth reached to the 12.5 mm limit. Section IV also failed due to rutting, but accompanying with some medium-severe alligator cracks developed in the surface. Section VI was classified as a cracking failure.

"Pavement life" is defined as the number of ALF load repetitions required for a test section to fail. This number can be translated into an ESAL number (Equivalent Single Axle Load of 18,000 lbf) based on the fourth power law (Huang, 1993). In this study, the pavement lives for sections I through VI were found to be 3,914,000-, 2,349,000-, 411,000-, 121,000-, 538,000- and 356,000- ESALs, respectively.

Table 1 presents an "ESAL advantage" comparison for various base and subbase materials studied. The "ESAL advantage" is defined as a ratio of pavement lives between two similar sections. Basically, the "ESAL advantage" implies that one material performs how many times better than another material in terms of pavement lives. For example, by comparing the "lives" of sections I and IV, the BCS/Slag in section I was found to have 32:1 ESAL advantage over the stone base in section IV. Overall, Table 1 indicates that BCS/Slag performed significantly better than its counterpart material of BCS/fly-ash (with an ESAL advantage of 1.7:1). Meanwhile, both stabilized BCS base materials appeared to marginally out-perform the rest base materials (i.e. two FA bases and the stone). The stone base was found to have an ESAL advantage of 1.3: 1 and 1.5:1, respectively, as compared to FA1 and FA2, whereas, FA1 had a 1.2: 1 ESAL advantage over FA2. In addition, the CTB subbase was found to have an ESAL advantage of 4.5: 1 over the LTS subbase.

Table 1. "ESAL advantage" for various base and subbase materials

	BCS/Slag	BCS/Fly	Stone	FA1	FA2	CTB	LTB
BCS/Slag	1:1	1.7→1	32→1	42→1	49→1		
BCS/Fly		1:1	19→1	25→1	29→1		
Stone			1:1	1.3→1	1.5→1		
FA1				1:1	1.2→1		
FA2					1:1		
CTB						1:1	4.5→1
LTB							1:1

Field Instrumentation Results. Two Geokon 3500 pressure cells were installed at a middle station along each section's center line at two depths: the bottom of base layer and the bottom of subbase layer. One multi-depth deflectometer (MDD), called SnapMDD™ (a patented device manufactured by the Construction Technology Laboratories, Inc, Illinois) with six potentiometers was installed on each test section at a station of 1.4-m (4.5-ft) away from the pressure cell station in the center lines.

Figure 2 presents the typically measured vertical compressive stresses at each test section. As expected, stiffer materials tend to possess a larger load distribution angle than less stiff materials, which would result in a lower vertical stress at its bottom. As shown in Figure 2, the vertical compressive stresses at bottoms of various base layer were 6.2-, 35.8-, 65.5-, 235.0-, 130.0- and 62.0- kPa, respectively, for sections I through VI. This implies that the BCS/Slag in section I had a significantly larger load spreading angle than any other base materials in this study. Similarly, the BCS/Flyash base in section II was found to distribute the load better than the FA and stone bases. The larger load distribution angles of the stabilized BCS materials are believed responsible for the longer pavement lives found on sections I and II, as described above. On the other hand, both FA bases seemed to have a larger load distribution angle than the stone base (if comparing the vertical compressive stresses at bottom of base layers on sections III, V and VI). However, both FA sections had a shorter pavement life than section V with the stone base. Further field survey indicated that section III was failed due to large shear flows developed inside the unstable FA1 materials (i.e. weak aggregate skeleton), whereas, section VI was failed due to over-asphalting of FA2 mixture. The above results indicate that only a better load distribution (or a higher compressive stiffness value) can not guaranty a better performance for a FA stabilized material; the combination of stability and sound aggregate structure must be considered in a FA mixture design.

By comparing sections VI and V, a smaller vertical stress was observed at bottom of the stone base on section V than that on section IV. This indicates that the stone layer on section V have a larger stiffness value than that in section IV. This may be explained by the so-called “stress hardening” phenomenon for stone materials. The stiffer CTS subbase on section V provided a stronger support (or higher confining pressure) to the stone base than the LTS subbase on section IV, which resulted in a larger stiffness value for the stone base on section V than that on section IV. Such observation indirectly explained why section V had a long pavement life than section IV.

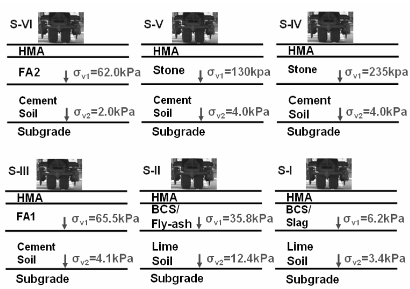


Figure 2. Typical pressure cell measured vertical stresses

The MDD measured vertical permanent displacements were used to calculate the permanent deformation developed in each pavement structural layer in a test section. Figure 3 shows the permanent deformation developed in each pavement structural layer on test sections IV and V. As can be seen in Figure 3, a significant amount of permanent deformation was developed on the stone bases of both sections. However, the permanent deformations developed on the cement-treated subbase of section V were negligible as comparing to the lime-treated base on section IV. This observation implies that the cement-treated subbase had a higher strength and higher load carrying capability than the lime treated subbase layer in this study.

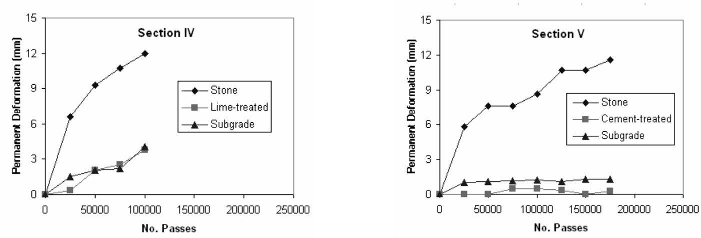


Figure 3. MDD measured permanent deformation at each section

FWD Results. Figure 4 presents the average measured FWD deflections on each test sections before the APT loading. The deflections were normalized to 40 kN (9,000-lb) for comparison. In general, a higher central deflection (d0) value indicates a weaker pavement structure. Similarly, a larger d8-value (deflection measured 1.8 m away from the FWD load) implies a weaker subgrade. As shown in Figure 4, the numerical orders in terms of d0-values (from low to high) are as follows: sections I, VI, II, V, III and IV. However, based on the statistical ranking analysis of the eight FWD deflection bowls measured at each section , the d0-values on section I were found significantly smaller (or pavement stronger) than any other test sections in this study, whereas, that of section IV appeared to be the largest (pavement structure is the weakest). The d0-values on sections II, III, V and VI were found not statistically different. On the other hand, the average d8-values were found statistically similar among various sections, indicating a same natural subgrade used for all test sections.

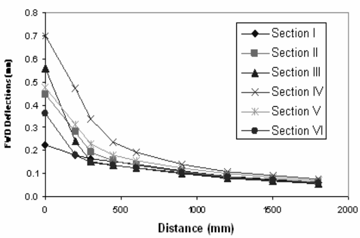


Figure 4. FWD deflections

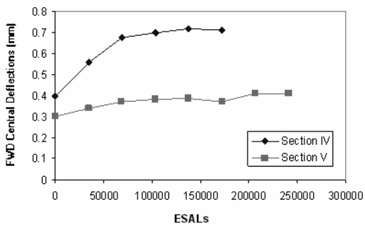


Figure 5. Progression of FWD central deflections under APT loading

EVERCALC backcalculation program (Pierce and Mahoney, 1996) was used in the FWD modulus backcalculation. Due to the thin asphalt surface layer (50-mm), both moduli of asphalt concrete and lime treated layers were set to a fixed value based on engineering judgment and laboratory test results. The backcalculation results are presented in Table 2. The FWD results generally confirmed that the stabilized BCS materials are stiffer than other base materials. The backcalculated modulus of the stone layer on section V was found higher than that on section IV due to a larger confining pressure provided by the CTS layer. Both FA materials seemed stiffer than the stone, at least under such FWD test condition. Further analysis revealed that FWD deflections can be used to monitor the loaded induced pavement deterioration. As load repetitions increase, the measured central deflection appeared to be increased correspondingly. Figure 5 presents the load-induced progression of the central deflections measured on sections IV and V. It is noted that the central deflections were temperature- and environmental- corrected based on FWD deflection results measured at two *unloaded* stations of each test section during each FWD testing cycle.

Table 2. FWD backcalculation results (note- * means fixed modulus value)

	HMA	Backcalculated Modulus (MPa)			RMS Error
		Base	Subbase	Subgrade	(%)
Section I	5,000*	8,410	480*	128	2.1
Section II	5,000*	599	480*	138	3.7
Section III	5,000*	510	2,845	119	3.82
Section IV	5,000*	190	480*	93	3.8
Section V	5,000*	397	1,087	102	2.22
Section VI	5,000*	625	3,372	122	1.08

Conclusions

- The test section with a furnace slag stabilized BCS base performed significantly better than its counterpart section with a fly-ash stabilized BCS base;
- Both stabilized BCS sections appeared to marginally out-perform the pavement sections with the foamed-asphalt stabilized bases and the crush limestone bases;
- The two foamed-asphalt treated bases did not perform well compared to a stone base. Both had an inferior structural capacity than the stone. Field survey indicated that one foamed-asphalt treated base had a shear flow problem possibly due to its poor aggregate skeleton, while the other one seemed to be over-asphalted;
- A cement-treated soil subbase was found to have a better load carrying capacity than a lime-treated subbase;
- FWD deflection data was found to reveal the structural deterioration fairly well.

Acknowledgments
This study was supported by the Louisiana Transportation Research Center and the Louisiana Department of Transportation and Development. The authors would like to express thanks to all those who provided valuable help in this study.

References
1. Huang, Y.H. (1993). *Pavement Analysis and Design*. Prentice Hall, Inc.
2. Pierce, L. M. and Mahoney, J.P. (1996). "Asphalt concrete overlay design case studies." *Journal of the Transportation Research Board.* No. 1543, 3-9.
3. *Wirtgen Cold Recycling Manual* (1998). Wirtgen GmbH, Hohner Straβe 2, D-53578 Windhagen.
4. Wu, Z, Zhang, Z., Mohammad, LN., King, B., Gautreau, G., Tao, M. and Herath, A. (2005). *Accelerated Loading Evaluation of a Sub-base Layer on Pavement Performance.* Interim Report. Louisiana Transportation Research Center.
5. Zhang, Z. and M. Tao. (2006). *Stability of Calcium Sulfate Base Course in a Wet Environment*. Final Report No. FHWA/LA-06/419, Louisiana Transportation Research Center.

Subject Index

Page number refers to first page of paper

Author Index

Page number refers to first page of paper