Bin Jiang
Xiaobai Yao
*Editors*

The GeoJournal Library 99

# Geospatial Analysis and Modelling of Urban Structure and Dynamics

Geospatial Analysis and Modelling of Urban Structure and Dynamics

## The GeoJournal Library

Volume 99



For further volumes:
http://www.springer.com/series/6007

Bin Jiang · Xiaobai Yao
Editors

# Geospatial Analysis and Modelling of Urban Structure and Dynamics

Foreword by Michael Batty

Springer

*Editors*
Bin Jiang
University of Gävle
Department of Technology and Built Environment
Division of Geomatics
SE-801 76 Gävle
Sweden
bin.jiang@hig.se

Xiaobai Yao
Department of Geography
University of Georgia
Athens GA 30602
Room 204, GG Bldg.
USA
xyao@uga.edu

ISSN 0924-5499
ISBN 978-90-481-8571-9 e-ISBN 978-90-481-8572-6
DOI 10.1007/978-90-481-8572-6
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2010922459



Printed on acid-free paper

Springer is part of Springer Science + Business Media (www.springer.com)

# Foreword

## A Coming of Age: Geospatial Analysis and Modelling in the Early Twenty First Century

Forty years ago when spatial analysis first emerged as a distinct theme within geography's quantitative revolution, the focus was largely on consistent methods for measuring spatial correlation. The concept of spatial autocorrelation took pride of place, mirroring concerns in time-series analysis about similar kinds of dependence known to distort the standard probability theory used to derive appropriate statistics. Early applications of spatial correlation tended to reflect geographical patterns expressed as points. The perspective taken on such analytical thinking was founded on induction, the search for pattern in data with a view to suggesting appropriate hypotheses which could subsequently be tested. In parallel but using very different techniques came the development of a more deductive style of analysis based on modelling and thence simulation. Here the focus was on translating prior theory into forms for generating testable predictions whose outcomes could be compared with observations about some system or phenomenon of interest.

In the intervening years, spatial analysis has broadened to embrace both inductive and deductive approaches, often combining both in different mixes for the variety of problems to which it is now applied. Moreover, the focus has become more explicitly geographical although the term spatial still has a wider usage for many of the statistics and models that form the arsenal of techniques in this area are applicable to spatial systems other than the obvious geographies – such as ecologies, climatic regimes, and even astronomies. In this collection of papers, however, the authors use the term "geospatial" to ground their systems of interest geographical conceptions of cities and regions, but they also show how many of the advances over the last 40 years are now part of the background knowledge that constitutes this field. In fact, the concerns here are with major extensions of analysis and modelling which reflect new themes in geographical thinking that are resulting from changes in our perceptions of city systems. These are largely due to demographic, technological and behavioural change which are driving new problems and the need for new methods, as well as the enormous changes that have taken place in our ability to collect, process and visualise digital data.

The way the editors have put together these chapters which are drawn from a wide and active community of scholars, reflects these concerns. The themes developed are based on new ways of data capture leading to data analysis orders of magnitude larger than anything hitherto in this field and at even finer scales, new ideas about the intrinsic complexity of the systems that we are dealing with and the limits to prediction, agent-based models of how individuals move and perceive their environments which are reflected in new ideas about transportation systems and new idea about measuring nearness in terms of accessibility, and new ways of incorporating time and dynamics in our models. The treatment is set of course against a background where the media for analysis and modelling is ever more visual and is fast moving from the desktop to the web and thence into real time contexts, unimaginable even when computers had become all pervasive by the mid 1990s.

It is worth elaborating several of these themes which emerge from the various chapters as they illustrate how standard theory and analysis is being supplemented by new ideas. Omer for example, shows how the long standing problem of spatial aggregation, often referred to as the "ecological fallacy" or the "modifiable aerial unit problem" influences and distorts our interpretation of the way space divides up, in his case residential segregation in the Israel town of Jaffa. His work blends issues about perception with the resolution at which the analysis takes place. Kostakos tackles the perception problem in a different way, building on movement patterns at an equally fine scale, involving the methods of space syntax, but showing how new technologies based on data capture using mobile phones can enrich our entire understanding of local space and its analysis. In some senses, these first two papers in the book point directions to a world of immensely rich data that analysts in this field will have to grapple with in the very near future, and the techniques developed are amongst the first rudimentary tools that will be needed as this new world enfolds. These are given real substance in the chapter by Laube and his colleagues where the notion of decentralised spatial computing in real time environments is demonstrated with respect to accessing location-based services that depend on high quality mapping databases.

A second theme dominating our thinking about city systems involves their complexity. Cities like most human systems can be understood on many levels and although it sounds extreme, every individual will perceive this complexity differently. This is in and of itself the very definition of what complexity means. Nevertheless, progress is being made in mapping structures of complexity, and in the last decade, network science has been revived to make sense of such structures through notions of hierarchy and scaling. Blumenfeld–Lieberthal and Portugali develop a simple but effective agent-based network model of city growth that generates the classic scaling of city sizes that mirrors Zipf's well-known rank size rule while Chen takes a different perspective, embedding fractal scaling of city sizes into Christaller-like central place hierarchical structures. This again demonstrates a really nice feature of this book where many of the authors illustrate that the current frontier is built on the older foundations of spatial analysis and location theory, notwithstanding the massive variety that now characterise developments in this and similar fields.

The editors here draw a fine line between methodological and domain-based knowledge and this is nowhere clearer than in the general area of transportation analysis and modelling. Network structures, agent-based models, and real-time allocation in terms of location-based services are developed for a series of different transportation examples. Martens and colleagues develop a very innovative agent-based model for parking in real-time contexts using a car-following allocation scheme. Chaker, Moulin and Theriault develop a much larger scale application for simulating travel activities and set this in a virtual urban environment. Omer and Jiang examine segregation in street networks, again for the town of Jaffa using Atkin's Q-analysis which reflects a bipartite topological network approach providing an alternative and different view of residential segregation. Lu and her colleagues examine the perceptions of levels of transport service in Beijing from survey data while Mandloi and Thill extend their analysis of transport networks from the street to the building, illustrating how routing problems need to be considered across many scales. In all these chapters, resolution, scale, networks and cognition figure strongly, impressing once again the importance of the new themes generated in these various contributions.

The other profound change in this field has been the widespread underpinning of spatial analysis with GIS and remote sensing. Many of the chapters here use GIS as an enabling technology but GIS continues to have a major effect on the types of models and analytical techniques that are being developed. Combined with the continuing focus on urban growth, particularly in the developing world, various models have been developed which are closely linked to GIS and to the simulation of land use change, often supported by data that is sensed remotely. This domain of applications blends cellular automata with pattern analysis, often originating from ecology as Chen and her colleagues demonstrate for Greater Toronto. Integrating these styles of analysis and simulation is another important area of research as Yang illustrates in an analysis of growth in the Atlanta metro area. Research of considerable importance that is likely to see many more applications in the next decade as climate change becomes ever more significant are simulations of the urban water cycle as presented by Shepherd and his team. In examining urban change, techniques for smoothing spatial data are important in deriving consistent interpretations as Varanka shows, while embedding all these kinds of growth process in more formal space-time frameworks with appropriate ontologies is addressed by Yao.

Most of the contributions collected here have obvious policy relevance and some directly link analysis and modelling to policy questions. Urban complexity and form are issues addressed by Pagliardini and his colleagues who develop the "city as organism" argument in a spirited essay on what they call "urban seeding", which they argue represents a much more careful and less intrusive paradigm for urban planning and design than anything adopted hitherto. Besides urban form, spatial differences in disease and treatment increasingly occupy an important place in spatial analysis and Mu, Wang and McLafferty illustrate how important spatial relations, adjacencies and correlations are to the interpretation of cancers at different spatial scales, once again involving issues involving aggregation. Accessibility issues and the relationship to prices are crucial to questions of housing market provision and efficiency with Hwang and Thill providing an excellent demonstration of how careful spatial econometrics can be used to unravel complex effects. Efficiency versus equity are issues that Horner and Widener examine spatially in the provision of disaster relief while methods of aggregation, geographical weighting, and locality are explored by Wheeler in a visual analysis of spatial pattern, providing essential diagnostics for both interpretation and the construction of spatial policies.

The contributions in this book present an excellent profile of the state of geospatial analysis and modelling after the first decade of work in the twenty first century. Compared to 10 years ago, there are now some really explicit drivers to this field which are picked up and exploited here. Amongst the many origins and themes, new ways of capturing data digitally at the individual level, the development of systems based around networks (of which spatial locations are a subset), the notion of linking hierarchy to networks to morphology as in complexity theory, and the development of new ways of integrating diverse urban processes through simulation paying careful attention to the basic econometric and statistical principles of spatial analysis, are all represented here. Jiang and Yao do a great job in both selecting a comprehensive range of papers while at the same time, emphasising these key themes. Readers will enjoy what follows and the book will provide both a pointer to the field and an inspiration for further research.

London Michael Batty

# Acknowledgments

The publication of this book owes a great deal to many individuals and organizations, to Daniel Z. Sui for his confidence in us by bringing this book project into the Springer book series in GeoJournal Library, to Efrat Blumenfeld-Lieberthal, James B. Holt, Sungsoon Hwang, Hiroyuki Iseki, Daniel Koch, Lan Mu, Stephan Winter, and Changshan Wu, for reviewing parts of the manuscripts, to Junjun Yin and Petra Norrland for assisting us in formatting and editing the chapters, and to all the authors for their quality contributions and timely revisions.

The International Cartographic Association Commission on Geospatial Analysis and Modeling, and the NordForsk funded Nordic Network in Geographic Information Science provide international platforms for networking active researchers in this dynamically evolving field of geographic information science. Both University of Gävle and University of Georgia, in particular the division or department with which we have been closely working, provide us friendly yet stimulating working environments.

# Authors Biographies

## Editors

**Bin Jiang** is Professor at University of Gävle, Sweden, and is affiliated to the Royal Institute of Technology (KTH) at Stockholm via KTH Research School. He received PhD in GeoInformatics from University of Utrecht and International Institute for Geo-Information Science and Earth Observation (ITC), the Netherlands. He worked in the past with The Hong Kong Polytechnic University and the Centre for Advanced Spatial Analysis of University College London. He is the founder and chair of the International Cartographic Association Commission on Geospatial Analysis and Modelling. He has been coordinating the NordForsk-funded Nordic Network in Geographic Information Science. His research interest is geospatial analysis and modelling, in particular topological analysis of urban street networks in the context of geographic information systems. He is primary developer of two software tools: Axwoman and AxialGen for space syntax related research. He is currently an Associate Editor of the international journal Computer, Environment and Urban Systems.

**Xiaobai A. Yao** is Associate Professor of Geography at the University of Georgia. She received BS in 1990 from Wuhan University in urban planning and management, before taking the position as a GIS specialist in Shanghai Municipal Bureau of Urban Planning and Management. She obtained her Master's degree in 1997 from the International Institute of Aerospace and Earth Science in the Netherlands and her PhD in Geography from the State University of New York at Buffalo in 2002. Dr. Yao joined the faculty at the University of Georgia in 2002 where she has been conducting research and teaching ever since. She received a best student paper award in 2002 at the Association of American Geographers (AAG) annual meeting and the ESRI-UCGIS Young Scholarship Award in 2005. She has co-organized two national or international research workshops and published 15 peer-reviewed research articles. Her current research interests center on Geocomputation, spatial analysis, and urban transportation.

## Contributors

**Itzhak Benenson** is Professor of Geography at the Department of Geography and Human Environment of Tel Aviv University, Israel, and Head of Spatial Modeling and GIS laboratory there. He received his MS degree in Mathematics from the Urals State University, Sverdlovsk, USSR, and his PhD in Mathematical Biophysics from Moscow University in 1985. His main areas of interest are urban and regional modeling and simulation, spatial population dynamics, GIS and spatial analysis. He serves on the editorial board of several journals and is an author of 90 papers and 3 books, the latest, written together with Paul Torrens, titled "Geosimulation: Automata-Based Modeling of Urban Phenomena" (Wiley, 2004). Itzhak's recent research focuses on the Geosimulation of urban and regional phenomena, including Agent-Based modeling of the use of public transport and parking in the city; modeling vehicle-pedestrian interactions and road accidents at junctions; high-resolution analysis of urban residential dynamics; analysis of the long-term impact of local and regional plans on actual land-use dynamics; modeling vulnerability of communities and territories; and archaeological GIS and analysis of the form of ancient vessels.

**Efrat Blumenfeld-Lieberthal** is an architect and a lecturer of Architecture at David Azrieli School of Architecture Tel Aviv University. She received her B. Arch, and PhD at the Faculty of Architecture and Town Planning, at the Technion – Israel Institute of Technology. Her first Post doctorate research was at the Center for Advanced Spatial Analysis (CASA), University College London. She has recently completed her second post doctorate research as a Porter Fellow at the Environmental Simulation Laboratory (ESLab), Tel Aviv University. Her research interests are applying theories of complexity to urban environments; urban morphology; size distribution of entities in complex systems; and complex networks in urban systems.

**Walid Chaker** is a PhD student in the Department of Computer Science and Soft-ware Engineering at Laval University. He received his BA degree in Mathematics, Physics and Computer Science from the University of Tunis in 1998 and his MSc in Computer Science from Laval University in 2001. He was the main designer and developer of the MAGS platform when he worked at Laval University as a research scientist from 2001 to 2003. The MAGS platform was later used in and extended for various simulation applications (crowd behaviors, consumer behaviors in shopping malls, wildfire fighting, train traffic, West Nile virus propagation). Walid's current research interests are related to multi-agent approaches applied to spatial simulations with a special focus on multiscale representations of urban environments. He published 10 articles in refereed journals and conference proceedings from 1999 to 2008.

**Dongmei Chen** is an associate professor at Department of Geography and director of the Laboratory of Geographical Information and Spatial Analysis (LAGISA) at Queen's University, Kingston, Canada. She received her BA degree in economic geography from Peking University, China, MSc in GIS and remote sensing from the Institute of Applied Remote Sensing, Chinese Academy of Science, and a joint doctorate degree in geography from San Diego State University and University of California at Santa Barbara. She has published more than 30 peer-reviewed articles and book chapters on GIS, remote sensing, spatial analysis, and their applications in environmental monitoring, land use planning, ecology, crime, health, etc. Her current research focuses on the understanding of dynamic spatial/temporal patterns and structures in environmental and health applications using GIS, remote sensing, spatial statistics, dynamic modeling and simulation.

**Jing Chen** is a research assistant at Beijing Research Center for Information Technology in Agriculture, Beijing, China. She received a B.E. degree in urban planning from the Department of Urban Construction of Jiangxi Normal University in China in 2004. In 2007, she got her M.S. degree in human geography from the Department of Geography and Environment Sciences of Jiangxi Normal University, through a joint graduate program between Jiangxi Normal University and Beijing Union University. Jing was a research fellow at Yen Ching Real Estate Institute in 2005. She was a trainee of Beijing Sunland Real Estate Brokerage Co., Ltd. in 2006. From 2007 to 2008, she was a teaching assistant of Beijing Geely University. Her current research interests include geo-information system, agriculture information, and livable city study.

**Yanguang Chen** is associate professor at Peking University, Beijing, China. He received his BA degree from Huazhong Normal University in 1987, his master's degree in Northeast Normal University in 1995 and his PhD in geography from Peking University in 2004. He was elected member of Academic Council of the Geographical Society of China (GSC) from 2005 to 2009, and committeeman of Professional Council of Policy Simulation in the Association of Science of Science (∇ϕϕ) and Science and Technology Policy of China from 2007 to 2010. He published over 5 books and 160 articles including 150 articles in Chinese and 13 articles in English. His major publications in Chinese is as follows: Cities and Urban Geography (2003, collaborated with Prof. Yixing Zhou); Fractal Urban Systems: Scaling, Symmetry, and Spatial Complexity (2008); Geo-Mathematical Methods: Principles and Applications (2009); Geographical Analysis and Modeling Using MS Excel (2009); Geographical Analysis and Modeling Using Mathcad (2009). His current research interests center on fractal cities, self-organized network of cities, allometric scaling analysis, symmetry and symmetry breaking in geography.

**Matt Duckham** is a Senior Lecturer in GIScience at the Department of Geomatics, University of Melbourne, Australia. Before moving to Australia, Matt worked at the US National Center for Geographic Information and Analysis (NCGIA) at the University of Maine. He holds degrees from the Universities of Edinburgh, Leicester, and Glasgow, in the UK. Matt has published three books, including co-authoring with Mike Worboys the second edition of the major GIScience text book “GIS: A computing Perspective”. His research is in the area of distributed computation with uncertain spatial and spatiotemporal information. In recent years, his main research focus has been on robust, decentralized spatiotemporal algorithms, for example monitoring spatial events in a geosensor network using no centralized control.

**Michael F. Goodchild** is Professor of Geography at the University of California, Santa Barbara, and Director of spatial@ucsb. He received his BA degree from Cambridge University in Physics in 1965 and his PhD in geography from McMaster University in 1969. He was elected member of the National Academy of Sciences and Foreign Fellow of the Royal Society of Canada in 2002, and member of the American Academy of Arts and Sciences in 2006, and in 2007 he received the Prix Vautrin Lud. He was editor of Geographical Analysis between 1987 and 1990 and editor of the Methods, Models, and Geographic Information Sciences section of the Annals of the Association of American Geographers from 2000 to 2006. He serves on the editorial boards of ten other journals and book se-ries, and has published over 15 books and 400 articles. He was Chair of the National Research Council’s Mapping Science Committee from 1997 to 1999, and currently chairs the Advisory Committee on Social, Behavioral, and Economic Sciences of the National Science Foundation. His current research interests center on geographic information science, spatial analysis, and uncertainty in geographic data.

**Lauren Hand** is a Geographer and Emergency Management Specialist for Dewberry and Davis. Her work provides a complete range of programmatic, technical assistance and support to federal, state and local governments and to private clients in the field of hurricanes, tropical cyclones, and emergency management. Over the past year, Ms. Hand has supported the US Army Corps of Engineers and FEMA in the development and modernization of the National Hurricane Program’s Comprehensive Hurricane Emergency Management Strategies (CHEMS) that will provide decision makers at all levels of government with the information and tools to assist them to prepare for, mitigate against, respond to and recover from disasters of all types. Her other specialties include hydrologic and hydraulic modeling, data collection, outreach, training, and GIS. Her graduate research focused on urban climate and precipitation and her thesis, An Investigation of Warm Season Spatial Rainfall Variability in Oklahoma City: Possible Linkages to Urbanization and Prevailing Wind, was published in the Journal of Applied Meteorology and Climatology. Ms. Hand is an AMS/Industry/Government Graduate Fellow and Meyerhoff Scholar. She received her B.S. in Environmental

Science from the University of Maryland, Baltimore County and her M.S. in Geography from the University of Georgia.

**Mark W. Horner** is Associate Professor of Geography at the Florida State University. He received his PhD in geography from Ohio State University in 2002. Mark is past Chair of the Spatial Analysis and Modeling Group (SAM) of the Association of American Geographers (2007–2009). He is presently serving a 2-year term on the editorial board of Transportation, and is a member two Transportation Research Board (TRB) standing committees (ADD20 – Social and Economic Factors in Transportation and ABJ60 – Geographic Information Science and Applications). His current research interests are in geographic information science, transportation, and urban geography.

**Sungsoon Hwang** is an Assistant Professor of Geography at DePaul University. She received her BA in Geography Education from Seoul National University in 1997, her MSc in GIS from University of Nottingham in 1999 and her PhD in Geography from SUNY at Buffalo in 2005. She was one of recipients of International Geographic Information Fund Award by Association of American Geographers in 2005. She has published in the areas ranging from GIScience to urban geography on the subject of housing and transportation. Her current research interest lies in examining the role of space influencing urban sustainability by geocomputational methods and developing GIS curricula for sustainable development.

**Tony Joyce** is a research scientist at Ordnance Survey, Great Britain's national mapping agency. Tony holds an MSc (GIS, 1995) from the University of Leicester, United Kingdom. His research at Ordnance Survey covers geospatial data modeling, including spatio-temporal data modeling and historical data visualization. Currently, his most active research projects are concerned with the concept of space, modeling vague extents and the semantic web.

**Vassilis Kostakos** is Assistant Professor in the Department of Mathematics & Engineering at the University of Madeira, Portugal, and Adjunct Assistant Professor in the Human Computer Interaction Institute at Carnegie Mellon University. He received his BSc degree in Computer Science from the University of Bath in 2001, and his PhD degree in Computer Science from the University of Bath in 2004. His post-doctoral work was on the Cityware project in the UK. He was a visiting scientist at UCSD in 2007, and visiting research fellow at UCL and Hewlett Pack-ard between 2005 and 2008. In 2006 he received a Role Model award in science by the Engineering and Technology Board (UK), and was declared Science Ambassador by Stemnet (UK). He reviews for a number of academic conferences and journals, and his work has been published in international conferences and journals. His work has also been reported by popular media (The Guardian, BBC, New Scientist). His research interests are:

pervasive and ubiquitous systems, human-computer interaction, security and privacy, social networks, and transport.

**Patrick Laube** is a research fellow with the Geomatics Department at the University of Melbourne, Australia. Patrick holds an M.Sc (Geography, 1999) and a Ph.D. degree (Sciences, 2005) from the University of Zurich, Switzerland. His thesis covered the analysis of movement data, presenting an approach for spatio-temporal data mining basing on pattern detection and visualization. Recently he was a research fellow at the Spatial Analysis Facility at the University of Auckland, NZ, and a visiting scholar at the GeoVISTA Center at Penn State University, PA, USA. His research interests include spatio-temporal geographic information, especially everything related to movement, as well as decentralized spatial computing techniques for geosensor networks.

**Nadav Levy** is a master student in Geography at the Department of Geography and Human Environment of Tel Aviv University. He holds a BA degree in Geography from Tel Aviv University, Israel. Nadav's research interests include agent-based modeling and its application to the analysis of urban dynamics, among which parking in the city.

**Wenbao Liu** is a PhD Candidate in Geography at Queen's University, Kingston, Ontario, Canada. He received his BSc degree from Shandong University of Science and Technology in Surveying Engineering in 1987 and his MSc in geography from the University of Guelph in 2005. His current research interests are spatial analysis and impact of climate change, land use change, and population growth on surface water quality and quantity.

**Yongmei Lu** is an associate professor with the Department of Geography, Texas State University – San Marcos. She received her PhD from University at Buffalo, State University of New York in 2001, and her MS and BS degrees from Peking University in 1994 and 1991 respectively. Dr. Lu's major research interests are the integration of GIS with spatial analysis and modeling on socio-economic and environmental studies. She is especially interested in GIS and spatial analysis of urban and environmental issues, including crime patterns, urban transportation and urban growth, public health and environmental pollution. Dr. Lu's research has been supported by the US Department of Justice, the US Department of Defense, Texas State Governor's Office, and Texas State University faculty development grant and library research grant. She is the author/coauthor of a number of publications including major journal articles, book chapters, and papers in refereed conference proceedings. She has co-edited a book on GIS and urban hazard/disaster studies. Dr. Lu has served as a reviewer for a number of highly recognized journals as well as multiple programs with the National Science Foundation.

**Peter D. Luciani** is a Research Associate and PhD Candidate at the Laboratory of Geographical Information and Spatial Analysis (LAGISA) in the Geography Department at Queen's University, Kingston, Ontario, Canada. After receiving a BA degree in Urban and Regional Planning in 1993, Peter procured employment as a geo-environmental scientist in the private consulting sector, later receiving a MSc in Environmental Applied Science and Management in 2005. Both degrees are from Ryerson University in Toronto, Ontario, Canada. Peter's current research interests are focused upon distributed water quality modeling targeting heavy metal contamination related to urbanization using tools of spatial analysis, environ-mental modeling and remote sensing.

**Deelesh Mandloi** works as a Product Engineer at ESRI Inc head quarters in Redlands, California. He spends his time designing, documenting and testing the Network Analyst extension for ArcGIS. He received his BS in Civil Engineering from Birla Institute of Technology and Science, Pilani, India in 2003 and his Master's in Geography from State University of New York, University at Buffalo, NY in 2006. His current interests include designing intuitive and powerful GIS software for analysis of Transportation networks and performing network based three dimensional analyses inside built environments such as high rise buildings. He has also worked on the integration of Operations Research techniques with GIS software to solve problems in logistics and transportation.

**Michael Manyin** is a scientific programming specialist in the Atmospheric Chemistry and Dynamics Branch, at NASA's Goddard Space Flight Center. He works with chemistry transport models and remotely sensed data, supporting research focused on the behavior of stratospheric ozone and the trace gases which affect it. He has previously worked with Dr. Shepherd and others in the study of urban climate, employing regional atmospheric models. He has also supported research in convective/stratiform rain estimation, and was principal programmer of Goddard's Image SpreadSheet visualization software. He received an M.S. in computer science from Cornell University.

**Karel Martens** is Assistant Professor in Transport Planning at the Department of Geography, Planning and the Environment, Radboud University Nijmegen, the Netherlands. He received his MA degree in Spatial Planning in 1991 and his PhD in Policy Sciences in 2000, both from the Radboud University Nijmegen. He has over 15 years of experience as an academic and practitioner in the fields of transportation planning and urban planning, in the Netherlands, Israel and, recently, Belgium. He has worked for universities in the Netherlands and Israel, for two international consultancy companies, as a director of a non-governmental organization, and as a private consultant. His fields of expertise include the nexus between transport and land use, multi-modal travel, transport and justice, and participatory governance. Karel's research focus takes sustainable transport as a starting point. From this perspective, he is working on a number of research projects, including agent-based modeling of the use of public transport and parking in the city, multi-modality and cooperation in freight transport, transport modeling and justice, cost-benefit analysis and equity, the interrelationship between infrastructure investments and spatial development and a supply-side approach to car sharing.

**Sara McLafferty** is Professor of Geography at the University of Illinois at Urbana-Champaign. She obtained her B.A. from Barnard College and M.A. and Ph.D. degrees from the University of Iowa. Her research investigates place-based disparities in health and access to health services and employment opportunities for women, immigrants and racial/ethnic minorities in the United States. She has also written about the use of GIS and spatial analysis methods in exploring inequalities in health and access to health care. Her books include GIS and Public Health (with Ellen Cromley) and Geographies of Women's Health (with Isabel Dyck and Nancy Lewis). She has published in a wide range of geography, epidemiology, and urban studies journals and currently serves on the editorial boards of Annals of the Association of American Geographers, Geographical Analysis, Transactions in GIS and Health and Place.

**Dmitry Messen** is a Socioeconomic Modeling Program Manager at the Houston-Galveston Area Council (H-GAC). Dmitry leads H-GAC's efforts in developing long-range small-area demographic, economic, and land use forecasts for the Houston region. Prior to joining H-GAC in 2004, Dmitry worked at the Center for Energy Studies at Louisiana State University doing research in applied economics as well as teaching GIS, statistics, and economic geography. Dmitry is a quantitative social scientist with extensive expertise in modeling, regional and urban economics, geographic information systems, and data management. Dmitry holds a Ph.D. in Geography from Louisiana State University and a M.A. in Geography from Moscow State University.

**Bernard Moulin** is a full professor at Laval University, Québec, Canada. He is teaching in the Computer Science and Software Engineering Department. He is also a member of the Research Center in Geomatics at Laval University and an active researcher of GEOIDE, the Canadian Network of Centers of Excellence in Geomatics. He received his engineering degree from l'Ecole Centrale de Lyon (1976), his Master in economics from the University Lyon II (1976) and his Ph.D. in computer science from the University Lyon I (1979) (all in France). He leads several research projects in various fields: Multi-agent geo-simulation, Design methods for multi-agent systems and software-agent environments; representation of temporal and spatial knowledge in discourse; modeling and simulation of conversations between artificial agents; modeling and design approaches for knowledge-based systems and multi-agent systems, as well as several projects at the intersection of geomatics and artificial intelligence. These research projects are funded by the Natural Science and Engineering Council of Canada, the Canadian Network of Centers of Excellence in Geomatics GEOIDE, the Defense Research Establishment at Valcartier and several other organizations and private companies.

**Lan Mu** is Assistant Professor of Geography at the University of Georgia. She received her B.S. in environmental science from Peking University, M.A. in geography and planning from University of Toledo, and Ph.D. in environmental planning from University of California, Berkeley. Her research topics focus on geographic information science (GIScience), spatial analysis, computational geometry in GIScience, and cartography and visualization. She is also interested in developing methods and tools for GIScience research such as multiplicatively weighted Voronoi Diagram (WVD), shape-based buffering, Modified Scale-Space Clustering (MSSC), and Weighted Difference Barrier (WDB). She has published in geography, GIScience, and urban studies journals.

**Itzhak Omer** is an urban-social geographer. He got his PhD in Urban Geography from Tel Aviv University (with distinction) in 1996 and is now a senior lecturer at the Department of Geography and Human Environment, Tel Aviv University. He is Head of the Urban Space Analysis Laboratory of Tel Aviv University and a member of Israeli Geographic Association Council. He is also a Head of Urban and Regional Modeling Working Group of the Association of Geographic Information Laboratories Europe (AGILE). The areas of academic interest of Itzhak Omer include: Simulation models of urban and environmental dynamics, Virtual environments, Spatial Cognition, Social Geography, and Urban morphology. His current research focuses on effects of the functional and morphological aspects of urban environment on the spatial cognition and behavior of people.

**Pietro Pagliardini** is an architect and urbanist born in Arezzo, Italy. As senior partner of Pagliardini, Rupi, Andreoni & Gazzabin Architecture Studio he has designed social housing and completed the design of over 1000 dwellings and over 50 industrial buildings. He participated in the preliminary phases of 17 different Integrated Plans for mixed-use social and private housing throughout Italy, and planning three of those neighborhoods. For the project in Arezzo he was also responsible for the architectural design. In an important architectural competition for the reconstruction of the historic city center of Arezzo, Italy, he collaborated with world-famous New Urbanists Léon Krier and Peter Calthorpe. Again with Léon Krier as consultant and Peter Calthorpe as supervisor working for the Arezzo Town Council, Pagliardini created the urban plan for social and private housing for the rural village of Vignale, studying and applying local building typologies. Pagliardini has spearheaded the movement to rediscover traditional Italian urbanism, coordinating various groups already working in that direction previously working in isolation. He has been influential in alert-ing the government to these possibilities, appearing on public discussions with important politicians such as Vittorio Sgarbi, and arguing for the value of preserving historic built heritage.

**Sergio Porta** architect, Ph.D. is assistant professor at the Milan Politechnic, Italy, and adjunct professor at the King Fahd University of Petroleum and Minerals in Dhahran, Saudi Arabia. In Milan he is director of the Human Space Lab, which is currently involved in street design schemes and applied research in Italy and abroad. Dr. Porta is coordinator of an international network of space analysis and design named UStED – Urban Sustainability through Environmental Design, and is a member of ESRG – Environmental Structure Research Group. Dr. Porta also works as a consultant on architectural design and urban planning in Milan and Reggio Emilia, focusing on the social dimension of architecture and public space design. Dr. Porta was a visiting scholar at the Institute of Govern-mental Studies of the University of California at Berkeley, and is an associate researcher at the Institute for Sustainability and Technology Policy at Murdoch University in Perth, Western Australia, where he teaches a short course on sustainable urban design. Dr. Porta's latest research is oriented towards sustain-able/human/adaptive urban analysis and design, ranging from GIS-based space analysis to sustainable community design, transportation planning, and traffic calming techniques, to strategies for safety and livability in the public scene.

**Juval Portugali** is a professor of Geography at the department of Geography and the Human Environment, Tel Aviv University. He is the founder and head of the ESLab (Environmental Simulation Laboratory), the Porter School of Environmental Studies, founded at Tel Aviv University in 2001, head of the Environment and Society Graduate Program. He holds Ph.D at the Department of Geography, The London School of Economics and Political Sciences. He specializes in theories of complexity and self-organization related to agent-based modeling, urbanism, socio-spatial change, cognitive maps and spatial archaeology. Current research includes (i) The city as a complex self-organizing system. 2D and 3D cellular automata and agent base urban simulation models. (ii) Inter-representation nets and the construction of cognitive maps. His recent books include: Complex Artificial Environments, Springer Complexity Series, 2005 (Ed.); Self-Organization and the city, Springer, 2000; The Construction of Cognitive Maps, Kluwer Academic Publishers, Dordrecht, 1996 (Ed.); Implicate Relations: Society and space in the Israeli-Palestinian conflict, Kluwer Academic Publishers, 1993.

**Nikos A. Salingaros** MA, PhD, ICTP, ICoH, ICA is the author of "Anti-Architecture and Deconstruction" (2004), "Principles of Urban Structure" (2005), and "A Theory of Architecture" (2006), as well as numerous scientific papers. Both an artist and scientist, he is Professor of Mathematics at the

University of Texas at San Antonio, and is also on the architecture faculties of universities in Holland, Italy, and Mexico. He designed the Commercial Center in Doha, Qatar in collaboration with Hadi Simaan and José CorneliodaSilva. Dr. Salingaros' theoretical work under-pins and helps to link new movements in architecture and urbanism, such as New Urbanism, the Network City, Biophilic Design, Self-built Housing, and Sustainable Architecture. He is working with the Peer-to-Peer Foundation to promote self-built housing for the developing world. Dr. Salingaros collaborated with Christopher Alexander, helping to edit the four-volume "The Nature of Order" during its 25-year gestation. In recognition of his efforts to under-stand architecture using scientific thinking, he was awarded the first grant ever for research on architecture by the Alfred P. Sloan Foundation, in 1997. Dr. Salingaros is a member of the INTBAU College of Traditional Practitioners, and is on the INTBAU Committee of Honor. Dr. Salingaros was included as one of the "50 Visionaries" selected by the UTNE Reader in 2008.

**Willis Shem** has just completed a 1-year appointment as a Lecturer in Meteorology in the department of Geology, Geography and physics at the University of Tennessee in Martin. Prior to that he was a post-doctoral research associate in the department of Geography at the University of Georgia. His research work at the University of Georgia was mainly on the impact of urbanization on the precipitation component of the hydrological cycle with focus over the Atlanta region. Part of this work was recently (2009) published in the "Atmospheric Research" Journal. Dr. Shem has made several presentations in conferences organized by the American Meteorological Association (AMS) and Association of American Geographers (AAG) and has membership in both organizations. Dr. Shem has a multi-disciplinary background having received his BS (Meteorology) from Nairobi University (Kenya), MEng. (Water Resources Technology) from Vrije Universiteit Brussels (Belgium) and PhD (Earth and Atmospheric Science) from Georgia Institute of Technology (USA). Dr. Shem re-joined the University of Georgia starr in the summer of 2009 as a post-doctoral researcher.

**J. Marshall Shepherd** is an associate professor of geography/atmospheric sciences at the University of Georgia. He conducts research, advising, and teaching in atmospheric sciences, climatology, water cycle processes and urban climate systems. Prior to joining the UGA faculty, Dr. Shepherd spent 12 years as a research meteorologist at NASA. Dr. Shepherd was also Deputy Project Scientist for the Global Precipitation Measurement (GPM) mission. For his work on urban climate, Dr. Shepherd was honored in 2004 at the White House with the Presidential Early Career Award for pioneering scientific research. Dr. Shepherd is a Fellow of the American Meteorological Society and has over 50 publications (refereed articles, book chapters, reports). Dr. Shepherd has served on the American Meteorological Society (AMS) Executive Council and as contributing author on the 2007

Inter-governmental Panel on Climate Change (IPCC) AR4 report amongst other activities. He serves on the NOAA Climate Working Group and is a Project Associate for the Urbanization and Global Environmental Change project. He is an editor for the Journal of Applied Meteorology and Climatology and co-section editor (climatology) for the journal, Geography Compass. Dr. Shepherd received his B.S., M.S. and PhD in physical meteorology from Florida State University.

**Marius Thériault** is geographer and full professor at the Graduate School of Land Planning and Regional Development (ESAD), Laval University, Québec, Canada. Since 1979, he was teaching geographical information science, spatial analysis, applied statistics, geo-simulation, transportation and applied computing. He was Director of the Land Planning and Regional Development Research Centre at Laval University, from 2000 to 2006. He is also member of the Research Centre in Geomatics at Laval University. He is involved in several research networks (including GEOIDE), multi-disciplinary research projects and peer reviewed journals in Canada and Europe. His research activities are about combining GIS, spatial statistics, geo-simulation and spatiotemporal modeling to further studies related to real estate markets, mobility behavior, accessibility to urban amenities, environ-mental impacts assessment, urban and regional planning. Findings of his research are published in more than 40 academic journals related to urban economics, housing, geography, geographical information science, transportation planning, computer science, environment, health science and regional development. He has edited several books.'

**Jean-Claude Thill** is the Knight Distinguished Professor of Public Policy at the University of North Carolina at Charlotte. He holds a doctorate in Geography from the Catholic University of Louvain, Belgium. Dr. Thill's multi-prong research has centered on the spatial dimension of mobility systems and their consequences on how space is used and organized in modern societies, statistical and computational methods of spatial analysis, and most recently urban land-use dynamics. Since 2008, he has been Editor-in-Chief of the international journal Computers, Environment and Urban Systems, and serves on the editorial boards of several other regional, national, and international journals of geography, regional science, and spatial systems. He is the Executive Director of the North American Regional Science Council. He received the 1988 Philippe Aydalot Prize, the 1996 Geoffrey J. D. Hewings Ward, and the 2008 David Boyce Service Award from the North American Regional Science Council. He has held faculty positions at Florida Atlantic University, The University of Georgia, and The State University of New York at Buffalo. He joined the University of North Carolina at Charlotte in 2006.

**Jie Tian** is a postdoctoral fellow in the Laboratory of Geographic Information & Spatial Analysis at the Geography Department of Queen's University, Kingston, Ontario. He received his BSc degree from Beijing (Peking) University in Earth Science in 2001 and his Master's (2004) and PhD (2008) in geography from the University of Western Ontario and Queen's University, respectively. His research interests broadly include GIS, remote sensing, geostatisticas, landscape ecology, and spatial epidemiology.

**Dalia Varanka** is a Research Geographer with the U.S. Geological Survey. She received her BA degree from the University of Wisconsin-Green Bay in Regional and Urban Analysis (1978), her M.A. degree in Geography at the University of Illinois at Chicago (1987), and her Ph.D degree in Geography from the University of Wisconsin-Milwaukee (1994). Dr. Varanka began her Federal career with the Bureau of Land Management in 1993 and has been with the U.S. Geological Survey since 1997, focusing on research interests in geographical analysis and urban ecology.

**Fahui Wang** is Fred B. Kniffen Professor at the Department of Geography and Anthropology and Director of Chinese Culture and Commerce, Louisiana State University. He earned his B.S. in geography from Peking University, China, and his M.A. in economics and PhD in city and regional planning, both from the Ohio State University. His studies cover the spatial and economic structure of systems of cities; urban and regional development in developing countries; job access, commuting and disadvantaged population groups; crime patterns; health care access, health policy and cancer. His research methods include GIS, spatial statistics, and computational methods. His work has been supported by several federal grant agencies including the National Institute of Justice, U.S. Department of Housing and Urban Development, U.S. Department of Health and Human Services, the National Cancer Institute, and the National Science Foundation. He has published over 40 refereed journal articles. He is the editor of Geographic Information Systems and Crime Analysis (IDEA Group Publishers, 2005) and the author of Quantitative Methods and Applications in GIS (Taylor and Francis, 2006).

**David C. Wheeler** is a Cancer Prevention Fellow at the National Cancer Institute. He received his BA from Syracuse University and a MA in geography, MS in applied statistics, and PhD in geography from The Ohio State University. In addition, he is earning a Masters in Public Health at Harvard University in 2009. His research interests center on developing and applying spatial statistical methods to problems in public health and disease ecology. Methodological interests include statistical learning, Bayesian statistics, and visualization. He has recently published articles in Statistics in Medicine; Journal of Agricultural, Biological, and Environmental Statistics; Environment and Planning A; Journal of Geographical Systems; and The Professional Geographer.

**Michael J. Widener** is presently pursuing a Master's of Science in Geography at Florida State University. He is currently serving as the graduate student board member of the Spatial Analysis and Modeling Specialty Group (SAM) of the Association of American Geographers. Michael's research interests include urban geography, social theory, and computational methods.

**Mike Worboys** has a PhD in mathematics, and is Professor and Chair of the Department of Spatial Information Science and Engineering and a member of the National Center for Geographic Information and Analysis, University of Maine. Mike is a Distinguished Scientist of the Association for Computing Machinery, and life member of the London Mathematical Society. He has held posts at several UK universities, the Rutherford-Appleton Laboratory at Oxford and has held visiting professorships at the University of Melbourne, Australia, University of Marseille, France, and the Technical University of Vienna, Austria. He serves on the editorial boards of several leading research journals and book series, and is co-author with Matt Duckham of the textbook "GIS: A Computing Perspective". Mike has worked for many years at the boundary between computer science, mathematics, and geographic information science. His current research interests include the development of ontologies and data models for dynamic geographic phenomena, sensor informatics, approaches to reasoning with uncertainty in geographic phenomena, and connections and transitions between indoor and outdoor spaces.

**Xiaojun Yang** is a tenured Associate Professor of Geography at Florida State University. He completed his higher education in China, Holland and USA. His research interests include the development of remote sensing and geospatial technologies with applications for urban and environmental studies. His research has been funded with competitive grants from EPA, NSF, and NASA. He has authored or co-authored more than 100 publications including four journal theme is-sues and one book on coastal remote sensing. He was Guest Editor for ISPRS Journal of Photogrammetry and Remote Sensing, Photogrametrical Engineering and Remote Sensing, International Journal of Remote Sensing, and Computers, Environment and Urban Systems. He currently serves as Chair of the International Cartographic Association (ICA) Commission on Mapping from Satellite Imagery.

**Weihong Yin** is an associate professor of Geography at the Beijing Union University. She is the director of Center for Science of Human Settlements of BUU. Professor Yin received her MS degree in geography from Peking University in 1994. She is the standing director of Beijing Land Science Society, and a member of the Geographical Society of China. She has published 5 books and over 30 articles. Her current projects including "Spatial Database and Planning Research on Livable Beijing Interior Dwelling Environment" (supported by Beijing Municipal Natural Science

Foundation (8073024)) and "Evaluation and Construction on Livable Beijing Dwelling Environment with the Public Participation" (supported by Planning Program (Major) of Beijing Philosophy and Social Science (06AbCS001)). Her research interests center on livable city, spatial analysis of urban geography, real estate and land science.

# Contents

# Contributors

**Michael Batty** CASA, University College London, 1-19 Torrington Place, London WC1E 6BT, UK, m.batty@ucl.ac.uk

**Itzhak Benensonb** Geosimulation and Spatial Analysis Lab, Department of Geography and Human Environment, Tel-Aviv University, Ramat Aviv, Tel Aviv, Israel 69978, bennya@post.tau.ac.il

**Efrat Blumenfeld-Lieberthal** The Yolanda and David Katz Faculty of the Arts, The David Azrieli School of Architecture, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel, efrat.lieberthal@googlemail.com

**Walid Chaker** Department of Computer Science and Software Engineering, Laval University, Quebec G1V 0A6, Canada; Centre de Recherche en Aménagement et Développement, Laval University, Canada, walid.chaker@ift.ulaval.ca

**Yanguang Chen** Department of Geography, College of Urban and Environmental Sciences, Peking University, Beijing 100871, China, chenyg@pku.edu.cn

**Jing Chen** Beijing Geely University, Beijing, China

**DongMei Chen** Department of Geography, Queen's University, Kingston, ON K7L 3N6, Canada, chendm@queensu.ca

**Matt Duckhama** Geomatics Department, The University of Melbourne, Australia

**Michael F. Goodchild** Center for Spatial Studies, Department of Geography, University of California, Santa Barbara, CA, USA, good@geog.ucsb.edu

**Lauren Hand** Dewberry, Atlanta, GA, USA

**Mark W. Horner** Department of Geography, The Florida State University, Tallahassee, FL 32306-2190, USA, mhorner@fsu.edu

**Sungsoon (Julie) Hwang** Department of Geography, DePaul University, Chicago, IL 60614, USA, hwang.sungsoon@gmail.com

**Bin Jiang** Division of Geomatics, Department of Technology and Built Environment, KTH Research School, University of Gävle, SE-801 76 Gävle, Sweden, bin.jiang@hig.se

**Tony Joyce** Ordnance Survey of Great Britain, Southampton, UK

**Vassilis Kostakos** Madeira Interactive Technologies Institute, Funchal, Madeira, Portugal, vk@m-iti.org

**Patrick Laubea** Department of Geography, University of Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland, patrick.laube@geo.uzh.ch

**Nadav Levy** Geosimulation and Spatial Analysis Lab, Department of Geography and Human Environment, Tel-Aviv University, Ramat Aviv, 69978, Tel Aviv, Israel, nadavlev@post.tau.ac.il

**Wenbao Liu** Department of Geography, Queen's University, Kingston, ON K7L 3N6, Canada

**Yongmei Lu** Beijing Normal University, Beijing, China; Texas State University, San Marcos, TX, USA, yl10@txstate.edu

**Peter Luciani** Department of Geography, Queen's University, Kingston, ON K7L 3N6, Canada

**Deelesh Mandloi** ESRI, Inc., Redlands, CA 92373, USA, deelesh@gmail.com

**Michael Manyin** NASA-GSFC, SSAI, Inc., Greenbelt, MD, USA

**Karel Martens** Institute for Management Research, Radboud University Nijmegen, Nijmegen, HK 6500, The Netherlands, k.martens@fm.ru.nl

**Sara McLafferty** Department of Geography, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

**Dmitry Messen** Houston-Galveston Area Council, Houston, TX, USA

**Bernard Moulin** Department of Computer Science and Software Engineering, Laval University, Quebec G1V 0A6, Canada, bernard.moulin@ift.ulaval.ca

**Lan Mu** Department of Geography, University of Georgia, Athens, GA 30602, USA, mu@UGA.EDU

**Itzhak Omer** The Environmental Simulation Laboratory, Department of Geography and Human Environment, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel, omery@post.tau.ac.il

**Pietro Pagliardini** Pagliardini Rupi Andreoni and Gazzabin, Studio d'Architettura, Arezzo 52100, Italy

**Sergio Porta** UDSU, Department of Architecture, University of Strathclyde, Glasgow G4 ONG, UK

**Juval Portugali** Environmental Simulation Laboratory (ESLab), The Department of Geography and Human Environment, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel

**Nikos A. Salingaros** Department of Mathematics, University of Texas at San Antonio, San Antonio, TX 78249, USA, yxk833@my.utsa.edu

**Willis Shem** University of Tennessee-Martin, Martin, TN, USA

**Marshall Shepherd** University of Georgia, Athens, GA, USA, marshgeo@uga.edu

**Marius Thériault** Centre de Recherche en Aménagement et Développement, Laval University, Canada, marius.theriault@crad.ulaval.ca

**Jean-Claude Thill** Department of Geography and Earth Sciences, University of North Carolina at Charlotte, Charlotte, NC 28223, USA, jean-claude.thill@uncc.edu

**Jie Tian** Department of Geography, Queen's University, Kingston, ON K7L 3N6, Canada

**Dalia Varanka** Mid-Continent Geographic Science Center, U.S. Geological Survey, Rolla, MO 65401, USA, dvaranka@usgs.gov

**Fahui Wang** Department of Geography and Anthropology, Louisiana State University, Baton Rouge, LA 70803, USA

**David C. Wheeler** Harvard School of Public Health, Harvard University, Boston, MA, USA, wheelerdc@mail.nih.gov

**Michael J. Widener** Department of Geography, The Florida State University, Tallahassee, FL 32306-2190, USA

**Mike Worboys** National Center for Geographic Information and Analysis, University of Maine, Orono, ME, USA

**Xiaojun Yang** Department of Geography, Florida State University, Tallahassee, FL 32306, USA, xyang@fsu.edu

**Xiaobai (Angela) Yao** Department of Geography, University of Georgia, Athens, GA 30602, xyao@uga.edu

**Weihong Yin** Beijing Union University, Beijing, China

# Part I
# Introduction

# Chapter 1
# Geospatial Analysis and Modeling of Urban Structure and Dynamics: An Overview

**Bin Jiang and Xiaobai Yao**

## 1.1 GIS = GISystems, GIScience, GIServices and GIStudies

Geographic information research and technologies have experienced over four decades of development, from the mainframe to the workstation to the desktop, and to today's laptop and mobile devices. Every important GIS development is driven by a significant breakthrough of mainstream information technology. For example, the 1980s was characterized by the popularity of personal computers that were increasingly becoming affordable to university departments, governmental agencies and private sectors. Many university GIS programs were established during that time period, and the NSF-funded NCGIA played an important role in coordinating the development of course curriculum and related research activities. The next decade can be named the age of GIScience. GIScience is the science behind GISystens, dealing with fundamental questions raised by the use of GISystems and technologies (Goodchild 1997). It occurred at the time when the Internet and the World Wide Web started to change the way we led our lives and ran our businesses. It was the Internet and the Web that made the GIS community think of a service oriented approach to GIS, namely GIServices (Günther and Müller 1999). Instead of owning a GIS, end users can be served by GIS functionalities from a remote GIService center. GIServices aim to develop distributed or decentralized GIS to serve individuals and communities for spatial planning and decision making, as well as for their daily life. Another perspective of GIS is GIStudies for studying the impacts of geographic information and technologies on society.

The above GIS related terms reflect from a multi-dimensional perspective how GIS has evolved from a computer-based centralized system, to an internet-or-web-based decentralized service; from the technologically dominated view to the increasingly science oriented view, and to a broader societal perspective.

B. Jiang (✉)
Division of Geomatics, Department of Technology and Built Environment, KTH Research School, University of Gävle, SE-801 76 Gävle, Sweden
e-mail: bin.jiang@hig.se

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_1,

Along the technological dimension, GIS has evolved from initial 2D maps to 3D representations, from static maps to animated visualization, from stationary computers to mobile devices, and from being professionally oriented to catering to the general public. The continuous evolution of GIS is still ongoing. A significant driver is the search engine giant Google. The launch of Google Maps and Google Earth suddenly changed the general public's perception of GIS professionals. For example, nowadays it is much easier to explain what GIS is by simply referring to Google Maps. Indeed, Google Maps and similar online GIServices have democratized GIS and map making activities (Butler 2006) in an unprecedented manner. Hundreds of thousands of Google Maps mashups have been created to serve various professionals as well as the general public. All these are created under the rubric of Web 2.0, where individuals can voluntarily generate web content, which can be then again be shared by individuals. That said, no matter how GIS has been evolving, the core of GIS remains unchanged. As commented by the president of the leading GIS vendor ESRI, Dangermond (2003), … *the real heart of GIS is the analytical part, where you actually explore at the scientific level the spatial relationships, patterns, and processes of geographic phenomena, cultural phenomena, biological phenomena and physical phenomena. … this is the area that holds the greatest promise for creating insights into how our world works and how it is evolving, connecting and changes.*

The ever increasingly urbanized world has created various problems of environment, climate, consumption of resources, and public health, which are closely linked to the side-effects of urbanization such as sprawl, congestion, housing affordability and loss of open space. These problems need to be addressed at various levels of spatial planning and decision making. For instance, in urbanization management, special and specific consideration should be given to the relationships between land use, transportation, and the environment. Fundamental to the urban problems are two separate yet related issues: urban structure and urban dynamics. Both issues can be seen from physical and socio-economic perspectives. From the physical perspective, urban structure has its overall geometric shape, and the shape evolves and changes over the time. On the other hand, from the socio-economic perspective, urban systems demonstrate certain structures and dynamics. Thomas Schelling's well known segregation model illustrates the underlying mechanism of human movement from a social and ethical sense. More importantly, many of the components of urban structure and dynamics can be understood from an interweaving perspective involving both physical and socio-economic factors. For instance, human activities and mobility are intimately influenced by the spatial configuration of urban land use and transportation networks.

Geospatial analysis and modeling, combined with the powerful capability of GIS in data storage and visualization, have become important and indispensable tools for understanding urban structure and dynamics. Primarily, the geospatial tools contribute to obtaining useful information and knowledge from massive geographic information.

## 1.2 Individual-Based Data Capture for Modeling Urban Structure and Dynamics

Current GIS technologies have collected massive data for cities, and these data have been served to the public. Commonly cited examples are the above mentioned Google Maps and Google Earth that have integrated terabytes of satellite imagery, aerial photos, and GIS data. Many value-added services have been developed on platforms enabled by geographic information technologies. With the advancement of computer technologies, finer resolution geospatial data and individually-based geospatial data have become available for modeling urban structure and dynamics. *Itzhak Omer* in Chapter 2 presented one such case at the individual household level for examining residential segregation. Information communication technologies are providing unprecedented ways of mobile data collection (Giannotti and Pedreschi 2008). Nowadays, many urban residents carry mobile devices with built-in Bluetooth. It has been observed that about 7% of the mobile devices are set on, which provides a way to collect data about people's whereabouts. In this regard, Chapter 3 by *Vissilis Kostakos* developed a novel method for collecting traffic data in urban space, and provides a novel use of space syntax (Hillier and Hanson 1984) for the design and development of pervasive systems. Interestingly, the author developed a method of tracking people's whereabouts in virtual space (Kostakos and O'Neill 2008), providing an excellent example of studying human activities in both physical and virtual space.

A big problem emerging from Bluetooth detection is the invasion of privacy. In this regard, Chapter 4 contributed by *Patrick Laube and his colleagues* proposed the concept of decentralized spatial computing, which is able to safeguard privacy in the context of location based services. They developed some decentralized query algorithms for this purpose. In terms of data capture, it is worthwhile to mention volunteered geographic information created, assembled, and disseminated voluntarily by individuals (Goodchild 2007; Sui 2008). For example, OpenStreetMap (OSM), started in July 2004 by Steve Coast in London, is a wiki-like collaboration to create a free editable map of the world, using data from portable GPS devices, aerial photography and other free sources. Many value-added services like routing have been developed using OSM, and it provides a rich data source for studying urban structure and dynamics.

## 1.3 Modeling Urban Complexity and Hierarchy

Cities are complex systems that demonstrate a hierarchical structure. This is evident in Kevin Lynch's classic work – *the image of the city* (Lynch 1960). People capture those dominant city elements in mental representations, while a majority of city elements were filtered out in the process of perception and cognition. Not only city elements within a city, but also cities within a country or region are hierarchically organized. This has been well exemplified in city size ranking in the classic work by Zipf (1949). This topic has received a revival of interest in recent years (Pumain 2006) in research on complexity networks followed by the major discovery of small world and scale free networks (Watts and Strogatz 1998, Barabási and Albert 1999). Underlying the major discovery is network thinking, in which relationship between things is modeled from a topological perspective. This kind of topological perspective also underlies the phenomenal success of the Google search engine. Using crawling, the search engine captures the topological structure by modeling individual pages as nodes and hotlinks as links of a huge web graph, the world's largest graph. This kind of network thinking has been adopted to characterize the structure of urban street networks, which are considered to be self-organized in nature (Jiang et al. 2008). It has been found that a majority of streets are trivial, while a minority of streets are vital, and the minority of streets account for a majority of traffic (Jiang 2007, Jiang 2009). In essence, cities are much like biological entities, although they are man made. Along the line of urban complexity and hierarchy, many efforts have been made using an interdisciplinary approach (e.g., Salingaros 2005, Albeverio et al. 2008).

Under the same paradigm of network thinking, yet using agent based simulation modeling (Benenson and Torrens 2004, Batty 2005), *Efrat Blumenfeld-Lieberthal and Juval Portugali* in Chapter 5 developed an urban simulation model to study urban dynamics. The agents in the model are able to think globally and act locally, and eventually the local interactions of agents give rise to the global urban structure that in turn affects agents' behavior. The model succeeds in generating some interesting distributions like power law distribution of city size as observed in real city systems. Hierarchy is closely related to the spatial recursive subdivision and network structure, so mathematical formalization underlying urban hierarchy is of value to better understand urban systems. Chapter 6 by *Yanguang Chen* studied the hierarchy of cities that has a cascade structure. This structure can be described by a set of exponential laws. A significant finding of this study is that the hierarchy of cities in the large or middle scale complies with the $2^n$ rule, while the hierarchy at a small scale follows the $3^n$ rule. However, both the $2^n$ rule and $3^n$ rule are equivalent to the rank-size rule with exponent -1.

## 1.4 Simulating and Modeling Urban Transportation Systems

Urban transportation systems constitute the fundamental and critical part of urban structure and dynamics. Many urban problems, such as traffic congestion, energy consumption and environmental pollution, safety, and even security and emergencies, are largely related to the transportation systems. With the increasing population burden, we are facing an unprecedented challenge to renew or update our transportation systems in order to accommodate more mobility. An urban transportation system is too complex to be experimented on. On the other hand, the state-of-the-art computers are powerful enough to build up a mirror world to carry out simulations, by adding different what-if scenarios and testing their feasibility. This simulation technology is commonly known as agent-based modeling or simulation. Many agents interact and communicate with each other locally, and the interaction results in some interesting and surprising structure or patterns, namely emergence.

In recognizing the shortcomings of existing models of parking search, Chapter 7 by *Karel Martens, Itzhak Benenson and Nadav Levy* developed a non-spatial model of parking search and an agent-based model, namely PARKAGENT, for exploring parking behavior and dynamics. The authors used the two models to analyze the phenomena of searching for parking, and compared their results with a focus on the spatial effects on parking dynamics. They demonstrate some interesting outcomes that are of value for traffic engineering and for parking policies. *Walid Chaker, Bernard Moulin and Marius Theriault* in Chapter 8 introduced a multi-scale and multi-modal virtual urban environment that is combined with a synthetic population creation mechanism. It is able to generate various kinds of agents (e.g. pedestrians or vehicles) within a multi-scale and multi-mode transport network. Such a software platform can generate a massive amount of agents and is able to simulate massive trips within one minute. It makes an important step toward more comprehensive simulation and modeling of complex urban phenomena.

How urban streets are perceived by human beings and how human beings perceive the convenience of urban transportation systems constitute two basic questions of the next two papers. *Itzhak Omer and Bin Jiang* investigated city legibility and imageablity from the perspective of urban streets. They adopted a topological measure developed from Q-analysis (Atkin 1974) to characterize predominance of streets, and compared it with graph–theoretic measures. It provides an alternative yet complementary perspective toward the understanding of urban street networks. In Chapter 10 by *Yongmei Lu, Weihong Yin and Jing Chen*, transportation convenience of an urban transportation system is assessed from both subjective and objective perspectives. On the one hand, they examined how local residents perceive the transportation convenience subjectively. On the other hand, the transportation convenience is measured objectively from road capacity, public transit services and real traffic flow. Furthermore, they compare the subjective and objective assessment, and reveal some displacement between the patterns of transportation convenience. Such a study sheds substantial light on effective management and planning of urban transportations. *Deelesh Mandloi and Jean-Claude Thill* in Chapter 11 presented an object-oriented data model of multi-modal, and indoor/outdoor transportation networks, which is targeted for route planning and navigation as well as for other network analyses. This model integrates transportation both inside buildings (3D in nature yet using a 2.5D approach) and outside buildings (multi-modes in nature involving streets, sidewalks and public transit routes).

## 1.5 Analyzing and Modeling Urban Growth, Urban Changes and Impacts

For the first time in history, more than half of the world's population lives in urban areas. Remarkably, up to 80% of the populations are urban residents in North and Latin America as well as in Europe. This situation gives rise to an increasing need to research urban growth and its impacts on various aspects of urban structure and dynamics. Remote sensing (RS) imagery has been a major data source for analyzing urban growth and land use changes, for it continuously provides time series of imagery for tracking urban expansion and land use changes. Interestingly, the integration of GIS, RS, and dynamic models has been an essential part of geospatial analysis and modeling for exploration, simulation, prediction, and model verification. Chapter 12 by *Xiaojun Yang* reported a study in this direction. Based on a case study of Atlanta, the study illustrates that urban growth and land use change are highly correlated with population and economic growth, as well as accessibility conditions. In contrast to the previous work, *Dongmei Chen and her colleagues* in Chapter 13 took the Greater Toronto Area as a case study, and investigated the relationship between landscape change and population increase, as well as the ecological impacts of urbanization. They show that landscape change is negatively correlated with population growth.

Not only ecology, but also precipitation of urban area can be affected by urbanization. *Marshall Shepherd and his coauthors* (Chapter 14) presented a comprehensive review of studies on the "urban rainfall effect": their findings and methods focused on numerical modeling. A series of recommendations is offered to improve the modeling of the urban rainfall effect, and future research advances are indicated. Chapter 15 by *Dalia Varanka* presented a spatial model for assessing population pressure by interpolating a consumption variable using the Kriging method. Eventually, a surface trend is generated as a visualization tool that is useful to better understand urban structure and dynamics. *Xiaobai Yao* in Chapter 16 investigated the concept of place and specifically focused on cities as places. She presented a spatio–temporal model that links the spatial representations of cities with spatio–temporal processes. This model is built on top of a place ontology that distinguishes between the static view of places and the dynamic view of spatial–temporal processes. This model is further assessed for its potential use in spatio–temporal analysis, in particular in exploring spatio–temporal topological relations.

## 1.6 Studying Other Urban Problem Using Geospatial Analysis and Modeling

The set of papers presented in this section show that a diversity of urban problems can be studied using geospatial analysis and modeling. As indicated in the chapter title, *Pietro Pagliardini, Sergio Porta and Nikos Salingaros*

(Chapter 17) outline the relationship between geospatial analysis and urban planning and design. They argued that geospatial analysis should be used to extract morphological rules from traditional settlements, and the rules can be used to redevelop modern cities in order to generate living cities. They believe that predictive and dynamic spatial models can help in constructing the new discipline of "urban seeding" that will substantially re-frame the existing discipline of architecture and urban planning. Their convincing discussions lead to some interesting insights towards urban structure and dynamics regarding urban planning and design. In seeking possible associations between risk factors and late-stage breast cancer diagnosis, *Lan Mu, Fahui Wang and Sara McLafferty* (Chapter 18) adopted a modified scale–space clustering method that considers adjacency in both spatial and attribute dimensions to minimize the loss of information in the clustering process. Their method is proved to be effective and better than existing methods in avoiding the modified areal unit problem (MAUP) effect in spatial analysis.

Toward a better understanding of the intimate relationship between land use and transportation, *Sungsoon Hwang and Jean-Claude Thill* (Chapter 19) examined the impact of job accessibility on housing prices in both Buffalo and Seattle metropolitan areas. The authors adopted a hedonic regression modeling framework, and investigated the impacts from both local and global perspectives. It was found that job accessibility is positively associated with housing prices at the global scale, and that the impact of job accessibility on housing prices varies locally. In Chapter 20, *Mark Horner and Michael Widener* dealt with allocating relief goods by considering equity factor as well as efficiency. Their results showed that population heterogeneity has an important role in influencing appropriate facility placement. *David Wheeler* in Chapter 21 promoted a visual analytics approach to dealing with urban problems. The study is not designed to tackle any urban problem per se, although census undercount is used as an illustrative case study. In essence, this paper is about visualization tools for diagnosing correlation in estimated regression coefficients. This work fits well with visual approaches, or more precisely the visual analytics approach (Thomas and Cook 2005), in geospatial analysis and modeling.

## 1.7 Conclusion

We have tried in this opening chapter to elaborate on the role of geospatial analysis and modeling in understanding urban structure and dynamics. The chapters demonstrated how geospatial analysis and modeling in the context of GIS contribute to the study of various urban problems. We have seen how GIS in general, and geospatial analysis and modeling in particular, have been evolving in their capabilities to gain insights into urban systems. Such insights provide a significant input to spatial planning and policy making for urban development in the future. In terms of future development, the increasing availability of geospatial information through positioning technology, geosensor networks, and human volunteers is adding enormous challenge for geospatial analysts. In this regard, geospatial analysis and modeling need to seek more collaborations and contributions from GIS related fields such as spatial data mining, geovisualization, and uncertainty analysis.

## References


Albeverio, S., Andrey, D., Giordano, P. & Vancheri, A. (2008). *The dynamics of complex urban systems: an interdisciplinary approach*. (Berlin: Springer)

Atkin, R.H. (1974). *Mathematical structure in human affairs*. (London: Heinemann)

Barabási, A.-L. & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286, 509–512

Batty, M. (2005). *Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals*. (Cambridge, MA: The MIT Press)

Benenson, I. & Torrens, P. (2004). *Geosimulation: automata-based modeling of urban phenomena*. (New York: John Wiley & Sons)

Butler D. (2006). Virtual globes: the web-wide world. *Nature*, 439(7078), 776–778

Dangermond, J. (2003). Foreword. In P. Longley & M. Batty (Eds.), *Advanced spatial analysis: the CASA book of GIS* (pp. vii–viii). (Redlands, CA: ESRI Press)

Giannotti, F. & Pedreschi, D. (Eds.) (2008). *Mobility, data mining and privacy – geographic knowledge discovery*. (Berlin: Springer)

Goodchild, M.F. (1997). What is Geographic Information Science? NCGIA Core Curriculum in GIScience, http://www.ncgia.ucsb.edu/giscc/units/u002/u002.html, posted October 7, 1997.

Goodchild, M.F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), 211–221

Günther, O. & Müller, R. (1999). From GISystems to GIServices: spatial computing in the Internet marketplace. In M. Goodchild, M.J. Egenhofer, R. Fegeas & C. Kottman (Eds.), *Interoperability in geographic information systems*(pp. 427–442). (Boston: Kluwer Academic Publishers)

Hillier, B. & Hanson, J. (1984). *The social logic of space*. (Cambridge: Cambridge University Press)

Jiang, B. (2007). A topological pattern of urban street networks: universality and peculiarity. *Physica A: Statistical Mechanics and its Applications*, 384, 647–655

Jiang, B. (2009). Street hierarchies: a minority of streets account for a majority of traffic flow. *International Journal of Geographical Information Science*, 23.8, 1033–1048, Preprint, arxiv.org/abs/0802.1284

Jiang, B., Zhao, S., & Yin, J. (2008). Self-organized natural roads for predicting traffic flow: a sensitivity study. *Journal of Statistical Mechanics: Theory and Experiment*, July, P07008, Preprint, arxiv.org/abs/0804.1630

Kostakos, V. & O'Neill, E. (2008). Cityware: urban computing to bridge online and real-world social networks. In M. Foth (Ed.), *Handbook of research on urban informatics: the practice and promise of the real-time city* (pp. 195–204). (Hershey, PA: IGI Global)

Lynch, K. (1960). *The image of the city*. (Cambridge, MA: MIT Press)

Pumain, D. (2006). *Hierarchy in natural and social sciences*. (Dordrecht: Springer)

Salingaros, N.A. (2005). *Principles of urban structure*. (Delft: Techne)

Sui, D.Z. (2008). The wikification of GIS and its consequences; or Angelina Jolie's new tattoo and the future of GIS. *Computers, Environment and Urban Systems*, 32(1), 1–5

Thomas, J.J. & Cook, K.A. (Eds.) (2005). *Illuminating the path: the R&D agenda for visual Analytics*. National Visualization and Analytics Center. (Richland, WA)

Watts, D.J. & Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. *Nature*, 393, 440–442

Zipf, G.K. (1949). *Human behavior and the principles of least effort*, (Cambridge, MA: Addison Wesley)

# Part II
# Individual-Based Data Capture for Modeling Urban Structure and Dynamics

# Chapter 2
# High-Resolution Geographic Data and Urban Modeling: The Case of Residential Segregation

**Itzhak Omer**

**Abstract** The increasing availability of geographic data at high resolution and good quality has improved our ability to investigate a city's social geography thanks to the data's enhancement of our capacity to integrate three fundamental dimensions: residential distributions, the built-up environment's properties, and individuals' perceptions. This paper discusses the benefits and constraints of using such data in current approaches to the investigation and modeling of urban residential segregation. Based on several studies conducted on residential segregation in Jaffa, a mixed area in Tel-Aviv, we conclude that high-resolution data, especially at the house-level, has significant potential to enrich our understanding of the involvement of the built-up environment and individuals' spatial preferences in ethnic residential segregation. In addition, the constraints associated with using such data are discussed. On the methodological level, they refer to privacy considerations and cartographic presentation; on the conceptual level, they touch upon the gap between observed behavior, spatial preferences, and the dynamics of social residential segregation.

**Keywords** High-resolution spatial data · Urban modeling · Social residential segregation · Agent-based simulation models

## 2.1 Introduction

*All geographic data leaves its users, to some extent, uncertain about the nature of the real world* (Goodchild 2005, p. 14). This situation, which often stems from data resolution constraints regarding space, time and objects, has affected socio-geographic research of the urban environment. Until recently, socio-geographic research was limited to the use of large-scale aggregate data based on administrative areas, which frequently do not capture the micro-scale

I. Omer (✉)
Department of Geography and Human Environment, Tel Aviv University, Israel
e-mail: omery@post.tau.ac.il

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_2,

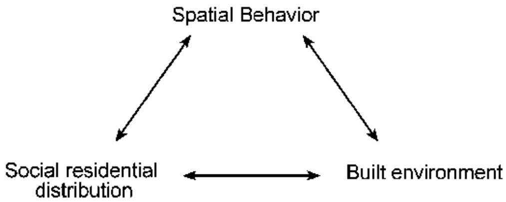



**Fig. 2.1** Illustration of the relationship between urban social geography's dimensions: spatial behavior – spatial cognitive representation, spatial preferences and the observed residential mobility of individuals; Social residential distribution – the spatial distribution of socio-demographic groups; Built environment – the spatial distribution of land uses (e.g. residence, commercial and public institutions) and morphological properties (e.g. street patterns and buildings' height)

situations where the three main dimensions of urban social geography – social residential distributions, the built-up environment's properties, and the individuals' spatial behavior – come together (Fig. 2.1).

The basic problem of using aggregate spatial data for geographical areas is the fact that the true distribution of social groups within the areas is unknown. As a result, it is difficult to investigate how objects in the built environment – such as building types, the street network, urban services, etc. – and specific local socio-spatial compositions affect residential mobility. It is also difficult, and sometimes even impossible, to combine aggregate with subjective data to better understand individual preference factors – for instance, social composition and environmental conditions – regarding residence choice.

This situation has nonetheless changed recently due to improvements in the construction of geographic databases and GIS technology, making it possible to obtain high-resolution data in urban locations. Such data actually invite the possibility of diminishing the gaps between the different kinds of data that represent the dimensions of social geography.

Two stages can be observed in this process. The first is collection of high-resolution geo-referenced data on the built environment by the location of urban objects, e.g., road networks or houses. These data have been available since the 1970s; they are currently available for almost any city in the western world. Precise estimation of spatial relations between urban objects is finally achievable. However, such data are, by themselves, insufficient to assess spatial behavior with respect to residential mobility when socio-economic and demographic attributes remain available only in aggregate form. The second stage involves the collection of high-resolution socio-demographic data. This stage is rather recent, having begun in the early 2000s with the release of individual geo-referenced data gathered by population censuses in several Western countries (Benenson and Omer 2003). These census databases contain the socio-economic attributes of householders, linked to house location.

The availability of geo-referenced house-level socio-demographic data in addition to built-up environmental data is also significant because the two types of data can be combined, at house-level, with subjective data that reflect perception and intention. This progress has important implications for empirical socio-geographic research. It is therefore now possible to more directly and reliably describe true social residential distributions, as well as investigate their relationship to the built environment and individuals' perceptions. As will be described later, this possibility supports new approaches for modeling the dynamics of social residential distribution.

Nevertheless, the use of high-resolution data comes with a price. Despite their potential, it is rather difficult to use these data in research, especially if they capture sensitive social attributes and personal attitudes requiring privacy considerations. The need to protect privacy has implications for access, analysis, and presentation. Furthermore, high-resolution data also raise methodological constraints when attempting cartographic presentation.

The aim of this paper is to discuss the benefits and constraints of using and integrating different kinds of high-resolution data in the investigation of social residential segregation in cities. In the next section, we examine these issues by focusing on the case of Jaffa, an area in Tel-Aviv. Use will be made of house-level data made available by the Israeli Census of Population and Housing for 1995, the first census containing geo-referenced socio-demographic data at the level of householders and houses. Based on a discussion of this case, the second part will turn to the potential of these high-resolution data for current approaches to the modeling of urban residential segregation.

## 2.2 The Potential Implications of High Resolution Data for Socio-geographic Research

The aim of this section is to discuss the benefits and constraints of using high-resolution spatial data when investigating residential segregation within the framework of socio-geographic research. Special attention will be given to the possibility of integrating different aspects of this process. For this purpose, the discussion will concentrate on social residential segregation in Jaffa. Jaffa is an area of about 7 km$^2$; its population in 1995 was about 39,000, with the Jewish majority representing about 70% and the Arab minority the other 30% of the population. The data for the study were obtained from several sources: a detailed geo-referenced house-level socio-demographic survey conducted by the Israeli Census of Population and Housing in 1995 (ICBS 2000), detailed infrastructure data and subjective data gathered through in-depth interviews, surveys, and field observation conducted in the area (for more details see: Omer 1996, 2003).

The use of high-resolution spatial data in socio-geographic research has several advantages. The first and most direct advantage is as a description of social residential distributions at the house-level. This capacity prevents – or at least reduces – two basic methodological problems arising form the employment of aggregate data in socio-geographic research, especially when the research deals with local phenomena. This first problem is the modifiable areal unit problem (MAUP) – different results are obtained for the same area when using aggregate data sets at different scales or with different geographical partitions (Green and Flowerdew 1996; Openshaw and Rao 1995). The problem becomes especially significant when the target group's distribution does not correspond with the geographical units of analysis. MAUP has been extensively deliberated in the social residential segregation literature with respect to the divergent conclusions reached regarding the degree of residential segregation found in a selected area (see for example: Benenson and Omer 2003; Omer and Benenson 2002; Wong 2003). The second problem is the "ecological fallacy," which occurs when conclusions about individual spatial behavior are inferred from aggregate data, a conceptual leap that impacts on the reliability of a study's results (Wrigley et al. 1996).

Figure 2.2 illustrates the advantages of using high-resolution data by comparing administrative statistical areas with house-level aggregation during graphic presentation of ethnic residential distributions in Jaffa. This comparison illustrates a situation where the residents of two spatially separated groups of houses, almost exclusively populated by Arabs or Jews, appear to be living in an ethnically mixed statistical area when using aggregate data exclusively. That is,

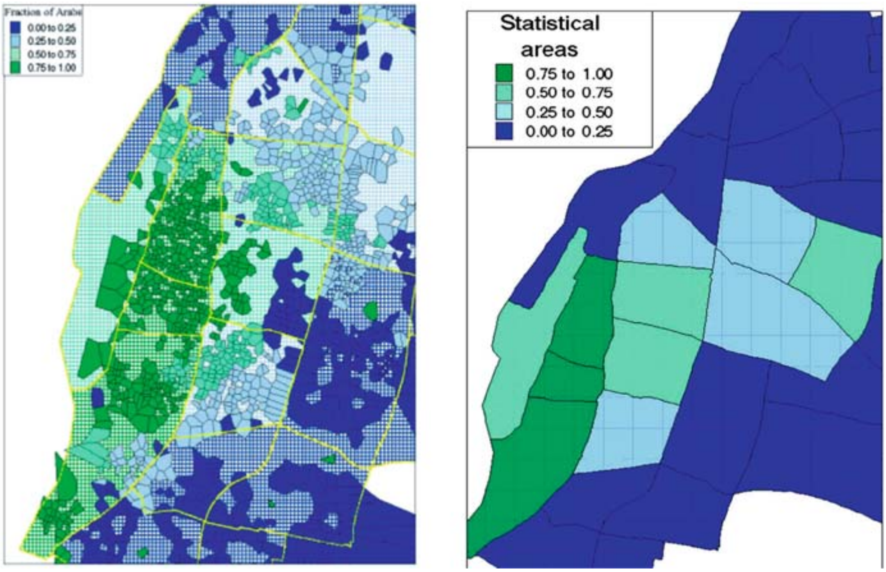



**Fig. 2.2** The Arab-Jewish residential distribution in Jaffa, Tel-Aviv, at the resolution of administrative statistical areas (*right*) and at house-level (*left*)

the level of segregation of the Arab and Jewish population in Jaffa is much higher than one can conclude on the basis of aggregate data arranged according to administratively defined statistical areas.

The second advantage of high-resolution spatial data is the ability to observe the relationship between social residential distribution and land use distributions in a city. Such an investigation is difficult to be conduct with aggregate data because such data are based on the relationship between an average representative of a designated social group in a given area and the average value of land uses in that area. Alternatively, house-level socio-demographic and physical data enable us to locate the true relationships between these variable in a given urban location and thus understand how environmental conditions affect residential segregation.

To illustrate this effect, Fig. 2.3a shows the Arab community's residential pattern in Jaffa. This pattern reveals diffusion at a rate that varies by direction, beginning from the center (Ajami) outward. Figure 2.3b presents a high-resolution view that includes several Jaffa neighborhoods, each characterized by a different rate of population change. High-resolution data enables exploration of the different factors active in the built urban environment that may

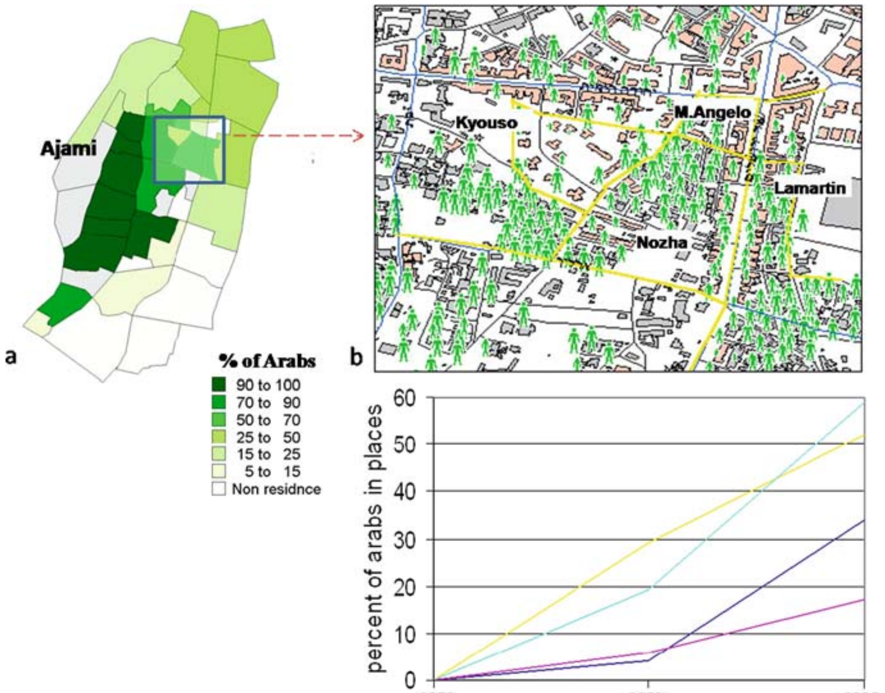



**Fig. 2.3** Rate of change in population composition in adjacent geographic units, Jaffa

encourage divergent forms of residential segregation. In the Jaffa case, for example, several factors affect the willingness of the Jewish and Arab populations to live in the same neighborhood; they include the availability of oriental-style houses and modern block buildings (the first building's architectural style is attractive for the Arab population while the second is attractive for the Jewish population), a high proportion of commercial and public land uses, which prevent identification of the urban environment with one of the group (Omer 2003).

The third advantage of high-resolution data is its capacity to integrate subjective data about individual intentions and perceptions with detailed objective data on population and infrastructure for the purpose of constructing individual preference indices. One of the more salient types of data regarding residential segregation is data on perceived neighbourhood territory, i.e., subjectively defined neighborhood territories (Pacione 1983; Brower 1996). Perceived neighbourhood territory plays a major role in individual residential decisions, yet, at the same time, is affected by that group's residential distribution (Omer 2007). This reciprocal effect can be illustrated by the findings from Jaffa. It was found that the perceived neighbourhood boundaries were highly correlated with the spatial units when defined by the changes in ethnic composition in these areas i.e., the neighbourhoods' perceived boundaries cut across areas that significantly differ in their ethnic composition (Omer 2003, 2007; Omer and Benenson 2002).

The fourth advantage of high-resolution data is the possibility of investigating relationships between fine-scale socio-geographic phenomena and macro-scale patterns. Local-global relations are considered to be the driving force behind social system dynamics in contemporary social-geography – i.e., structuration theory and the realist approach (Crang and Thrift 2000) – as well as in current approaches to modeling urban system dynamics (Batty 2005).

Batty (2000, p. 483) has pithily described the potential of fine-scale data for investigating local-global relations:

> Until now, ... [we have] never been able to develop theory of sufficient generality to illustrate how local decisions and actions are consistent with and determine large-scale urban structures. All this is changing. Quite suddenly, so it appears, a new kind of fine-scale geography is beginning to emerge from data which are sufficiently intensive to detect detailed patterns and morphologies but also sufficiently extensive to enable these patterns to be generalized to entire metropolitan areas.

Notwithstanding their theoretical and practical potential, house-level data does raise several methodological constraints that are creating new challenges for socio-spatial research. I refer here to the two main constraints discussed in previous research: privacy considerations and cartographic presentation.

In any presentation of sensitive public data, the need to safeguard privacy soon arises (Sabourin and Dumouchel 2007; Tamisier et al. 2005). Spatial data on a household's socio-economic characteristics – e.g., income, education, and ethnic identity – are clearly very sensitive and obligate the imposition of privacy considerations when gathering, analyzing, and presenting spatial information. This need has direct implications for access to sensitive geo-referenced socio-demographic census data in research and practice. For example, Israeli law restricts the use of geo-referenced census data on householders. Access to house-level data was conducted under the personal supervision of Israel Central Bureau of Statistics (ICBS) staff.

However, the constraint on the presentation of house-level data does not stem from privacy considerations only; cartographic constraints also come into play. Mapping house-level data over a large area prevents viewers from discerning local- spatial variance because these distributions create high densities of symbols. Alternatively, mapping such data on a detailed scale creates discontinuous, sparse coverage of buildings, which creates difficulties for viewers interested in distinguishing the patterns identified (Benenson and Omer 2003).

Previous studies have suggested overcoming these cartographic constraints by using thematic maps produced at different scales, based on local indices of spatial association (LISA) (Benenson and Omer 2003; Omer 2005). LISA is a localized version of the global autocorrelation statistics resulting from comparison of the characteristics of a spatially located object and its neighbors (Anselin 1995). Applying these measures helps the observer identify spatial variance and small clusters in their spatial socio-demographic distribution (Talen and Anselin 1998).

To illustrate this approach's contribution, we present here an example of the use of the local Geary index $K_1$ on the Arab-Jewish residential distribution in Jaffa (Benenson and Omer 2003). This index estimates the characteristic's variance within a neighborhood. In practice, the values of the local index are calculated for every building in an area over neighborhoods of increasing sizes (order 1, 2, etc.). To make the patterns visually comprehensible, while complying with privacy restrictions, they are based on Voronoi coverage of buildings, where each polygon of a Voronoi partition corresponds to a building and its surrounding area (Benenson and Omer 2003).

Figure 2.4 shows maps of the Geary index for neighborhoods at order 1 and order 8. The maps help to identify general spatial patterns and spatial variances, i.e., homogeneous and mixed Arab and Jewish spatial concentrations (the $K_{1i,n}$ is always positive; the higher its value, the higher the neighborhood's heterogeneity). Although employment of such measures can be quite beneficial and even crucial when mapping large geographical areas, or when the data is highly sensitive, they imply aggregation, which demands concessions to accuracy. These costs and benefits are illustrated in Fig. 2.4, which shows that higher-order patterns are clearer and therefore promote better identification of ethnic clusters. Nonetheless, they forego the accuracy associated with house-level data while forsaking the possibility of understanding and explaining the factors associated with residential segregation.

Two additional methodological constraints also characterize detailed reference data: timeliness and gaps in data. Although updated house-level demographic data can be retrieved from other sources, including municipal

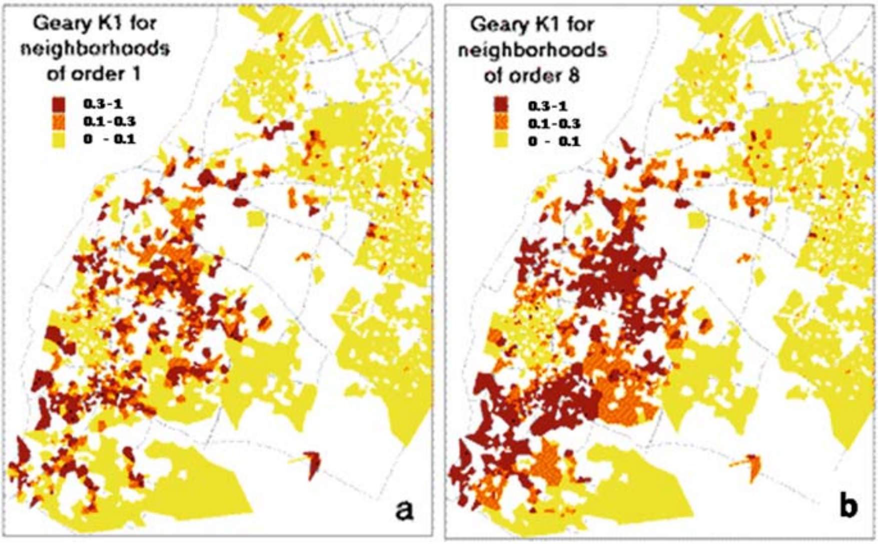



**Fig. 2.4** Thematic maps of the local Geary Ki,1 index constructed for Jaffa – ethnic (Arab–Jewish) distributions on the basis of (**a**) neighborhoods of order 1 and (**b**) of order 8

authorities, the national census is the only source for comprehensive data on social characteristics. The time lapse between censuses, generally a decade, introduces inconsistencies when using the data to describe changes in residential distributions. They may therefore not be relevant for operative planning. In addition, gaps in the data, a well-recognized phenomenon in socio-geographic research, can be very significant at house-level. In general, as much of the data is found at higher resolutions, it is less accessible and less frequently updated; privacy and cartographic issues thus become all the more acute. In the following section, the implications of using high-resolution data for residential segregation modeling are discussed.

## 2.3 Simulation of Social Residential Segregation

Due to the improvements in geographic database construction, access and use, as well as the power of computers, urban modelling now has the ability to simulate the simultaneous behaviour of large numbers of individuals within an urban environment. However, beyond improvements in computer and geographic information technology, the increasing availability of high-resolution spatial data does impact the characteristics of urban models and their development. This is not surprising because the data's resolution has implications for all aspects of the model's components: geographical partitions, objects, variables, parameters, and logical–mathematical structures. Here we focus on the potential influence that the increasing availability of high-resolution data has on residential segregation models.

The story of computerized dynamic models of residential segregation began with Morrill's diffusion model, which simulated Afro-American residential patterns in Seattle, Washington (Morrill 1965). This model was built on the Monte Carlo simulation (Hagerstrand 1952, 1967). Followed Morrill's work, several diffusion models were constructed to simulate the expansion of other "Afro-American ghettos" in the US; two major examples are the work of Hansell and Clark (1970) in Milwaukee and of O'Neill (1985) in Chicago. This model was also applied by Woods (1980) to simulate the dynamics of Asian and West Indian population distributions in UK cities. All these models were based on the ecological invasion–succession mechanism, in which the minority population is active, while the majority is reactive to events. In contrast, Rose (1970) suggested an alternative simulation model for Milwaukee that was based on the filtering-down mechanism, where the majority population is active and the minority population is passive.

These studies report successful simulations of minority residential expansion patterns. Yet, as Robinson points out:

> ...works employing Monte Carlo simulation reveal that neither the theoretical nor practical contribution of the approach has been great. Despite the early enthusiasm generated by the technique, an understanding of the exact causes of residential segregation is little nearer.... The reasons for this are many and varied, but center on the inherent weakness of the technique and its injudicious employment (Robinson, 1981, p. 143).

Spatial data resolution contributed significantly to this "inherent weakness of the technique". These models suffered from substantial deviations between the aggregate data regarding infrastructure (e.g., fractions of land usage, types of dwellings, average prices), population (e.g., fractions of socioeconomic or cultural groups), and the behavior simulated. This feature, which is actually a computerized version of the ecological fallacy (Klosterman 1994), was also criticized by Lee (1973) in his seminal paper "Requiem for Large-Scale Models". For the same reason, it was also difficult to estimate the validity of the model's definition at a given geographic scale. Yet, the most cutting criticism of the residential segregation diffusion model was directed at "employment of the approach for inappropriate and unachievable goals" (Robinson 1981, p. 159). Robinson was in fact referring to the use of simulation models for improving knowledge on the causes of residential segregation.

The next wave of urban modeling approaches was populated by the Cellular Automata (CA) models that arose during the 1990s. These models require that urban infrastructure be represented by means of a high-resolution cellular space. Each cell is assumed to be in one of several states that can change over time, according the state of the cell itself and the states of the neighboring cells. Similar to diffusion models, the cells in real CA models represent aggregate data.

Despite their extensive implementation in modeling the entire range of urban system dynamics, including those for real cities (Batty and Xie 1997; Clarke et al. 1997; Dietzel and Clarke 2006; Stevens et al. 2007; White and Engelen 2000; White et al. 1997), few attempts were made to model residential segregation (O'Sullivan 2002; Torrens 2007). This occurred because strict, classical CA does not consider – at least not explicitly – the dynamics of individuals with different goal–oriented behavior and interactions carried out among themselves (Benenson and Torrens 2004). In addition, due to their aggregative character, CA models suffer from the same limitations as the diffusion models discussed above. Thus, with respect to residential segregation, no significant development occurred until the recent wave of Agent-Based Simulation models (ABS).

Unlike CA models, ABS models explicitly refer to spatial behavior by mobile human entities that make decisions regarding spatial relations and environmental properties. That is, in addition to the cellular infrastructure or objects layer that represents infrastructure, an additional layer of agents move over the infrastructure layer according to its content while influencing it in turn. These characteristics allow easy and more "natural" simulation of residential segregation (Benenson 1998; Omer 1999; Portugali et al. 1994) as well as other instances of urban spatial behavior (Batty 2005; Benenson and Torrens 2004).

In ABS modelling, we directly interpret individual spatial choices in terms of the agents' tolerance (or preference) for specific social compositions. This allows us to explore the implications of our assumptions on the residential patterns, including novel patterns and bifurcations, which emerge at the macro level. Due to these characteristics, this modelling approach has been found to be convenient for studying the dynamics of residential segregation on the theoretical level, such as the hypothesized effect of neighbourhood size (Laurie and Jaggi 2003), minority group's internal structure (Omer 2005) and other demographic and behavioural parameters (Benenson 1998; Epstein and Axtell 1996; Flache and Hegselmann 2001; Hegselmann and Flache 1998; Omer 1999; Portugali et al. 1997; Portugali 1999). We should note that these studies are conceptually related to Schelling's model (1969, 1978), which can be viewed as the first non-computerized version of ABS modelling (Gilbert and Troitzsch 1999).

ABS offers new options for real-world modeling of residential segregation. First, its units are not identical with the $100 \times 100$ – $300 \times 300$ cells of the rectangular grid but with actual houses, an element that makes the model's infrastructure appropriate for implementing observation of real-world human behavior. Second, this kind of model explicitly refers to relationships between the spatial behavior of individuals – householders, businesses, public services, and so forth – and the urban environment. Third, the models enable us to directly account for a variety of changes in population and built environment objects. Fourth, on an operational level, ABS models respond well to object-oriented paradigms for constructing modeling software (Torrens and Benenson 2005); in addition they interface relatively seamlessly with raster-based and vector-based GIS. Similarly, relational conventions are well-matched with object-oriented database-management systems and entity-relationship models (Benenson and Torrens 2004; Torrens 2007).

These qualities, which overcome the restrictions of CA as well as traditional diffusion models, make a significant difference for urban modeling. The application of ABS modeling to real urban systems has been delayed, however, largely because they are based on geo-referenced infrastructure and socio-demographic data that have become available only recently.

The availability of geo-referenced socio-demographic date from Israel's *Census of Population and Housing, 1995* enabled us to construct a real-world agent-based model. The model was designed to simulate the dynamics of Arab-Jewish residential distribution for the period 1955–1995 (Benenson et al. 2002). In this model, Jewish and Arab families (the latter includes internal differences between Muslims and Christians) are represented by agents that change their residence locations according their ethnic composition. That is, the probability of an agent choosing a new residential location increases when the agent's ethnic identity is similar to the ethnic identity of other agents in the respective neighborhood and selected features of the built environment.

Estimates of the parameters and the model scenarios were based on a comprehensive experimental study of residential distribution in Jaffa (Omer 1996). For example, this study indicated a strong tendency for the Arab population to reside in oriental-style houses in comparison to the Jewish population, which prefers modern block buildings. Accordingly, a building's architectural style influences the agents' residential decisions in the model as well. In addition, due to the possibility of integrating subjective data gathered through intensive field study, we were able to use the neighborhood's territory as perceived by the two populations to define neighborhood boundaries in the model. Operationally, we defined the neighborhood by using Voronoi coverage constructed for the layer of house foundations, with two houses considered neighbors if their Voronoi polygons shared a common edge. Migration into and out of Jaffa was also included in the model, based on empirical data. The simulation results were then compared with detailed geo-referenced population data obtained from the Israeli's 1995 population census.

The simulation began with Arab and Jewish agents randomly distributed over the buildings in areas that had been populated by Arabs in 1995 by a ratio of 1:2 (Arabs to Jews). The model succeeded in simulating the experimental data and produced the segregated spatial distribution observed in Jaffa in 1995 (Benenson et al. 2002). Figure 2.5 shows the observed residential distribution in 1995 and the simulation output for the corresponding period. The fairly good approximation of Jaffa's residential dynamics during that period permitted review of a variety of scenarios, each based on different assumptions regarding the degree of tolerance between Jewish and Arab residents in the area as well as to study their consequences. These experiments indicated that the best fit was obtained with the scenario in which Arabs and Jews maintained asymmetric relations: Jews are intolerant toward Arabs whereas Arabs were almost neutral toward Jews. These preferences were maintained over the entire model period (40 years). In addition, the model was robust to quantitative changes in parameters and worked well independently of other parameters, such as

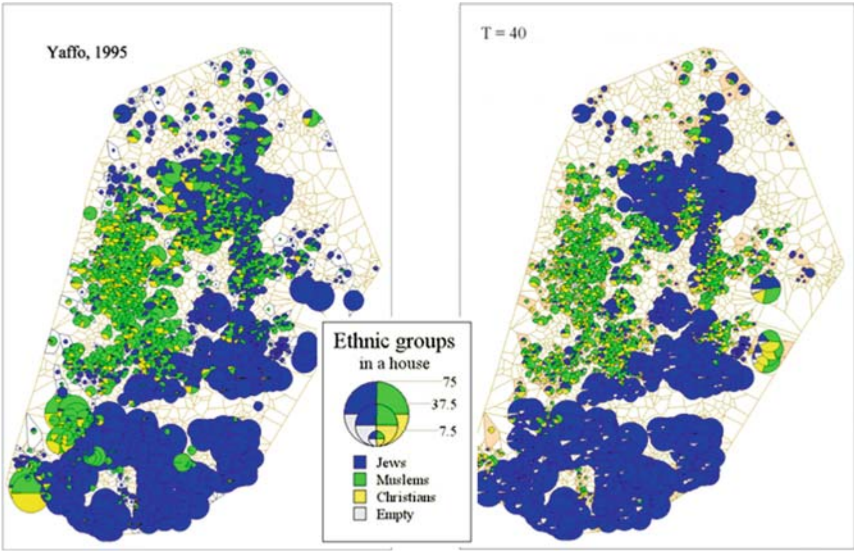



**Fig. 2.5** Dynamics of residential distribution in Jaffa: Simulated (at year 40) and observed residential distributions (year 1995) at house-level (in a scenario of asymmetric relations)

demographic and income variables. Thus, the ABS model of Jaffa's residential patterns clearly demonstrates which features of agent (householder) behavior are of primary importance for understanding residential segregation in the area.

## 2.4 Discussion and Conclusions

The increasing availability of geographic data at high resolution and good quality has improved our ability to investigate a city's social geography due to the capacity to integrate between three fundamental dimensions: social residential distributions, the environment's built properties, and individual perceptions. As the case of ethnic residential segregation shows, availability of geo-referenced data capturing these dimensions at the same level – the house-level – is essential for identifying and understanding its main components.

This possibility enhances our ability to model the dynamics of real social residential distributions with the Agent-Based Simulation (ABS) approach. Based on our experience with residential segregation data for Jaffa, we can conclude that the capabilities provided by ABS modeling improve the simulation of real socio-spatial processes in an urban system. This ability can contribute to our understanding of the involvement of the built environment as well as individual spatial preferences in creating ethnic residential segregation. It can also be employed for formulating relevant housing policy.

Even if the application of ABS modeling is limited for now due to its dependence on good quality, high-resolution data of infrastructure objects, and human agents, it is reasonable to assume that we are only at the beginning of the model's broader applications, which are expected to expand with the data's increasing availability.

Yet, one should be aware of the limitations of this approach. On the one hand, high-resolution databases supply researchers with accurate information on micro-scale socio-geographic contexts, a capacity that critically improves our ability to evaluate a model's parameters. On the other hand, the same databases still do not consider the implications of the observed residential behavior, nor the intentions and preferences of decision makers. It is well known, as indicated by the Schilling model, too, that we cannot infer individuals' intentions solely from their observed spatial distribution. Therefore, the gap between subjective and objective data remains the "Achilles heel" of ABS modeling.

Another constraint to be contended with is the gap between individual data and the dynamics of residential segregation. Namely, the availability of high-resolution data on spatial phenomena may be insufficient to explain why the dynamics of these events. Sayer (1984, p. 108) expressed the limitation of data disaggregation quite succinctly:

> It is often assumed that a useful way of understanding a complex object is to break it down into its constituent parts . . . or "disaggregate" the aggregate statistics in the hope that complexity and regularity might be reduced. . . . Many researchers have been seduced by the simple idea that if only individuals and their attributes, etc., were understood, the macro patterns of society would become intelligible. But it is not always so straightforward. . . . In such cases objects are said to have "emergent powers", that is, powers or liabilities which cannot be reduced to those of their constituents.

Residential segregation, like any other complex socio-geographic phenomenon results, after all, from self-organizing processes composed of a huge amount of interactions. Hence, despite the advantages of real ABS modeling, we still need to face the traditional scientific limitations inherent in the modeling method.

## References

Anselin, L. (1995). Local indicators of spatial association – LISA. *Geographical Analysis*, 27 (2), 93–115

Batty, M. (2000). The new urban geography of the third dimension. *Environment and Planning B*, 27, 483–484

Batty, M. (2005). Cities *and complexity: understanding cities with cellular automata, agent-based models, and fractals*. (Cambridge, MA: MIT Press)

Batty, M. & Xie, Y. (1997), Possible urban automata. *Environment and Planning B*, 24, 175–192

Benenson, I. (1998) Multi-agent simulations of residential dynamics in the city. *Computers, Environment and Urban Systems*, 22, 25–42

Benenson, I. & Torrens, P. M. (2004). *Geosimulation: automata-based modeling of urban phenomena*. (London: Wiley)

Benenson, I., Omer, I. & Hatna, E. (2002). Entity-based modeling of urban residential dynamics: the case of Yaffo, Tel Aviv. *Environment and Planning B: Planning and Design*, 29, 491–512

Benenson, I. & Omer, I. (2003). High-resolution census data: a simple way to make them useful. *Data Science Journal*, 2(26), 117–127

Brower, S. (1996). *Good neighborhood:A study of in-town and suburban residential environments*. (Westport, CT: Praeger)

Clarke, K., Hoppen, S., & Gaydos, L. (1997). A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area. *Environment and Planning B*, 24, 247–261

Crang M. & Thrift N.J. (2000). *Thinking space (critical geographies)*. (London: Routledge)

Dietzel, C. & Clarke, K. (2006). The effect of disaggregating land use categories in Cellular automata during model calibration and forecasting. *Computers, Environment and Urban Systems*, 30(1), 78–101

Epstein, J.M., & Axtell, R. (1996). *Growing artificial societies*. Cambridge, MA: MIT Press

Gilbert, N.S. & Troitzsch, K.G. (1999). *Simulation for the social scientist*. (Buckingham: Open University Press)

Goodchild, M.F. (2005). GIS, Spatial Analysis, and Modeling Overview. In D. Maguire, M. Batty & M.F. Goodchild (Eds.), *GIS, Spatial Analysis, and Modeling* (pp. 1–17). (Redlands: ESRI Press)

Green, M, & Flowerdew, R. (1996). New evidence on the modifiable areal unit problem. In P. Longley & M. Batty (Eds.), *Spatial analysis: modeling in a GIS environment*(pp. 41–54). (Cambridge: Geoinformation International)

Hagerstrand, T. (1952). The propagation of innovation wave. *Lund Studies in Geography B, Human Geography*, 4, 3–9

Hagerstrand, T. (1967). *Innovation diffusion as a spatial process*. (Chicago: University of Chicago Press)

Hegselmann, R., & Flache, A. (1998). Understanding complex social dynamics: A plea for cellular cutomata based modeling. *Journal of Artificial Societies and Social Simulation, 1*, http://www.soc.surrey.ac.uk/JASSS/1/3/1.html

Hansell, C.R. & Clark, W.A.V. (1970). The expansion of the negro ghetto in Milwaukee: a description and simulation model. *Tijdschrift voor Economische en Sociale Geografie*, 61, 267–277

Hegselmann, R. & Flache, A. (1998). Understanding complex social dynamics: a plea for cellular automata based modeling. *Journal of Artificial Societies and Social Simulation*, 1(3)

Israeli Central Bureau of Statistics (ICBS) (2000). *The 1995 census of population and housing, No. 8*, Socio-economic characteristics of population and households in localities and statistical areas. Jerusalem: State of Israel, Central Bureau of Statistics Publications.

Klosterman, R.E. (1994). Large scale urban models: retrospect and prospect. *Journal of the American Planning Association*, 60(1), 3–7

Laurie, A.J. & Jaggi, N.K. (2003). Role of `vision' in neighborhood racial segregation: a variant of the Schelling segregation model. *Urban Studies*, 40, 2687–2704

Lee, D.B. (1973). A requiem for large scale modeling. *Journal of the American Institute of Planners*, 39(3), 163–178

Morrill, R.L. (1965). The negro ghetto: problems and alternatives G *eographical Review*, 55, 339–369

Omer, I. (1996). *Ethnic residential segregation as a structuration process* (in Hebrew) Unpublished Ph.D. Thesis, Tel-Aviv University

Omer, I. (1999). Ethnic residential segregation and demographic processes. *Discrete Dynamics in Nature and Society*, 3(2–3), 171–184

Omer, I. (2003). The Jewish-Arab residential distribution in Yaffo, 1961–1995, *Merhavim* Studies in the Geography of Israel and the Middle East 3 (in Hebrew), Tel Aviv University

Omer, I. (2005). How ethnicity influences residential distribution: an agent-based simulation. *Environment and planning B: Planning and Design*, 32, 657–672
Omer, I. (2007). A GIS framework for evaluating the effect of social homogeneity on perceived neighborhoods. *AGILE: Geographic Information Science*, 204–212
Omer, I. & Benenson, I. (2002). Investigating fine-scale residential segregation by means of local spatial statistics.*Geography Research Forum*, 22, 41–60
O'Neill, W. (1985). Estimation of a logistic growth and diffusion model describing neighbourhood change. *Geographical Analysis*, 389–397
O'Neill, W.D. (1985). Estimation of logistic growth model describing neighborhood change. *Geographical Analysis*, 13, 391–397
Openshaw, S. & Rao, L. (1995). Algorithms for re-engineering 1991 census geography. *Environment and Planning A*, 27, 425–446
O'Sullivan, D. (2002) Towards micro-scale spatial modeling of gentrification. *Journal of Geographical Systems*, 4, 251-274
Pacione, M. (1983). The temporal stability of perceived neighborhood areas in Glasgow. *The Professional Geographer*, 35(1), 66-72
Portugali, J. (1999). *Self-organization and the city*. (Berlin: Springer)
Portugali, J., Benenson, I. & Omer, I. (1994). Socio-spatial residential dynamics: stability and instability within a self-organized city. *Geographical Analysis*, 26(4), 321–340
Portugali, J., Benenson, I. & Omer, I. (1997). Spatial cognitive dissonance and sociospatial emergence in a self-organizing city. *Environment and Planning B*, 24, 263–285
Robinson, V. (1981). Segregation and simulation: a re-evaluation and case study. In P. Jackson & S. Smith (Eds.), Social *interaction and ethnic segregation*. (London: Academic Press)
Rose, H.M. (1970). The development of an urban sub-system: The case of the Negro Ghetto. *Annals Association of American Geographers*, 60, 1–17
Sabourin, M. & Dumouchel, B. (2007). Canadian national consultation on access to scientific research data. *Data Science Journal*, 6, 26–35
Sayer, A. (1984). *Method in social science*. (London: Hutchinson)
Schelling, T.C. (1969). Models of segregation. *American Economic Review*, 59(2), 488–493
Schelling, T.C. (1978). *Micromotives and macrobehavior*. (New York: WW Norton)
Skjaeveland, O. & Garling, T. (1997). Effect of interactional space on neighbouring. *Journal of Environmental Psychology*, 17(3), 181–198
Stevens, D., Dragicevic, S. & Rothley, K. (2007). *iCity*: A GIS-CA modelling tool for urban planning and decision making. *Environmental Modelling & Software*, 22, 761–773.
Talen, E. & Anselin, L. (1998). Assessing spatial equity: an evaluation of measures of accessibility to public playgrounds.*Environment and Planning A*, 30, 595–613.
Tamisier, T., Felt, F. & Riegel, S. (2005). Socioeconomic database online.*Data Science Journal*, 4, 159–164
Torrens, P. (2007). A geographic automata model of residential mobility. *Environment and Planning B: Planning and Design*, 34, 200–222
Torrens, P.M. & Benenson, I. (2005). Geographic automata systems. *International Journal of Geographical Information Science*, 19, 385–412
White, R. & Engelen, G. (2000). High-resolution integrated modelling of the spatial dynamics of urban and regional systems. *Computers, Environment and Urban Systems*, 24(5): 383–400
White, R., Engelen, G. & Uljee, I. (1997). The use of constrained cellular automata for high-resolution modeling of urban land-use patterns.*Environment and Planning B*, 24, 323–343
Wong, D.W.S. (2003). Spatial decomposition of segregation indices: a framework toward measuring segregation at multiple levels. *Geographical Analysis*, 35(3), 179–184
Woods, R.I. (1980). Spatiotemporal models of ethnic segregation and their implications for housing policy. *Environment and Planning A*, 13, 1415–1433
Wrigley, N., Holt, T., Steel, D., & Tranmer, M. (1996). Analyzing, modelling, and resolving the ecological fallacy. In P. Longley & M. Batty (Eds.), *Spatial analysis: modelling in a GIS environment* (pp. 25–40), (Cambridge, UK: Geoinformation International)

# Chapter 3
# Space Syntax and Pervasive Systems

**Vassilis Kostakos**

**Abstract** In this paper we describe our novel use of space syntax for the design and development of pervasive systems. Pervasive systems are computer systems that are designed to be "invisible" to users because they are designed to blend in with their environment and become part of the fabric of everyday life. Pervasive systems consist of fixed, mobile and embedded components, each of which may entail interactive capabilities and intelligence. Due to their close relationship with the build environment, pervasive systems are an ideal domain for adopting a space syntax methodology. The contribution of this paper is two-fold. First, we present an adaptation of the space syntax methodology aimed at researchers and designers of space and pervasive systems. By developing an adaptation of the space syntax methodology we intend for space syntax to be added the arsenal of tools and theories that researchers use to understand and design pervasive systems. To exemplify how this can be achieved, we present three case studies that demonstrate crucial ways in which space syntax analysis can aid the design of pervasive systems. The case studies presented here show how space syntax can be used in the development of pervasive system as an application development tool, as an exploratory tool, and as a modelling tool.

**Keywords** Space syntax · Pervasive and ubiquitous computing · Observation · Methodology · Applications

## 3.1 Introduction

The relationship between computers and space is a topic that has persisted throughout the various advances in computer science, as early as when the first Multi User Dungeons (MUDs) were developed (Muramatsu and Ackerman

V. Kostakos (✉)
Madeira Interactive Technologies Institute, Funchal, Madeira, Portugal
e-mail: vk@m-iti.org

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_3,

1998). With the advent of personal computers, researchers began considering the relationship between physical space and "digital space" (Dix 2000; Dix 2003) and examined how to optimise this relationship and convey it to users. Advances in graphics enabled the development of virtual reality systems, in turn sparking a wave of interest in "virtual space" and its potential for replacing real space (e.g. Shiode and Kanoshima 1999). The popularity of the Internet, the World Wide Web, and "online" systems gave rise to research on cyberspace (e.g. Spinello 2000), interaction space (O'Neill et al. 1999), and social space (Harrison and Dourish 1996). All such work suggests that computer systems generate various types of space that as designers we must understand, manipulate and optimise.

> Interestingly, while some theories consider digital and virtual space as potential alternatives to physical spaces, today we see that physical space is more important than ever before. With urbanisation levels reaching more than 50% for the first time in history (United Nations 2007), visions suggesting the end of cities and the rise of cyber cities seem more and more out of touch. As Stephen Graham suggests, "We are not experiencing some wholesale, discrete break with the urban past ushered in by the 'impacts' of new technology. Rather, we are experiencing a complex and infinitely diverse range of transformations where new and old practices and media technologies become mutually linked and fused in an ongoing blizzard of change" (Graham 2005, p. 96).

Urban space has become more relevant than ever, especially in light of pervasive systems operating at the urban scale. As the field of pervasive systems matures, the models, theories, and tools used in our research evolve. Hence, as computers move away from the desktop and into our cities and our lives, we need new theories and tools to help us understand the relationship between urban space and our systems. Urban space is now the infrastructure on top of which our pervasive systems operate (Dourish and Bell 2007).

In an attempt to systematically incorporate this infrastructure in our systems, we here present our novel adaptation of space syntax (SS) and its use in the design and development of pervasive systems. To acquaint the reader with SS's concepts and methodology, we first describe SS and present its main ideas from the perspective of pervasive systems. We then present three case studies demonstrating our novel application and extension of SS in the course of designing pervasive systems.

Our main contributions are, first, the adaptation and operationalisation of SS for use in pervasive computing, and second the use of SS to develop pervasive systems.

## 3.2 Space Syntax

Space syntax, much like pervasive computing, is a relatively young and multidisciplinary domain. The central concept behind SS is that the configuration of space, rather than space itself, is the driving force behind how cities operate. Architectural and urban design, both in their formal and spatial aspects, are seen as configurations in that the way the parts are put together to form the whole is more important than any of the parts taken in isolation. The first major publication on SS came in 1984 where Hillier and Hanson set out a theory of space as an aspect of social life (Hillier and Hanson 1984). In his second major publication Hillier (1996) brought together an array of configuration analysis techniques and applied them to issues of architectural and urban theory. What sets SS apart from other architectural theories is that SS aims to explain how cities are, rather than how cities ought to be. The basic motivation of SS is the realisation that in addition to functioning as bodily protection, buildings operate socially in two ways: they constitute the social organisation of everyday life, and they represent social organisation.

The initial research finding that motivated the development of SS is that, *ceteris paribus*, movement in cities is generated by the configuration of the cities' streets. This fundamental relationship underlies many aspects of cities including the distribution of land use and the spatial patterning of crime. SS research has shown that ineffective configuration in cities or housing estates can cause social segregation and antisocial behaviour. Such examples have given rise to the notion of good, or bad, syntax. To better illustrate this point, Hillier provides an analogy between architecture and language.

> Language is often naively conceptualised as a set of words and meanings, set out in a dictionary, and syntactic rules by which they may be combined into meaningful sentences, set out in grammars. This is not what language is, and the laws that govern language are not of this kind. This can be seen from the simple fact that if we take the words of the dictionary and combine them in grammatically correct sentences, virtually all are utterly meaningless and do not count as legitimate sentences (Hillier 1996, p. 7).

Similarly, successful architecture is not just about simply constructing streets and connecting buildings, and neighbourhoods; it is about how we connect the different parts of a city, neighbourhood or building.

## *3.2.1 Operationalisation*

We note that while SS covers a multitude of topics, including vision, isovists, and indoor space, here we focus on the application of SS to urban space. While cities are extremely rich in many respects, SS focuses on a single property: spatial configuration. All claims, findings, and understandings of SS derive from the analysis of spatial configuration. While SS is applied to both indoor and outdoor spaces, here we focus on the outdoor and urban use of SS.

### 3.2.1.1 Axial Maps

The fundamental premise of SS is that every city can be represented as an "axial map". A city's axial map can be derived by simply drawing the set of least and longest straight lines (axial lines) such that all open spaces have an axial line

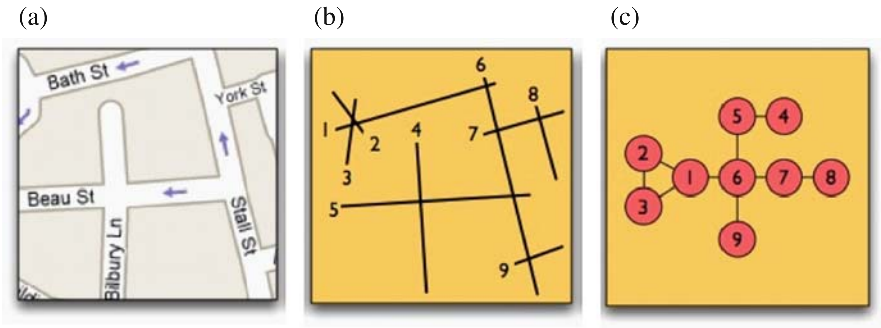



**Fig. 3.1** (**a**) map of city, (**b**) axial map, (**c**) graph representation

passing through them. The next step in the analysis is to transform the axial map into a graph. This transformation is a unique aspect of SS as it results in graphs that, at first, seem counter-intuitive (Fig. 3.1).

An axial map is transformed into a graph by representing each axial line as a node, and linking the nodes (i.e. axes) that intersect in the axial map. This graph representation is the edge-vertex dual of how we intuitively represent maps: street intersections as nodes and streets as edges. SS uses the opposite: streets become nodes, and intersections become links.

It is worth highlighting the fact that in SS analysis streets become nodes. Each street, irrespective of its width, length, and location, is represented as a single node in the graph. This is an extremely powerful representation, because it enables researchers to identify concrete metrics for each street in a city. This is achieved by carrying out analysis of the graph, and deriving numerous metrics for each node (i.e. street).

#### 3.2.1.2 Metrics

A city graph can be analysed in terms of its properties such as the integration of nodes. In SS, integration describes the deepness or shallowness of a node in relation to the other nodes in the graph. To arrive at a highly integrated street, pedestrians require on average fewer changes of direction (assuming a change of direction means changing streets). On the other hand, less integrated streets are relatively isolated from other streets, and require more changes of directions to get to them. Typical examples of highly integrated streets are high streets, while cul-de-sacs and alleyways are usually less integrated.

Readers may realise that the metric of integration is identical to the concept of distance in traditional graph analysis. In fact, SS utilises many traditional graph metrics, but has developed its own jargon when referring to them. Additionally, SS researchers have developed new metrics. In Table 3.1 we show the most popular metrics used by SS, their conventional name (if one exists), and their definition and description.

**Table 3.1** Table basic space syntax metrics. For each metric we give the corresponding graph theory terminology (if it exists), along with a definition and description of the metric

| Space syntax | Graph theory | Description/Definition |
|---|---|---|
| Integration | Closeness | The mean distance between an axis and all other axes of the system. |
| Connectivity | Degree | The connectivity of axis i is the number of axes that it intersects. |
| Choice | Betweeness | The number of times axis i is used when calculating the shortest paths between all pairs of axes in a system. |
| Control | – | The degree to which axis i controls "access" from and to the axis it intersects. |
| Global metric | – | A metric for an individual axis, calculated using the whole system. |
| Local metric | – | A metric for an individual axis, calculated using the axis' neighborhood (e.g. axes up to three intersections away). |
| Intelligibility | – | The correlation between axes' connectivity and global integration |
| Synergy | – | The correlation between axes' local and global integration. |
| PageRank | PageRank | The popularity of an axis i as determined by the number of popular axes intersecting axis i (recursive) |

Connectivity is a metric used to describe the number of streets that any particular street intersects. (Note that in terms of SS, streets and axes are identical, since all metrics are calculated based on the axial map representation shown in Fig. 3.1b).

Control is a metric that describes the extent to which an axis, or street, controls access to and from streets that it intersects. For instance, a street i connected to multiple cul-de-sacs has high control, as anyone wishing to go to the cul-de-sacs must use street i.

Similarly, choice is a metric that describes the importance of a node in the context of a greater system. An axis i with high choice indicates that many shortest paths between pairs of axis use axis i. In other words, axis i is utilised by many pedestrians wishing to follow the shortest path to their destination.

The metrics we describe here may be applied globally or locally. Global metrics consider the whole system when calculating a property of an axis, while local metrics consider only the immediate neighbourhood of an axis. For instance, global integration describes the mean distance between an axis and all other axes of the system. Local integration, on the other hand, describes the mean distance between an axis and its immediate neighbourhood. Thus, local integration of "radius 3" considers an axis neighbourhood up to three intersections away.

Calculating the correlation of one metric with another generates a number of metrics. For instance, a city's intelligibility is the correlation between axes' connectivity and global integration. In other words, intelligibility is a measure of how well we can predict an axis' global integration given its connectivity, and vice versa. Practically, intelligible urban grids offer a good correlation between what pedestrians can see, and what they cannot see.

A similar metric is synergy, which describes the correlation between axes' local and global integration. This metric considers how well we can understand a neighbourhood given an understanding of a city, and vice versa.

More recently, the PageRank algorithm (originally used by Google search) has been shown to provide a very good correlation with pedestrian movement (Jiang 2009; Jiang et al. 2008). It provides a means for relatively ranking the importance of an axis *i* by considering the importance of the axes pointing to *i*.

### *3.2.2 Data Collection and Observation*

An integral part of SS involves data collection, either through observation or analysis of publicly available data sources. While axial maps and graphs offer concrete metrics, it is only through observation and further data collection that one may understand how a city operates on an aggregate level.

In the course of their work, some researchers may be interested in the economic aspects of the city, while others might focus, for example, on the flow of pedestrians. It is through data collection and observation that SS can make claims about such aspects of a city. Using correlation, SS analyses observational data in relation to the structural properties of streets described in the previous section. We now give examples of the most popular data sources utilised by SS, and then describe the techniques used to gather empirical data.

#### 3.2.2.1 Data Sources

There is a virtually endless list of data sources that SS can utilise, provided that they are appropriately geo-tagged. This means that the data must be attributable to specific streets in a city. Examples of such data include statistics on crime, land use, land use change, movement & flows, economic growth, safety, social status, work habits, shopping behaviour, quality of life, demographics, prostitution, and human encounter. All these data sources can be utilised by SS, if they are fine-grained enough to enable researchers to attribute this data to specific streets of a city.

In addition to utilising existing datasets, SS employs three primary observational techniques, which enable researchers to gather data on people's use of cities. These techniques are gatecounts, static snapshots, and trails.

#### 3.2.2.2 Gatecounts

Gatecounts are used to establish the actual flows of people throughout the city over the course of a day. A gate is a conceptual line across a street, and gatecounts entail counting the number of people crossing that line. Using this technique the observer stands on the street and counts the number of people crossing the gate in either direction. A typical 2-day observation (complemented by a 1-day pilot) will involve about 100 gates throughout the city, as shown in Fig. 3.2 (taken from one of our own studies). Here we counted the number of


**Fig. 3.2** Gatecounts. In the *top*, a map of the city marked with the location of the gates. At the *bottom*, the observed flows of pedestrians, colour-coded from red (high flow) to blue (low flow)

people crossing each gate, and to do so we deployed multiple observers. Our 12 observers took 5-min samples from each gate in five cycles throughout the day, from 8:30 a.m. to 4:00 p.m. over 2 days.

In Fig. 3.2 we see a colour-coded graph of the observed flows of people, ranging from high flows of 2,750–4,000 people per hour (red) to low flows of 250 people per hour or less (blue). This enables us to correlate between the (SS) predicted and actual flows of people across the city. In this manner we are able to identify the general trend of the city, as well as some outliers - i.e. streets that are over-performing or under-performing (in terms of pedestrian flows) in relation to their integration.

#### 3.2.2.3 Static Snapshots

The next SS observation method we consider is static snapshots, in which various open spaces of the city are considered in detail (Fig. 3.3). The method can be used for recording both stationary and moving activities, and is useful

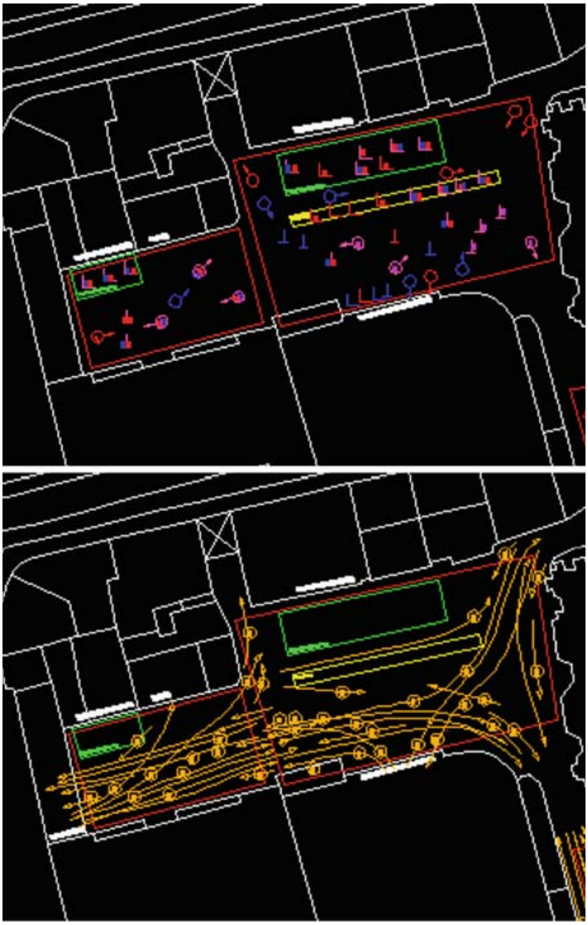


**Fig. 3.3** Static snapshots. The *top* shows a map of a city, with multiple "convex spaces" identified. The *middle* and *bottom* shows the static and mobile activity as recorded in one such space

when a direct comparison is being made between the two types of space use. A strength of this method is that it makes the patterns of space use in an urban area apparent. For each open space under consideration, the observers record the movements in and out of the space, as well as the type of activity taking place in the space.

In Fig. 3.3 we present a sample observation of people's movement within a particular city. The top of Fig. 3.3 shows the map of a city, with marked "convex spaces" where observation is to take place. For one such space we show our observation of static (middle of Fig. 3.3) and moving activities (bottom of Fig. 3.3). Here, we identify the entry and exit points that people used throughout the day. This gives us an understanding of how people perceive a particular space, and how it is utilised. For example, we may observe that a seating area in a park is actually not used for seating but for playing by children. A common observation is use of certain spaces by people making calls on the mobile phone or using their laptops. In Fig. 3.3 we see our observations of people's activities in this space. We differentiated between instances of males, females, children, or groups of people. For each instance we recorded if they were standing or sitting, and also commented on their behaviour and activities.

#### 3.2.2.4 Trails

A third method used for understanding the urban landscape is called trails, or people following. This involves tracing the movement of a specific person or group for a fixed time period (e.g. 3 min as shown in Fig. 3.4). This is an

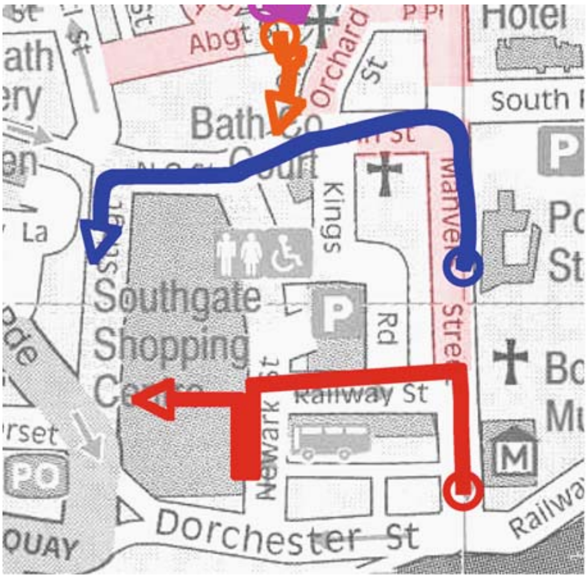



**Fig. 3.4** Trails. A set of 3-min trails are shown here, indicated by a beginning (*circle*) and end (*triangle*)

increasingly important technique for observing movement that disperses from a specific "movement distributor" – for example, a train station or a shopping mall. It can be used to investigate three specific issues:

- The pattern of movement from a specific location,
- The relationship of a route to other routes in the area, and
- The average distance people walk from the specific location (this can help determine the pedestrian catchment area of architectural spaces such as a retail facility or public square).

Trails give us a sense of how people use the city in different ways. A mother pushing a pram will follow different routes compared to a young student or a tourist. In Fig. 3.4 the two large trails at the bottom show two people walking at a fast pace who decided to take alternative routes for similar beginning and destination points. We also see at the top of Fig. 3.4 two slow-paced pedestrians originating from the same location but going to different destinations. As part of this method, observed trails are augmented by comments that the observer can attach to a trail. Typically, trails are collected by "people following" or manual observation, but GPS trails can also be used.

### *3.2.3 Empirical Results*

SS researchers have explored various aspects of space and cities in terms of axial maps. Aspects such as pedestrian flow, space use density, crime, cognitive maps, and intelligibility, have been considered in relation to SS. Here we provide a brief overview of their results.

The correlation between pedestrian movement and spatial integration has been elaborated in (Hillier and Hanson 1984), while further studies have identified this correlation in relation to a large sample of cities and towns (Hillier et al. 1987), galleries (Hillier 1996; Peponis et al. 2004), office buildings and workspaces (Sailer and Penn 2007; Peponis et al. 1997), and even research laboratories (Hillier et al. 1985). While the observational evidence provides strong correlations, causality has not been determined to date. The recent identification of PageRank as a metric that correlates extremely well with pedestrian movement may provide some insight into a possible causality (Jiang 2008; Jiang et al. 2008).

In an attempt to understand the effect of space on cognition, the authors in (Kim and Penn 2004) analysed people's cognitive maps using SS. Their results suggest that the image of spatial configuration maintained in cognitive maps is highly associated with that of the physical built environment. The strong correlation between sketch maps and purely configurational measures of spatial patterns suggest that configurational and relational properties of spatial patterns are perhaps more important for the purposes of navigation than is precise local information.

A further aspect of the relationship between space and movement is highlighted by the differences between weakly and strongly programmed spaces. Strong programmes exist when activity follows a strict pattern, while weak programs describe an "all-play-all" environment. Thus, while courts and schools are strongly programmed, research labs and shopping malls are weakly programmed. The study reported in (Hillier 1996) suggests that as the program becomes weaker, the distribution of space use and movement is defined less by the program and more by the structure of the layout itself. In other words, weakly programmed spaces exhibit a strong correlation between integration and density of space use.

Similarly, the relationship between space and crime has been reported in a number of papers (e.g. Hillier and Sahbaz 2005; Hillier and Shu 2000). The findings suggest a relationship between spatial morphology and various types of crime, such as street robbery and residential burglary. The researchers note that conventional approaches to preventing crime, such as the creation of cul-de-sacs, may in fact be reinforcing crime.

A fundamental difficulty in understanding space is that it is generally impossible to intervene in a systematic way. Because it is extremely difficult to modify aspects of cities and observe the effects, SS research utilises modelling experiments to observe the effect of spatial configuration. In such studies, virtual spaces are generated, analysed and modified. For instance, in a study reported in (Hillier 1996) the layout of a hypothetical deformed grid is shown to have a high correlation between integration and connectivity. The building blocks are then slightly shuffled, resulting in a weak correlation (Fig. 3.5). This

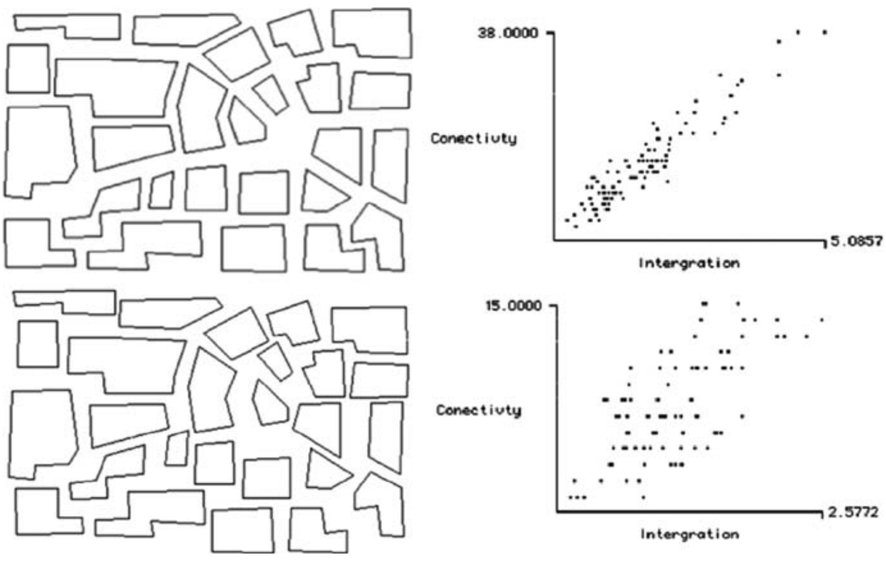



**Fig. 3.5** The *top* figure shows an "intelligible" urban layout, where integration and connectivity are highly correlated. The *bottom* figure shows the effect of a slight misplacement of the buildings, resulting in an unintelligible layout (Hillier 1996)

marginal reshuffling is shown to result in an unintelligent design, where what pedestrians can see from any location on the grid is not a good guide to what they cannot see.

While difficult to generalise from such artificial experiments, researchers have found universal properties of cities. In a recent study, 36 cities in 14 countries were analysed in terms of axial maps (Carvalho and Penn 2004). The authors show that the cities form two main clusters when considering the distribution of axial length, irrespective of the size of cities. Their results point to an underlying self-similar morphological property of cities, suggesting that cities are fractal in nature.

Some of the topics that SS-related research has recently begun to address relate to incorporating geometric and metric measures in the analysis of cities (Hillier et al. 2007), measuring spatial capital (Marcus 2007), understanding presence, co-presence and encounter (Hanson and Zako 2007; Kostakos and O'Neill 2007), preserving and rehabilitating space, settlements and cities (Amorim et al. 2007; Greene and Mora 2007; Karimi et al. 2007), designing spaces for improved evacuation (Choi et al. 2007), transportation (Kisimoto et al. 2007), and quality of life (Ribeiro and Holanda 2007), considering the relationship between space and prostitution (Hadjichristos 2007), and extending SS to three-dimensional analyses (Wang et al. 2007). Further work has considered automated ways of generating axial maps (Peponis et al. 1998). Most of these topics are in early development, but provide a good indication of the issues SS researchers are interested in further exploring.

### *3.2.4 Criticism*

The simplicity of SS, and its apparent ability to attribute complex phenomena to just spatial configuration, has induced a lot of criticism. A point of debate is whether or not SS is a theory. It is the view of the authors that SS is not a theory in the traditional scientific sense, as it cannot predict phenomena a priori. However, SS does provide invaluable tools to explore and understand real world phenomena, and, more importantly, to generate hypotheses.

A stream of criticism addressed at detailed aspects of SS was published in (Ratti 2004), where it was suggested that the same spatial configuration can result in more than one mapping, that SS cannot deal with regular grids, that SS ignores building height, discards metric information and land use information, and finally that axial maps are arbitrary.

A rejoinder to this criticism (Hillier and Penn 2004) addresses each of the above points in detail. The authors claim that minor changes in the built environment can give rise to alternative mappings from apparently similar configurations, but point out that, syntactically, those changes are important. Additionally, they provide morphological and behavioural evidence to support their claim. On the issue of SS's inability to deal with regular grids, the authors respond by claiming that, in practice, such grids do not occur, and therefore SS is applicable. In respect to SS's disregard for building height, metric information, and land use information, the authors explain that these are considered in their analysis, rather than in the axial map. In response to the criticism of axial maps being arbitrary, the authors provide statistics evidence from 28 major cities to suggest that axial lines are significant elements of cities, and that the errors associated with tracing them are not statistically significant.

## 3.3 Pervasive Systems and Space Syntax

So far we have described the components of SS analysis. We identified axial and graph analysis as the basis of SS analysis, and pointed out that a number of metrics have been developed to describe and understand cities. We also highlighted data collection and analysis as important steps in SS analysis. We now summarise how SS integrates all these aspects in a workflow, and provide a sample "outline" of how to carry out SS analysis.

The steps involved in SS analysis may be summarised as:

1. Generate an axial map of a city
2. Gather data (observation or data sources)
3. Identify correlations between data and SS metrics
4. (Optional) Evaluate alternative designs

The first two steps are a straightforward application of the processes we described in the previous sections: generating axial maps, and gathering data. Step three is in many ways the crucial step of SS analysis. Given a dataset on, say, crime, SS researchers try to correlate any of the multitude of metrics they have developed with instances of crime.

To illustrate our example of crime, consider a dataset that describes the number of burglaries that took place on every street of a city over a 5 year period. Thus, some streets may have zero burglaries, others one or two, while other streets may have fifty or more. Such data may be correlated with any SS metric that can characterise streets, such as integration, connectivity or choice. Thus, we are able to plot a graph where one axis describes the number of burglaries, while the other axis describes connectivity, integration, or any other SS metric. Each plotted point in such a graph represents a street. Overall, crime and any SS metric may display a strong correlation that may be described by a mathematical function.

Given the identification of a strong correlation, in step four we are able to evaluate alternative designs. For instance, let us assume a strong correlation between crime and integration. We can then explore the effect on crime if we introduce a new bridge, close a street, or replace a building block with a park. To evaluate such alternatives, we alter the axial map to reflect the proposed changes and assuming the correlation between crime and integration holds, observe how crime is distributed based on the newly calculated integration values for each street.

To incorporate SS in the design of pervasive systems, we have modified the four steps described above by incorporating our own systems in the procedure. The result is the following iterative sequence:

1. Generate an axial map of a city
2. Gather geo-data about the *pervasive system* (logs, usability, etc.)
3. Identify correlations between data and SS metrics.
4. (Optional) Evaluate alternative designs of the *pervasive system*

Our procedure describes a meaningful way of analysing data characterising our systems, and correlating them with the urban morphology of the environment in which our system is deployed. For instance, usability data can be captured about a system at various locations across a city, and then correlated with SS metrics just as crime data can be correlated with SS metrics. Furthermore, these correlations can provide a rationale for evaluating alternative designs and changes to our systems.

This modified procedure is a concise summary of how we have integrated SS in the design of our systems. We now describe in detail three case studies in which we used SS in our research.

## 3.4 Case Studies

In this section we describe three case studies where we used SS in the course of research and development. These case studies demonstrate the role of SS as a simple application module, as an explanatory tool, and as a modelling tool.

### *3.4.1 Space Syntax and Application Development*

In this case study we consider the redesign of a pedestrian way-finding application (Collomosse et al. 2006). The concept behind this application is that users can utilise their mobile phones' camera to take a picture that may be analysed in order to correctly identify a user's location and orientation. This would be useful for phones that do not have GPS, or where GPS reception is poor. Our analysis is made possible by a server database that holds geo-tagged images of a city. The application originally worked by uploading a photograph to the server, which searched through the database for a close match. The server had a virtual map of the city, and many photographs attached to specific locations in this map. Finding a match for the uploaded image meant identifying the user's location and orientation. However, the search through such a database of images is linear, and therefore expensive.

In our redesign, we suggest that the search can be optimised if we have some knowledge of the user's location. For instance, we can assume that users may have GPS that can be used to detect location, but this may not always function accurately within a city with tall buildings and narrow streets. Another source of location information may be proximity technologies, such as WiFi triangulation or Bluetooth beacons. Given such information, an obvious optimisation to the search process is to use the last known co-ordinates of a user's location in order to narrow down the search through the database. Thus, knowing that 10 min ago the user was at location x, the search through the database may be limited to the streets within a radius of 1 km from location x (depending on user's average speed).

In Fig. 3.6 (top) we show this optimisation in the form of a white circle enclosing all the streets to be searched during a particular query. A further optimisation is to focus on the paths that have a total distance of up to 1 km, such that only a portion of the white circle is searched. These are shown as dark axes highlighted with red.

In the bottom of Fig. 3.6 we show how SS was used to further optimise the database search. Our intention was to identify an optimum ordering of the axes to be searched in the database. Since SS can give us a list of axes order by the

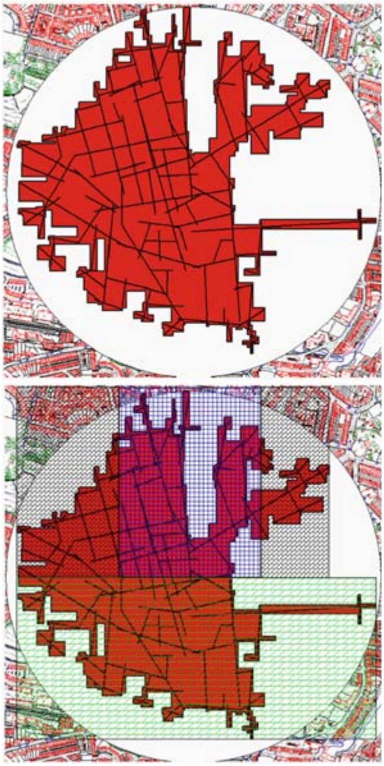


**Fig. 3.6** Space syntax optimisation of searching through a database of geo-tagged images. Assuming a pedestrian was heading north, we identify the streets on which they may have moved to after x minutes, and further identify three zones based on the pedestrian's last known trajectory

likelihood of where a user might be, a sequential comparison between a user's submitted image and the images associated with the highlighted axes is more likely to find a match sooner rather than later. Overall, therefore, the search effort is reduced.

Our SS optimisation was done through a series of iterations, each time measuring any obtained performance gains. Initially we ordered the axes using a radius-x SS metric, i.e. we considered a neighbourhood of diameter x segments around the user's last known location and ordered all of them with various SS metrics. However, this approach did not present significant performance gains.

After a series of iterations, we settled on the following solution. First, the radius-x axes are segmented in three zones (Fig. 3.6, bottom) based on the user's last known trajectory. Here, our assumption was that users are more likely to maintain their last known trajectory, rather than change it, as suggested by SS. Therefore we create three zones:

- zone A: assume the user kept his last known trajectory (blue, square-patterned grid in Fig. 3.6).
- zone B: assume the user deviated slightly from his last know trajectory, either to the left or to the right (black, zigzag-patterned grid in Fig. 3.6)
- zone C: assume the user greatly deviated from his last known trajectory (green, tile-patterned grid in the lower-half of Fig. 3.6).

Within every zone we rank each axis based on its local integration. The result is that rather than searching through the whole set of radius-x axes in an arbitrary order, the axes are ordered first based on the user's last known trajectory, and then based on each axis' local integration (as defined in Table 3.1).

Our optimisation reduced the search time by 75% on average (Mason 2007). The most successful aspect of our optimisation was zone ordering. Our further refinement of axes ordering based on local integration was successful, but to a smaller extent. This was in part due to the fact that the zones themselves where too small for any further ordering to have a considerable effect.

### *3.4.2 Space Syntax as an Explanatory Tool*

In recent years, a considerable amount of research has focused on collecting mobility traces using networks of WiFi or Bluetooth scanners (Chaintreau et al. 2006; Crawdad 2007). Such traces describe the presence of people at different locations and different times. Earlier work (O'Neill et al. 2006) has reported on a collection methodology, which involves augmenting the SS gatecount and static snapshot methods by employing a network of Bluetooth scanners.

Further work (Kostakos and O'Neill 2007) explains how mobility traces may be translated to sociographs describing users' encounters across a city and over time. Effectively, each user gets represented as a node in a sociograph, and then link together nodes (i.e. users) that have encountered each other at some point in time. Doing so allows us to represent the dataset of each scanning location, as well as of all scanning locations, as a sociograph. This has yielded some interesting results, including the identification of underlying equations describing the degree, betweenness, and closeness of the derived sociographs.

In Fig. 3.7 we see a visualisation of the process: for various scanning locations across the city, a number of sociographs are derived over time. While merging all scanning locations may generate a complete sociograph, for our purposes here we choose to differentiate between each scanning location. Here, we show what each individual scanner can "see".

An observation we have made from the comparison of sociographs from different scanning locations is that they are similar, yet distinct (Kostakos and O'Neill 2007). While visually it is difficult to accurately compare sociographs,

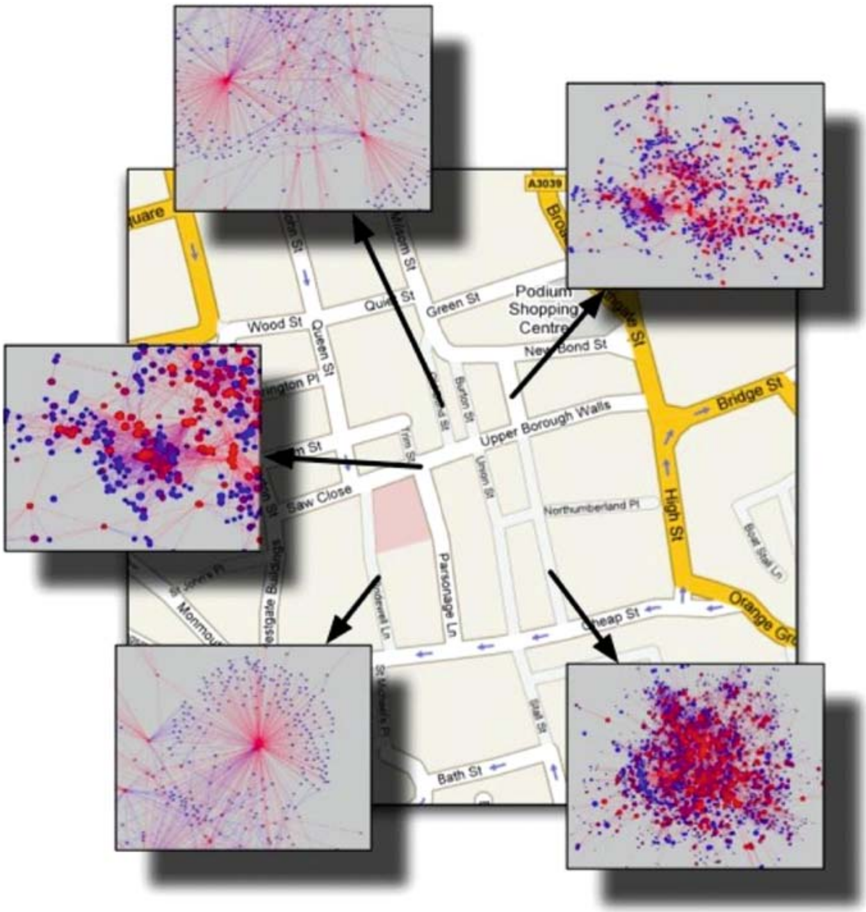



**Fig. 3.7** Visualisation of sociographs obtained from various locations in a city

we are able to compare them by plotting their various properties (such as degree distribution). In our research we have found that the sociographs are similar enough to make us hypothesise the existence of an underlying principle, yet distinct enough such that each location gives rise to unique sociograph properties.

Our long-term research is focused on identifying the reasons for those small, yet distinctive differences in the sociograph properties. We do so by applying our modified SS procedure described earlier. Currently, our hypothesis is that spatial configuration has an effect on the observed sociographs. Much like researching the relationship between crime and spatial configuration, we are currently researching the relationship between spatial configuration and urban encounters as represented by sociographs. The understanding of this relationship can be a great help in the *a priory* design, as well as deployment of pervasive systems in cities.

Our results suggest that spatial integration is correlated with sociograph density. However, there are many more relationships to be examined, since spatial layout and sociographs each have numerous properties, resulting in an exponential number of potential correlations to be examined. Hence, a key contribution of our work lies in developing a systematic way of identifying and examining these relationships.

Currently, however, the main obstacle in such a line of research is that many samples are required to derive strong correlations. In our case each sample corresponds to a scanning location: in Fig. 3.7, for example, we show five such samples. In order to derive meaningful statistics, one requires a large number of scanning sites, which will operate over long periods of time. This can be quite expensive in terms of both computational and human resources. A way of overcoming this limitation is by using modelling and simulation, as we describe next.

### 3.4.3 Space Syntax as a Modelling Tool

As a direct result of observing people's movement in space, the authors in (Turner and Penn 2002) have developed an agent simulation of pedestrian movement. Drawing on Gibson's ecological theory of perception (Gibson 1979), they developed a vision-based mental model of the world, based on the concept of axial lines, which drives agent behaviour. Their approach is distinct in the sense that their system employs an exosomatic visual architecture, thus their agents may be programmed with movement rules originating from Gibson's principle of affordance. By doing so, they make space, rather than agent intelligence, the primary causality of movement.

The resulting aggregate movement patterns of agents closely match those of humans. Based on these results, they have published a free simulation environment, called Depthmap (Turner 2001), which enables users to load city maps and observe aggregate pedestrian movement (Fig. 3.8).

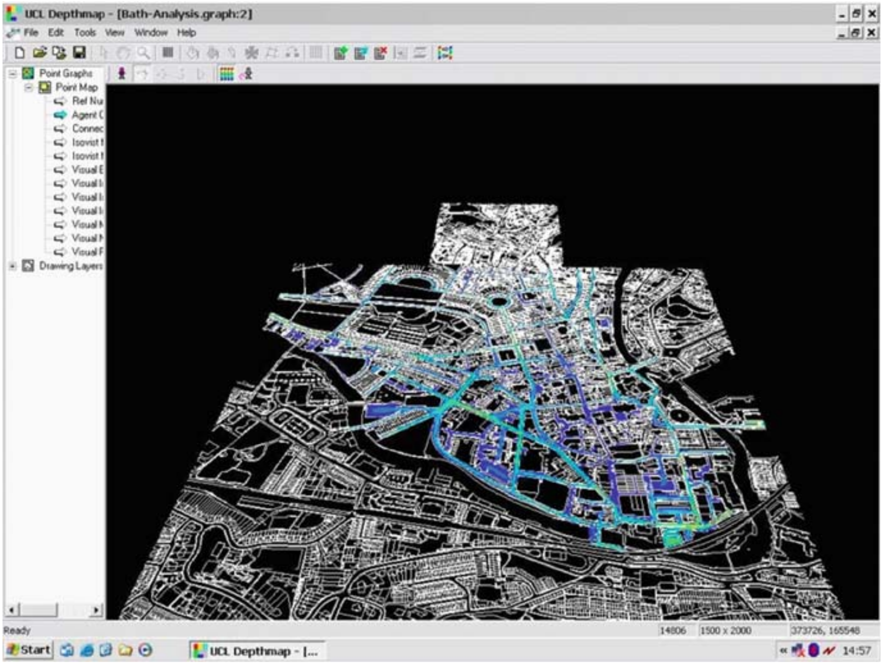



**Fig. 3.8** A map of the city of Bath with virtual pedestrians. Within this simulation we are able to establish virtual scanners that collect synthetic mobility traces

In our research we have extended Depthmap in order to generate synthetic mobility traces. As we noted in the previous case study, it is really difficult and expensive to carry out longitudinal Bluetooth or WiFi scans at multiple locations in a city. To address this problem, we have focused our efforts on carrying out Bluetooth scans within Depthmap.

By loading a map of the city of Bath inside Depthmap, we are able to establish virtual scanners in the exact locations where we carried out our real-world Bluetooth scans. At the moment, we are tweaking the Depthmap simulation rules so that the sociograph properties of the virtual scans match those of the real-world scans. Once we derive simulated data that match our empirical data, we will then extend the network of our virtual scanners within Depthmap. This will give us enough data to enable us to look for correlations between sociograph properties and spatial properties, as described in the previous section. Effectively, this approach will enable us to systematically identify the effect of spatial morphology on encounter and ultimately social behaviour. Moreover, these may be used to help us understand how to best integrate our systems in the fabric of everyday life.

## 3.5 Conclusion and Ongoing Work

In this paper we describe our novel use of space syntax (SS) in the design and development of pervasive systems. Our two novel contributions are our modification of the SS methodology, and our use of SS in building three distinct systems.

SS has a number of strengths that we find useful in our research. First, it helps us understand how space is, as opposed to how it ought to be. Additionally, the understanding and metrics derived from SS are in a form readily understood by computers. As such, SS can easily be used to design and augment pervasive applications. We demonstrate this in our case studies, where we used SS to simply augment and improve an existing application, as an explanatory tool, and as a modelling tool.

Our ongoing work, hinted throughout the case studies, involves the development of an urban space simulation environment. Such an environment will enable us to quickly test pervasive applications that draw on people's mobility and sociability.

**Acknowledgments** The author wishes to thank Eamonn O'Neill, John Mason, Alan Penn, Ava Fatah gen. Schieck, Shinichi Iida, Bill Hillier, Alasdair Turner, and all other members of the UCL space syntax Lab.

## References

Amorim, L., Loureiro, C. & Nascimento, C. (2007). Preserving space: Towards a new architectural conservation agenda. *Proceedings of the Space Syntax Symposium 2007*, Istanbul, 32, 1–13

Carvalho, R. & Penn, A. (2004). Scaling and universality in the micro-structure of urban space. *Physica A*, 332, 539–547

Chaintreau, A., Hui, P., Crowcroft, J., Diot, C., Gass, R. & Scott, J. (2006). Impact of human mobility on the design of opportunistic forwarding algorithms. *Proc. INFOCO*, IEEE CS Press

Choi, J., Kim, M. & Choic, H. (2007). Evacuation efficiency evaluation model based on euclidean distance with visual depth. *Proceedings of the Space Syntax Symposium 2007*, Istanbul, 41, 1–16

Collomosse, J.P., Al Mosawi, K. & O'Neill, E. (2006). Viewpoint invariant image retrieval for context in urban environments. *Proceedings of the Conference of Visual Media Production (CVMP)*, IEEE Press, 177

Crawdad (2007), Crawdad project. http://crawdad.cs.dartmouth.edu. Last access 22 August 2007

Dix, A. (2000). Welsh mathematician walks in cyberspace. *Proceedings of the 3rd International Conference on Collaborative Virtual Environments* (pp. 3–7). (New York: ACM Press)

Dix, A. (2003). Managing multiple spaces. *Proceedings of the Workshop on Space, Spatiality and Technology*, (Edinburgh: Napier University)

Dourish P. & Bell, G. (2007). The Infrastructure of experience and the experience of infrastructure: meaning and structure in everyday encounters with space. *Environment and Planning B: Planning and Design*, 34(3), 414–430

Gibson, J.J. (1979). *The ecological approach to visual perception*. (Boston: Houghton Mifflin)

Graham, S. (2005). Strategies for networked cities. In L. Albrechts & S. Mandelbaum (Eds.), *Planning in the network society* (pp. 95–109). (Routledge)

Greene, M. & Mora, R. (2007). Rehabilitation process in city centres: new residents in traditional areas. *Proceedings of the Space Syntax Symposium 2007,* Istanbul, 33, 1–14

Hadjichristos, C. (2007). Prostitution spatialised: Cyprians then and now. *Proceedings of the Space Syntax Symposium 2007*, Istanbul, 113, 1–6

Hanson, J. & Zako, R. (2007). Communities of co-rpesence and surveilalance: How public open space shapes awareness and behaviour in residential developments. *Proceedings of the space syntax symposium 2007*, Istanbul, 21, 1–21

Harrison, S. & Dourish, P. (1996). Re-place-ing space: the roles of place and space in collaborative systems. *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW)* (pp. 67–76). (Boston, MA, USA, ACM Press)

Hillier, B. (1996). *Space is the machine*. (Cambridge: Cambridge University Press)

Hillier, B., Burdett, R., Peponis, J. & Penn, A. (1987). Creating life: or, does architecture determine anything? *Architecture & Behaviour*, 3(3), 233–250

Hillier, B. & Hanson, J. (1984). *The social logic of space*. (Cambridge: Cambridge University Press)

Hillier, B. & Penn, A. (2004). Rejoinder to Carlo Ratti. *Environment and Planning B*, 31, 501–511

Hillier, B., Penn, A., Grajewski, T. and Burdett, R. & Musgrove, J. (1985). Space standards and configuration in research laboratories. *Technical report*. Bartlett School of Architecture and Planning, UCL, London, UK

Hillier, B. & Sahbaz, O. (2005). High resolution analysis of crime patterns in urban street networks: an initial statistical sketch from an ongoing study of a London borough. *Proceedings of the Space Syntax Symposium 2005*, 1, 451–478

Hillier, B. & Shu, S. (2000). Crime and urban layout: the need for evidence. In S. Ballantyne, V. MacLaren & K. Pease (Eds.), *Secure foundations: key issues in crime prevention, crime reduction and community safety* (pp. 224–248). (London, UK: Institute for Public Policy Research)

Hillier, B., Turner, A., Yang, T. & Park, H.T. (2007). Metric and topo-geometric properties of urban street networks: some convergences, divergences and new results. *Proceedings of the Space Syntax Symposium 2007*, Istanbul, 1, 1–21

Jiang B. (2009). Ranking spaces for predicting human movement in an urban environment. *International Journal of Geographical Information Science*. Preprint, arxiv.org/abs/physics/06-12011, 23(7), 823–837

Jiang B., Zhao S. & Yin J. (2008). Self-organized natural roads for predicting traffic flow: A sensitivity study. *Journal of Statistical Mechanics: Theory and Experiment*, July, P07008, Preprint, arxiv.org/abs/0804.1630

Karimi, K., Amir, A., Shafiei, K., Rafor, N., Abdul, E., Zhang, J. & Mavridou, M. (2007). Evidence-based spatial intervention for regeneration of informal settlements: the case of Jedda central unplanned areas. *Proceedings of the Space Syntax Symposium 2007*, Istanbul, 34, 1–14

Kim, Y.O. & Penn, A. (2004). Linking the spatial syntax of cognitive maps to the spatial syntax of the environment. *Environment and Behaviour*, 36(4), 483–504

Kisimoto, T., Kawasaki, S., Nagata, N. & Tanaka, R. (2007). Optimal location of route and stops of public transportation. *Proceedings of the Space Syntax Symposium 2007*, Istanbul, 75, 1–10

Kostakos, V. & O'Neill, E. (2007). Quantifying the effects of space on encounter. *Proceedings of the Space Syntax Symposium 2007*, Istanbul, 97, 1–9

Marcus, L. (2007). Spatial capital and how to measure it: an outline of an analytical theory of the social performativity of urban form. *Proceedings of the Space Syntax Symposium 2007,* Istanbul, 5, 1–11

Mason, J. (2007). *I spy with my little eye*. University of Bath, Department of Computer Science, undergraduate dissertation

Muramatsu, J. & Ackerman, M.S. (1998). Computing, social activity, and entertainment: A field study of a game MUD. *Computer Supported Cooperative Work*, 7(1–2), 87–122

O'Neill, E., Johnson, P. & Johnson, H. (1999). Representations and user-developer interaction in cooperative analysis and design. *Human Computer Interaction*, 14(1–2), 43–92

O'Neill, E., Kostakos, V., Kindberg, T., gen. Schieck, A.F., Penn, A., Fraser, D.S. & Jones, T. (2006). Instrumenting the city: developing methods for observing and understanding the digital cityscape. *Proceeding of the International Conference on Ubiquitous Computing (UbiComp) 2006*, 315–332

Peponis, J., Conroy-Dalton, R., Wineman, J. & Dalton, N. (2004). Measuring the effects of layout upon visitors behaviors in open plan exhibition settings. *Environment and Planning (B): Planning and Design*, 31, 253–273

Peponis, J., Wineman J, Bafna, S., Rashid, M. & Kim, S.H. (1998). On the generation of linear representations of spatial configuration. *Environment and Planning (B): Planning and Design*, 25, 559–576

Peponis, J., Wineman, J., Rashid, M., Kim, S.H. & Bafna, S. (1997). On the description of shape and spatial configuration inside buildings: convex partitions and their local properties. *Environment and Planning (B): Planning and Design*, 24, 761–781

Ratti, C. (2004). Space syntax: some inconsistencies. *Environment and Planning B*, 31, 487–499

Ribeiro, R.J.D.C. & Holanda, F.R.B.D. (2007). What integration adds to quality of life. *Proceedings of the Space Syntax Symposium 2007*, Istanbul, 81, 1–12

Sailer, K. & Penn, A. (2007). The performance of space – exploring social and spatial phenomena of interaction patterns in an organisation. *Proceedings of the International Architecture and Phenomenology Conference*, Haifa, Israel.

Shiode, N. & Kanoshima, T. (1999). Utilising the spatial features of cyberspace for generating a dynamic museum environment. *Proceedings of the Symposium on the Virual Reality Modelling Language (VRML) 1999* (pp. 79–84). (Paderborn, Germany: ACM Press)

Spinello, R.A. (2000). Excerpt from cyberethics: morality and law in cyberspace. *SIGCAS Computers and Society*, 30(1), 10–13

Turner, A. (2001). Depthmap: a program to perform visibility graph analysis. *Proceedings of the Space Syntax Symposium 2001*, 31, 1–9

Turner, A. & Penn, A. (2002). Encoding natural movement as an agent-based system. *Environment and Planning B,* 29(4), 473–490

United Nations (2007). World Urbanization Prospects: The 2005 Revision, [electronic version]. Retrieved from http://esa.un.org/unup/, last access August 28, 2007

Wang, J., Zhu, Q. & Mao, Q. (2007). The three-dimensional extension of space syntax. *Proceedings of the Space Syntax Symposium 2007*, Istanbul, 48, 1–16

# Chapter 4
# Decentralized Spatial Computing in Urban Environments

**Patrick Laube, Matt Duckham, Mike Worboys, and Tony Joyce**

**Abstract** This chapter presents the concept of *decentralized spatial computing* (DeSC) as a way to embed dynamic spatial data capture and processing capabilities within our built urban environment. The chapter illustrates the potential of DeSC for safeguarding privacy in a dynamic location-based services scenario: Mobile service users protect their potentially sensitive location by the use of a decentralized query algorithms, solely collaborating with peers close by and thereby excluding the privacy bottleneck of an omniscient global service provider. In an extensive set of consecutive experiments several decentralized query algorithms were tested, trading the level of privacy for the quality of service. The use of a real world test bed, – a small part of Ordnance Survey's OS MasterMap® Integrated Transport Network™ Layer for Southampton – underlines the experiments' validity. The chapter concludes with a research and development agenda for DeSC in the urban context.

**Keywords** Decentralized spatial computing · Ambient spatial intelligence · Mobile wireless sensor networks · Privacy · Location-based services

## 4.1 Introduction

In a world of dynamic, networked, and data-rich computing, the days of spatial data processing in a monolithic geographic information system are numbered. Ubiquitous, embedded, and highly distributed computing has led to the emergence of *pervasive computing*, *ubiquitous computing*, *ambient intelligence* (AmI), and "everyware" (Greenfield 2006). Similarly, spatial computing systems are becoming highly dynamic, multi-party networks, in which mobile human users interact in dynamic networks, and spatially distributed autonomous computing

P. Laube (✉)
Department of Geography, University of Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland
e-mail: patrick.laube@geo.uzh.ch

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_4,

nodes form the "cyber-infrastructure" of urban environments. These spatial applications of "everywhere", termed here *ambient spatial intelligence* (AmSI), are becoming increasingly important to the natural and built environments of the future.

The goal of AmSI is to embed spatial data capture and processing capabilities within the environment itself, for example using *wireless sensor networks* (WSN). Traditional forms of centralized, client/server spatial data processing are hardly capable of enabling AmSI. Amongst other issues, centralized approaches to the highly dynamic, real-time nature of AmSI data sources lead to information and communication overload, unmasking the unscalable nature of conventional GIS and spatial database architectures. Omniscient centralized databases furthermore present a potential privacy breach. Hence, AmSI rather requires adopting decentralized architectures, where spatially distributed but collaborating computing nodes autonomously take on responsibility for responding to spatial queries with no centralized control (Laube and Duckham 2009). For example, with respect to privacy protection, such decentralized architectures allow private and sensitive spatial data to be collected at different sites, and analyzed in a decentralized way without collating and storing personal data in a centralized GIS or spatial database.

The emergence of such *decentralized spatial computing* (DeSC) is so far most evident in the area of environmental monitoring, where wireless geosensor networks are challenging conventional ways of centralized modeling and detecting change (Duckham et al. 2005; Worboys and Duckham 2006). The same paradigm shift is now also reaching the urban context, where spatial information processing provides the backbone of a range of application domains, including location-based services (LBS), traffic management, and facilities management. Hence, this chapter explores the notion of DeSC for embedding AmSI in urban environments.

AmSI has also raised clear privacy concerns, due to its potential for real time monitoring and rapid integration of personal and sensitive location information (Dobson and Fisher 2003). The same technological advancement that allows AmSI to pervade our urban environments also leads to the availability of finer and finer granularities of location information about users of LBS. With ever finer spatiotemporal granularity, however, location information becomes a *quasi-identifier*, allowing re-identification of previously anonymized information (Bettini et al. 2005). Hence, when considering the future of AmSI, balancing the quality of provided services and the level of privacy sacrificed for those services, becomes a key challenge.

In this chapter, we examine the potential for using DeSC techniques to deliver high quality, dynamic location-based services within an AmSI environment, at the same time as protecting the privacy of mobile individuals accessing those services. Our approach is explicitly dynamic: instead of protecting individual location fixes, we aim to protect "trajectory privacy" – the degree to which aggregated knowledge of an individual's location over time can be used to invade that person's location privacy. In our decentralized approach, mobile individuals query their spatial neighbors for responses to spatial queries. Since the neighborhood of a mobile individual is constantly changing, the likelihood that a single hostile agent ever collects enough information to seriously threaten any particular individual's location privacy is decreased. In other words, mobile individuals "smear" their location information across spacetime in order to protect their privacy. For many common spatial and LBS queries (for example, $k$-nearest neighbors) this strategy can still enable relatively high quality of service because it can exploit the spatial structure of decentralized knowledge, where mobile agents that are closer in space are more likely to possess information relevant to one another.

In conclusion, the major contributions of this chapter are:

- a discussion of the potential of DeSC in an urban context;
- a case study applying and testing the concept of DeSC for safeguarding privacy in a LBS application; and
- a road map for further DeSC research and development in the urban context.

## 4.2 Related Work

Recent work in at least three distinct areas is particularly relevant to our work. This section provides an overview and synthesis of these three related topics: current paradigm shifts in process-oriented spatiotemporal modeling of urban environments; the fundamentals of decentralized spatial computing; and location privacy protection for LBS.

### *4.2.1 The City in Flux*

Urban environments are highly dynamic. On various spatial and temporal scales, urban events and processes range from urban sprawl through migration over commuter traffic flows to pedestrian movement in a mall. Or as Worboys and Hornsby (2004, p. 327) state figuratively, "processes of urban growth and decline, migration, and expansion, constitute the city in flux". LBS and intelligent travel assistance are just two examples that illustrate the potential of AmSI and DeSC in urban environments (Dillenburg et al. 2002). For example, Winter and Nittel (2006) have shown for shared ride trip planning that decentralization is not only scalable but can deliver near-optimal solutions with local knowledge only.

Any ontological treatment of urban environments must account for both their statics and dynamics (Galton 2001, 2003; Grenon and Smith 2004; Worboys 2001). The ontology is divided into four components

- *continuant* entities, often called *objects*: entities that exist in their completeness at any moment in time, have no temporal parts, but have qualities that might change through time. Examples include roads, vehicles, people, mobile agents, houses, and cities.

- *occurrent* entities, often called *events*: entities that happen, are situated in spacetime and have temporal parts. Examples include journeys, the construction of a house, or dynamic points of interest, such as traffic accidents.
- *processual* entities: entities having some of the characteristics of occurrents, but not anchored in spacetime. Examples include walking (oppose this to a specific walk event). Flows in networks, such as vehicle densities along a transportation link, are often placed in this category (Galton and Worboys 2005).
- *situational* entities: sometimes called *sites*. These are the spatiotemporal references of entities in the above categories, if they have them. A house is situated in a region of spacetime, a journey might be thought of as a trajectory in spacetime. Depending upon the entity, the temporal component can be more or less important.

When it comes to ontological analysis or conceptual modeling for urban DeSC, not only must the above categories be investigated in the specific scenario presented, but also relationships between them. For example, in an LBS system, a mobile service user (continuant entity) may participate in a commuter journey (occurrent entity) through urban space, which is situated along a specific route in a transportation network (situational entity). It is this detailed and rigorous analysis that can provide the foundation of the computational model of the system.

In the context of DeSC for urban environments, such ontological analysis is paramount as it offers a foundation for querying dynamic systems (Worboys 2005) and for abstracting movement patterns (Stewart Hornsby and Cole 2007). Galton and Worboys (2005) identified specifically the application domain of traffic (specifically Ordnance Survey's OS MasterMap® Integrated Transport Network™ layer, ITN) for their conceptual model for dynamic spatial networks.

### 4.2.2 Decentralized Spatial Computing

Advances in ad-hoc wireless networking and micro-fabrication have enabled a new way of capturing and processing spatiotemporal information. *Wireless sensor networks* (WSN) – networks of untethered, wireless, battery powered miniaturized computers – monitor their environment by sensing, processing, and communicating information in a collaborative way (Zhao and Guibas 2004). Recent research activity in the area of WSN has focused on the establishment and maintenance of the network (e.g. Braginsky and Estrin 2002; Cheng and Heinzelman 2005), including many ingenious techniques using the spatial characteristics of the network for that purpose (e.g. Karp and Kung 2000; Mauve et al. 2001; Yu et al. 2001). WSN are especially well suited to monitor dynamics in geospace (geosensor networks, Nittel et al. 2004b) and vehicular traffic (vehicular ad-hoc networks, VANET, Kosch et al. 2006).

In this chapter we use the term *decentralized* system to specifically refer to a distributed system, where no component of the distributed system "knows" the entire system state (Lynch 1996). In decentralized systems, individual elements must cooperate to complete some processing task, but both the task and the data remain distributed throughout the network. Hence, *decentralized spatial computing* aims at the development of algorithms that can operate using purely *local* knowledge, but are still able to monitor geographic phenomena with *global* extents (Estrin et al. 2000; Laube and Duckham 2009).

For several reasons decentralized (in-network) algorithms are increasingly important to AmSI. First, decentralization increases network scalability and robustness, which is paramount to dynamic urban applications involving potentially thousands or millions of phone users or vehicles. Second, unlike global approaches, decentralized algorithms facilitate fast local updating in dynamic networks, as they do not require hard-to-maintain global consistency. Third, in-network data pre-processing reduces the need for network-wide communication, and hence conserves critically limited energy and bandwidth resources in the WSN. In this chapter we explore the potential of DeSC for providing location based services in a VANET scenario, where energy constraints are negligible as vehicles have engines and hence almost unlimited energy resources. Scalability, robustness, and a constantly changing network topology, however, remain critical VANET issues to be addressed through DeSC.

### 4.2.3 Safeguarding Privacy

In pre-internet and pre-database times privacy was safeguarded through the fragmented nature of personal information sources (Rule et al. 1980). One's bank knew about personal finances, the police about one's crime record, and the grocer about shopping habits; data integration was impractical, inference nearly impossible. The current inexorable integration of previously disjoined data sources into centralized databases creates omniscient bottlenecks open to fraud.

Recent improvements in mobile computing and location-aware technologies allow for the capture and communication of fine-grained spatiotemporal information about mobile individuals. Trajectories contain information in the form of sensitive personal points of interest and movement patterns (Bettini et al. 2005). Hence, trajectory information may act as *quasi identifier*, allowing the reidentification of anonymized data, just as with a non-spatial social security number. The increasingly dense cyber-infrastructure of our built urban environments increases privacy concerns for unsuspecting users of AmSI.

Conventional approaches protecting the privacy of mobile individuals involve *regulation*, *privacy policy*, and various forms of *data hiding* (see Verykios et al. 2008 for a comprehensive overview). Regulatory approaches to privacy develop rules to regulate the use of personal information, most importantly legislation (Langheinrich 2001). Privacy policies rely on trust and stipulate allowed uses of location information (Kaasinen 2003). Most recently, considerable effort has been put into various strategies of hiding sensitive information, especially sensitive spatial information. Anonymity concerns the dissociation of information about an individual, such as location, from that individual's actual identity. $k$-anonymity – one of the most important anonymity approaches – protects an individual's privacy in that each sensitive release is hidden in at least $k$ equally matching individuals (Bettini et al. 2005). Kido et al. (2005) hide true users' trajectories by mixing them with synthetic "fake" ones, so-called "dummies." Finally, obfuscation involves the deliberate degradation of the quality of location information (Duckham and Kulik 2005).

## 4.3 Protecting Privacy with DeSC

One of the weaknesses in the previous approaches to location privacy protection, reviewed above, is that they typically adopt a static model of privacy, aiming primarily to protect an individual's privacy for a specific instant, query, or site. Given the importance of dynamic information in urban environments ("the city in flux"), this research adopts an explicitly spatiotemporal approach to privacy protection. We argue that the disclosure of the odd (static) locational fix is often acceptable, especially as this information rapidly becomes out-of-date. Instead, gathering of trajectory information over time is the most imperative threat to an individual's geospatial privacy, enabling hostile agents to infer spatiotemporal patterns, and analyze or predict behaviors. Thus the focus of this work is less to protect the individual's location at a specific point in time, but rather to protect the trajectory in its entirety as a spatiotemporal entity.

The approach to protecting trajectory information explored in this chapter achieves *spatiotemporal privacy protection* by using DeSC. As we have seen, ubiquitous and location-aware computing environments bring with them privacy threats as they can potentially be used to collect automatically more and more detailed spatiotemporal information about an individual's location. However, we argue that DeSC allows the inversion of that process. Where centralized spatial computing architectures make the collation of spatiotemporal data easier, in decentralized architectures sensitive knowledge is "smeared" across spacetime. Potentially, decentralization ensures that no single system component can accumulate detailed knowledge about any individual, and so privacy is protected.

### *4.3.1 An LBS Scenario*

Consider the following scenario for safeguarding privacy in an urban LBS application using decentralization. Multiple mobile individuals are moving through an urban street network, each carrying a mobile location-aware device (like a PDA or wireless sensor node). From this point on we simply refer to the combination of a mobile individual and his or her mobile location-aware device as a *mobile agent*. In a DeSC environment, nearby mobile agents are assumed to be able to communicate with each other using short-range communication (e.g., WiFi or Bluetooth). *Points of interest* (POIs) are also distributed throughout the urban network. These POIs may be static, such as retail stores or coffee shops, or dynamic, like wireless hotspots, meetings, or traffic jams. Whenever an agent encounters a point of interest (POI), the agent's device stores that POI's ID and location. In many cases, the process of identifying POIs may be completely automatic (e.g., a device might automatically identify a wireless hotspot POI via WiFi radio frequency signals, or a retail store via RFID). However, in some cases semi-automatic or manual generation of this information is conceivable (e.g., user tagging of POI).

Given this scenario, one important task is to be able to answer $k$-nearest neighbor queries, like "Where is my nearest POI?" Answering such questions is a basic function of conventional centralized LBSs. However, in this work we look at the extent to which DeSC can be used to provide the same function, at the same time as protecting a user's location privacy. An important simplifying assumption in the following discussion and subsequent experiments is that POIs are "semi-dynamic" in the sense that POI locations are initially unknown and must be discovered by mobile agents, but POIs change with relatively low frequency when compared with the movement of agents. In other words, in the context of our scenario, once a POI has been discovered by an agent it remains valid for a relatively long period of time (when compared with the frequency with which agents move into and out of the system). Such a scenario is suitable for many (but not all) dynamic phenomena (e.g., ad hoc WiFi hotspots, which while dynamic can be usually relied upon to still be active hours or even days after first observed by an agent).

Figure 4.1 shows a simplified example of the basic scenario for five mobile agents $a$ – $e$. Mobile agents store information about POIs they have directly

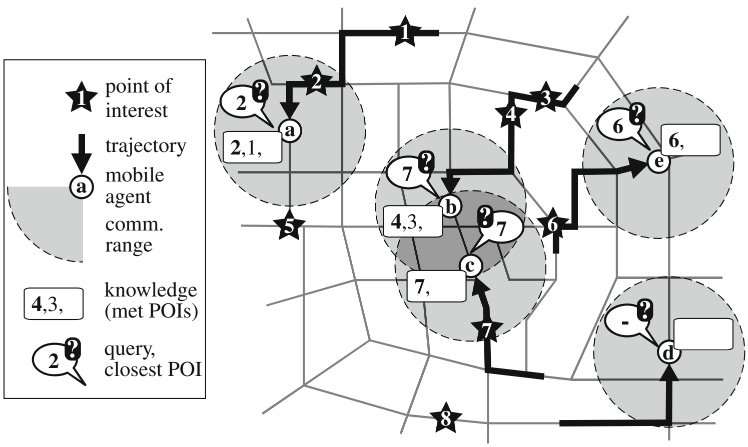



**Fig. 4.1** LBS scenario: mobile agents in an urban road network store information about POIs they have encountered, and can pose nearest neighbor queries to neighboring agents within ($n$-hop) communication range

encountered (e.g., agent *b* stores location and identity information about POIs 4 and 3). When an agent requires information about the nearest POI it can query its own stored data as well as query any neighbors within range (e.g., agent *b* can also query agent *c* for its closest POI). For clarity, the scenario in Fig. 4.1 shows a disconnected agent network, where only one-hop communication at most is possible. However, in more connected networks, the nearest neighbor query may involve multiple hops, – given that the number of hops lies below the predefined maximum and given that the network connectivity allows communication.

Thus, over time in our scenario, agents moving around the urban environment will discover more and more of the environment, enabling the system as a whole to answer nearest-neighbor queries more and more accurately; but at the same time agents will reveal information about where they are located when they pose a query, potentially reducing their location privacy.

### 4.3.2 Experimental Methodology

The performance of a decentralized, location-based nearest-POI service was explored empirically using a series of simulations. The goal of these simulations was to investigate the extent to which such decentralized LBS can provide useful services at the same time as effective spatiotemporal privacy protection. The simulation was programmed using a combination of Java and Oracle spatial DBMS, linked using JDBC (Java data-base connectivity) APIs. The experiments used a small part of Ordnance Survey's ITN data set for downtown Southampton, on the southern coast of the UK. Moving toward data models and databases that are capable of supporting AmSI is an important research problem for national mapping agencies like Ordnance Survey, especially within the context of next-generation intelligent transportation systems (ITS).

Each experiment was initialized with 100 randomly positioned POIs. A total of 100 agents were also randomly positioned in the network, initially with no knowledge of the location of the POIs. At each time step, agents move to adjacent nodes in the network. Any POIs encountered by (i.e., co-located with) an agent results in that POI's ID and location being stored in that agent's local memory. In this simplified simulation, no agents or POIs are ever destroyed or leave the simulation, and no agents or POIs are created after initialization of the simulation.

Over time, the mobile agents explore more of the network, discovering more POIs. Figure 4.2 shows a typical example of the state of the system after 50 time steps, with darker shaded circles indicating agents with knowledge of the location of more POIs. Agents can choose to engage in peer–to–peer communication with other nearby agents. The simulation allows a range of communication radii to be used, and single or multi-hop communication. In this way, agents can choose to share information about POIs they have encountered, or query nearby agents about the nearest POIs.

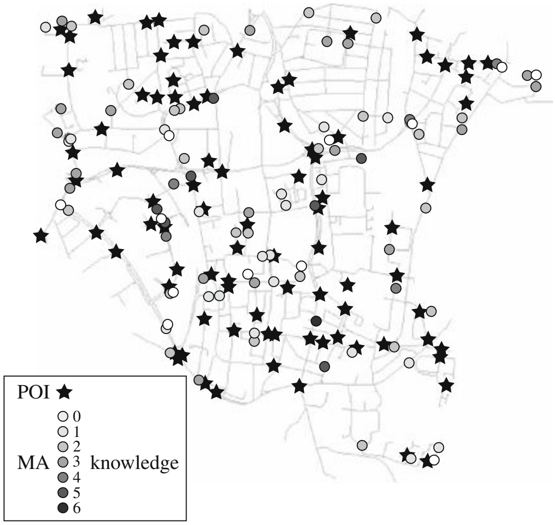



**Fig. 4.2** Example snapshot after $t = 50$ time steps of the state of a simulation, with mobile agents (MA) (*dots*) moving around the road network of central Southampton, UK. *Darker shading* of agents indicates an agent has knowledge of the location of more POIs (*stars*). Data is OS MasterMap® Integrated Transport Network™ Layer Ordnance Survey © Crown Copyright. All rights reserved

The following sections describe four increasingly sophisticated experiments that begin to uncover the behavior of a decentralized system for safeguarding trajectory privacy across a range of experimental parameters.

## 4.4 Experiment #1: Quality of Service versus Level of Privacy

The first experiments evaluated the levels of privacy and quality of service attainable by mobile agents using a decentralized nearest POI query. Initial simulations explored the situation where a *query agent* queried all other agents within its communication neighborhood at each time step for the query agent's nearest POI. The query agent and its neighboring agents used their own knowledge of POIs, if any, to compute the answer to the nearest neighbor query and return the POI location and network distance to the query agent. The best answer (in terms of shortest distance to POI) was then selected from all the responses. It was assumed that all agents possessed complete knowledge of the static road network (but, note, not the semi-dynamic POI locations). Although this is a somewhat simplifying assumption, it is not unrealistic as today many mobile devices currently store large amounts of (static) transportation network data.

In this dynamic scenario it is not possible to guarantee *exact* answers to queries. For example, from Fig. 4.1, agent $b$ may be closest to POI 6. Because none of the agents within communication range of $b$ at the time of query have encountered POI 6, agent $b$ may receive a sub-optimal nearest neighbor query response of POI 7. Given that the POIs are assumed to be (semi-)dynamic, it would similarly not be possible to guarantee exact answers to queries using a centralized approach (for example, where discovered POIs are reported back to a central server; in such a case POI 8 would still remain undiscovered). However, unlike a centralized approach, using decentralized queries enables a balance between quality of service and level of privacy to be struck. If agents choose to reveal their location information to more neighbors (e.g., using multi-hop communication or larger communication distances), then the likelihood of receiving an optimal nearest neighbor response increases, but leads to lower levels of privacy. Conversely, if an agent decides to reveal its location information to fewer neighbors, then the levels of privacy for that agent is expected to increase but at the cost of lower quality of service (i.e., fewer optimal query responses). Striking an acceptable balance between level of privacy and quality of service is one of the key goals of a good privacy protection system.

Figure 4.3 shows a typical example of how levels of privacy and quality of service (averaged across 100 simulations) vary for agents over 100 simulation time steps, referred to as a *QoS/LoP signature*. In Fig. 4.3, the quality of service is measured in terms of the inaccuracy in query responses. Two different measures of inaccuracy are used: the fraction of agents who do not receive the optimal query response, and the percentage of agents who do not receive any

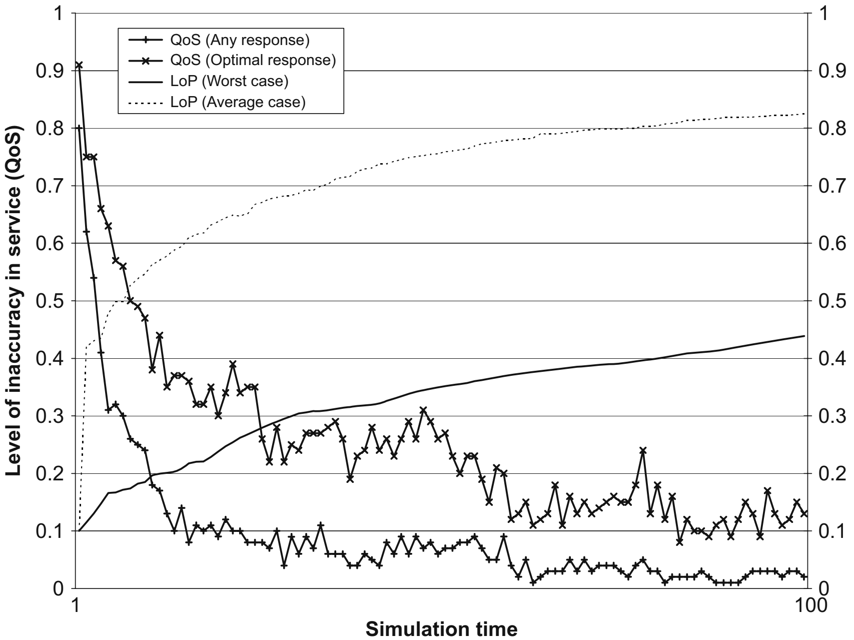



**Fig. 4.3** Typical QoS/LoP "signature", showing increasing average and worse case levels of privacy over simulation time, and increasing quality of service, in terms of decreasing inaccuracy in query responses

response at all. As would be expected, at the beginning of the simulations, levels of inaccuracy are extremely high, since the agents have not yet explored much of the environment, and so know relatively little about the POIs. However, as the simulation proceeds, quality of service rapidly improves, indicated by rapidly dropping levels of inaccuracy. At the end of the simulation, approximately 90% of agents are receiving the optimal query response (i.e., the closest POI).

The level of privacy of agents is also quantified with two different measures. The average level of privacy indicates what percentage of an agent's trajectory on average is unknown to any other arbitrarily chosen single agent. Figure 4.4 illustrates the general idea behind using knowledge of trajectories as a basis for measuring the level of privacy. In the figure, assuming agent *b* communicates its location to agent *c* during the entire duration they are in direct one-hop communication range, about 70% of agent *b*'s trajectory is unknown to *c* (a level of privacy for *b* of 0.7 with respect to *c*). The average level of privacy is initially very low, since after only a few time steps knowledge of only one or two locations for a query agent is likely to constitute a high percentage of that agent's entire trajectory, and so low levels of privacy. However, as simulation time allows the system to equilibrate, the level of privacy increases asymptotically. At the end of 100 time steps, an arbitrarily chosen node will, on average, only know less that 20% of a query agent's trajectory (indicated by an average level of privacy of more than 0.8).

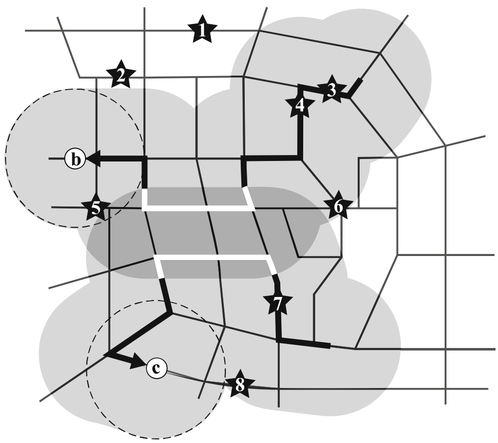



**Fig. 4.4** Level of privacy is measured as percentage of an agent's trajectory that is not known to other agents. Agent *b* does not know about ~60% of *c*'s trajectory (*c* has privacy level of 0.6 with respect to *b*); by contrast, *c* does not know about ~70% of *b*'s trajectory (*b* has privacy level of 0.7 with respect to *c*)

While average level of privacy does provide a good picture of the degree to which trajectory information about a mobile agent is "smeared" across the network, it can be argued that it is a poor overall measure of location privacy, since only *one* node is needed to potentially result in a breach in privacy. To better reflect the potential privacy risk of one agent invading another's privacy, the second measure of level of privacy is the worst case privacy, which indicates what percentage of a query agent's trajectory on average is unknown to the agent that knows *most* about that query agent. Although the worst case level of privacy is necessarily lower than the average case, the pattern is still the same, showing steadily increasing levels of privacy.

*Discussion*. Figure 4.3 shows that it is possible to achieve a balance of quality of service and level of privacy using decentralized spatial computing techniques. The levels of privacy provide a reflection of how information about an agent's location is "smeared" across spacetime. While 100% accuracy in quality of service cannot be guaranteed, relatively high qualities of service can be achieved (and, even for a centralized LBS, quality of service would never be perfect either, since typically some POIs may by chance remain undiscovered by any agent).

Most importantly, when explicitly protecting trajectory privacy and not location privacy, the suggested decentralized peer–to–peer query procedure allows individual agents the spatiotemporal accumulation of knowledge for a better QoS without compromising their privacy. In other words, when protecting trajectory privacy, levels of privacy and quality of service can *both* improve over time. This contrasts to most approaches protecting location privacy, where the collation of knowledge over time in centralized architectures leads to worsening privacy.

## 4.5 Experiment #2: Communication Network Effects

The results presented in 4.4 are typical of simulation results across a range of simulation parameters. Two such important parameters are related to the communication network used for locally querying nearby agents in a decentralized way. The communication range of agents and number of "hops" used to propagate queries through the network will affect both the balance of quality of service and level of privacy. Larger communication radii and more hops will tend to lead to improved quality of service (since it extends the communication neighborhood of a node, resulting in more nodes contributing their knowledge of the environment to the query response) as well as decreased levels of privacy (since more nodes will be informed of the location of a query agent).

The QoS/LoP signature in Fig. 4.3 showed the results for a simulation using a communication radius of about 1/16 of the diameter of the study area (1/16 being ~200 m), and using one-hop communication for queries. In other words, in Fig. 4.3 only agents that are at most 200 m from a query agent will respond to a query. Figure 4.5 summarizes the effects of changing communication radius on quality of service. A range of communication radii were tested, measured as a proportion of the size of the overall study area (so for example, a communication radius of 1/2 translates into approximately 1,600 m, half the diameter of the study area; 200 m translates into a communication radius of approximately 1/16).

As expected, quality of service increases (i.e., service inaccuracy decreases) asymptotically across all communication radii over simulation time, with larger communication radii generally leading to higher service qualities. It is noticeable that increasing the communication radius to above about 1/4 of the study area diameter leads to no further perceptible improvements in quality of service. In other words, at communication radii above 1/4 of the study area size, no additional information relevant to nearest POIs queries is being found.

*Discussion*. The effects of increasing the number of hops for queries is similar to increasing the communication range (since on average doubling the communication range means approximately as many nodes will be queried as by doubling the hop limit). Consequently, it is not surprising that similar results to those in Fig. 4.5 were obtained from similar studies of increasing hops in queries. The key message of Fig. 4.5 is that most of the information relevant to a nearest POI query can be expected to reside in close spatial proximity to the query agent. Thus, increasing hops or communication radius leads to diminishing returns in terms of increased quality of service. Conversely, level of privacy rapidly decreases for increasing communication radius and hops, such that for any communication radius of above 1/4 of the study area diameter or hop count above 3, there almost always exists at least one agent that has near-complete knowledge of a query agent's trajectory (i.e., the query agent has zero privacy). Thus, in decentralized LBS, the communication network is critical to

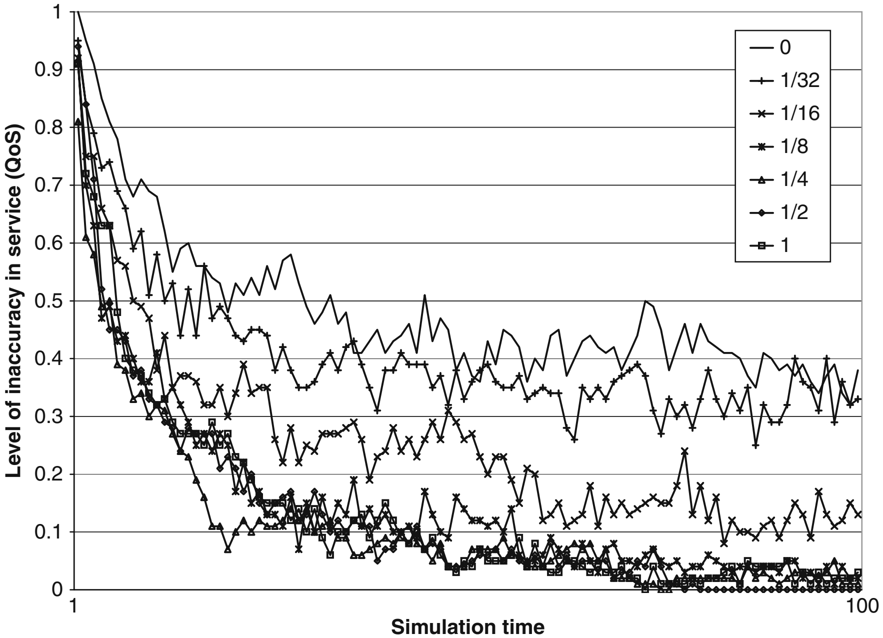



**Fig. 4.5** Effects of communication radius on quality of service. Communication radius is measured as a proportion of the size total study area, i.e., communication radius of 1 means agents can communicate directly with all other agents in the study area, radius of 0 means agents do not communicate

maintaining an acceptable balance of level of privacy and quality of service. Communication needs to be restricted enough to prevent privacy being compromised, while at the same time enabled enough to ensure reasonable quality of service.

Finally, it is also noticeable that in Fig. 4.5 a communication radius of 0 (i.e., no communication, agents can only query their own database of POIs they have already seen) still leads to relatively high qualities of service, of around 60% optimal answers. This can be ascribed to the random walk movement behavior of the agents, which means they tend to stay and thoroughly explore relatively small, spatially constrained regions of the environment. The following experiment addresses this shortcoming.

## 4.6 Experiment #3: Goal Directed Movement

As discussed in the previous section, the movement regime of agents can affect the observed quality of service and level of privacy. Random walk, used in the previous experiments, is not a realistic movement regime for most mobile humans accessing location-based services. To address this issue, the experiments were repeated using goal-directed movement, where agents move from their current location to a randomly selected destination using the shortest path. When an agent reaches its destination in our simulation, it is immediately re-tasked with a new randomly selected destination, to which it again moves along the shortest path. The effect of changing the movement behavior to goal-directed is expected to both increase level of privacy (since a query agent is expected to meet and reveal its location to a wider range of other agents less frequently) and decrease quality of service (since in general agents in a locality are less likely to have thoroughly explored and identified all the POIs in that area).

Figure 4.6 shows the signature from the same simulation scenario as Fig. 4.3 (communication radius 1/16 of the study area diameter, and one-hop communication), except where agents are using goal-directed movement instead of random walks. As expected, the levels of privacy are indeed increased, rising to 0.7 in the worst case (i.e., at the end of the simulation, there exists no agent that knows more than 30% of the trajectory of a query agent), and the quality of service is decreased somewhat when compared with agents performing random walks. It is interesting to note, however, that contrary to expectations the quality of service in terms of query agents who receive *some* answer (albeit not necessarily the optimal answer) actually *increases*. This is presumably because goal-directed movement results in greater mixing of agents, also ensuring good mixing of knowledge. Conversely, random walk tends to allow clustering of agents, and occasionally of ignorant agents who happen to have no knowledge of nearby POIs.

*Discussion*. Since goal-directed movement tends to increase mixing and aid spacetime “smearing” of an agent’s location information, agents engaged in such movement regimes can afford to reveal more about their location while still

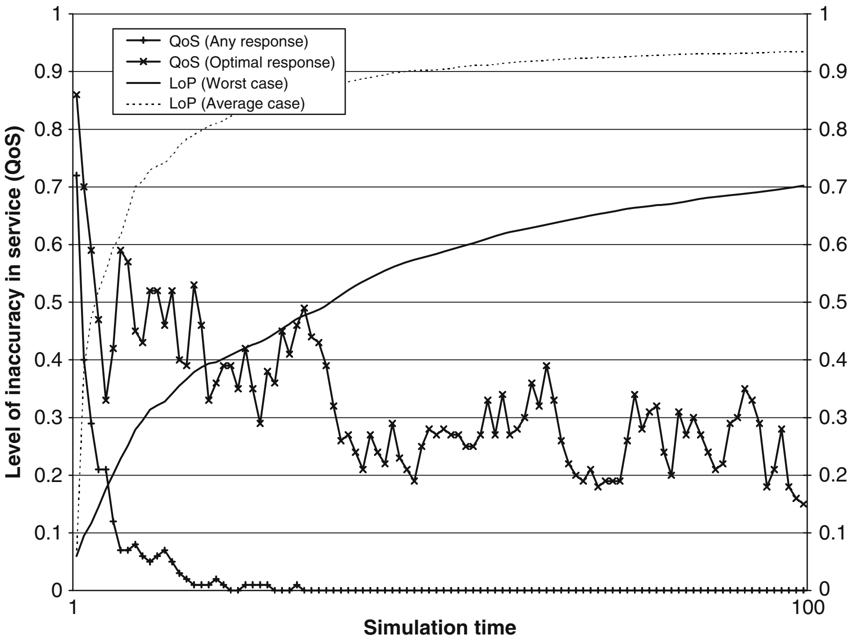



**Fig. 4.6** QoS/LoP signature for same scenario as Fig. 4.3, except where agents perform goal-directed movement rather than random walk

maintaining the same overall level of privacy. Experiments with a range of network communication parameters revealed that using one-hop communication with a communication radius of 1/8 of the study area yielded some very favorable results, with worst case levels of privacy of 0.5; average case privacy of 0.15; more than 80% of agents receiving an optimal answer; all agents receiving some answer; and more than 90% of agents receiving a nearest POI that was less than 50% further away than the optimal answer (i.e., relatively good suboptimal-answers). While the highly simplified simulation setup means that these values do not have any real meaning outside the simulation, this result does provide encouragement that by varying the decentralized query parameters it is possible to achieve high levels of privacy in concert with high qualities of service.

## 4.7 Experiment #4: Push and Pull Queries

In all the simulations discussed thus far, agents query using a "pull" strategy, where information is only ever exchanged between agents when a query is generated. However, it is also possible to design decentralized privacy protection systems that utilize "push" strategies, where information is opportunistically pushed to communication neighbors in case required at a later stage.

In contrast to the discreet pull strategy in Fig. 4.1, push is more loquacious. Using the push strategy, mobile agents synchronize their knowledge of POIs with any nearby agents they "meet" (i.e., agents that move into each other's *n*-hop neighborhood). Figure 4.7 illustrates the push strategy for the same mobile agents, POIs, and trajectories as Fig. 4.1. The key difference between the figures is that agents in Fig. 4.7 "push" information about POIs whether or not an explicit query has been received. As a result, mobile agents can accumulate information about remote POIs they have not directly encountered (e.g., agent *c* "hears" about POIs 3, 4, and 6 from agent *b*, who in turn hears about POI 6 from agent *e*).

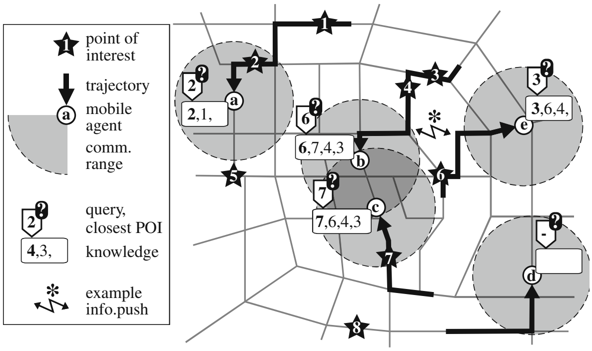



**Fig. 4.7** Push strategy: mobile agents in an urban road network store information about POIs they have encountered and share this information with other neighboring agents within (*n*-hop) communication range. Nearest neighbor queries can then be submitted to the agent's local POI database (cf. Fig. 4.1)

In a pure push strategy, agents who later require a response to a nearest neighbor query can simply query their own locally stored data about POIs without querying nearby nodes. The push strategy is similar to the concept of information dissemination in mobile ad-hoc geosensor networks (and specifically the "flooding" strategy) explored in more detail in Nittel et al. (2004a).

Figure 4.8 illustrates the typical effects of adopting a push strategy, like that explored in Nittel et al. (2004a), upon the QoS/LoP signature. Figure 4.8 uses the same basic simulation parameters as Fig. 4.6 (i.e., goal-directed movement, communication range of ~200 m). Using the push strategy, agents exchange at every opportunity all their knowledge about the locations of POIs with any other agent within direct one-hop communication range. Unlike the pull strategy, when a query agent subsequently requires information about the nearest POI, it merely queries its own stored POI data.

Figure 4.8 illustrates that the push strategy performs better in terms of QoS than the corresponding pull strategy. This is to be expected, as the push strategy leads to much more widespread dissemination of POI information, ensuring that agents are much more likely to receive high quality responses to queries. Indeed, Fig. 4.9 shows how using a push strategy can result in improved QoS almost irrespective of communication radius (since using a push strategy, all agents will "hear" about newly discovered POIs after a relatively small number of hops).

However, also as might be expected, the push strategy leads to lower levels of privacy than the pull strategy, as it requires that agents reveal information

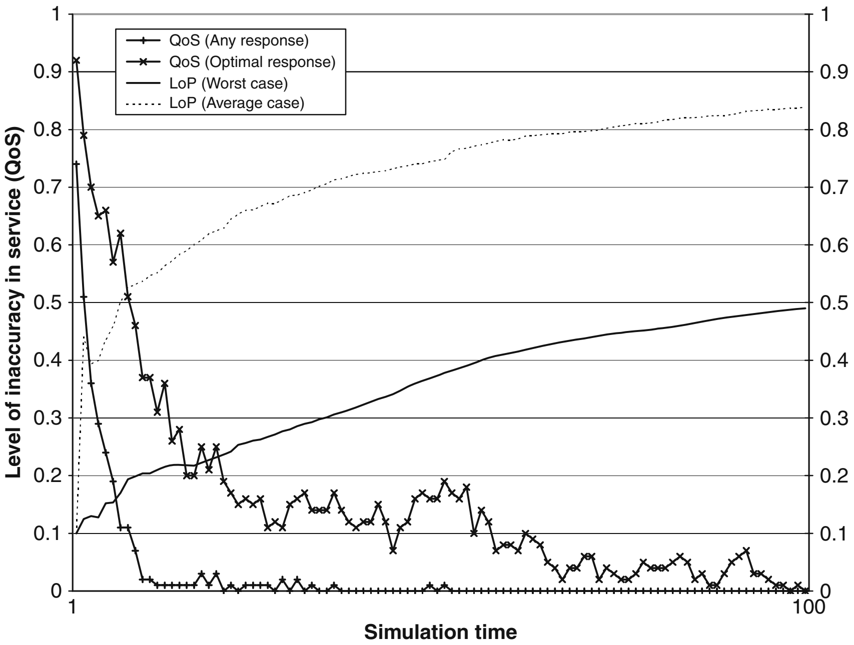



**Fig. 4.8** QoS/LoP signature for same scenario as Fig. 4.6, except where agents use push rather than pull strategy

about their location much more frequently (i.e., whenever they meet another agent). In practice, the discrepancy between the levels of privacy achievable with push and pull strategies is likely to be much greater than suggested by Figs. 4.8 and 4.6. Agents using a pull strategy are unlikely to require near-continuous responses to queries as in our simulations, needing instead to make occasional pull queries. Thus, in practice, the levels of privacy achieved by pull queries are expected to be much higher than in our simulations. However, because it is not known in advance what information will be subsequently required, there are no such obvious privacy optimizations that can be adopted using a push strategy.

*Discussion*. The results clearly indicate that while push strategies can improve QoS, equivalent pull strategies are expected to provide better privacy protection. While the focus of this work is on privacy protection, an important practical issue that is not addressed by these simulations is the communication and computational overhead of the different strategies. Push strategies generally lead to much higher volumes of data being stored and exchanged, placing strong demands on resources such as communication bandwidth, data storage space, and battery power. While these issues are not as important in most mobile computing environments as they are in, for example, wireless sensor networks, they are still a potential threat to system scalability.

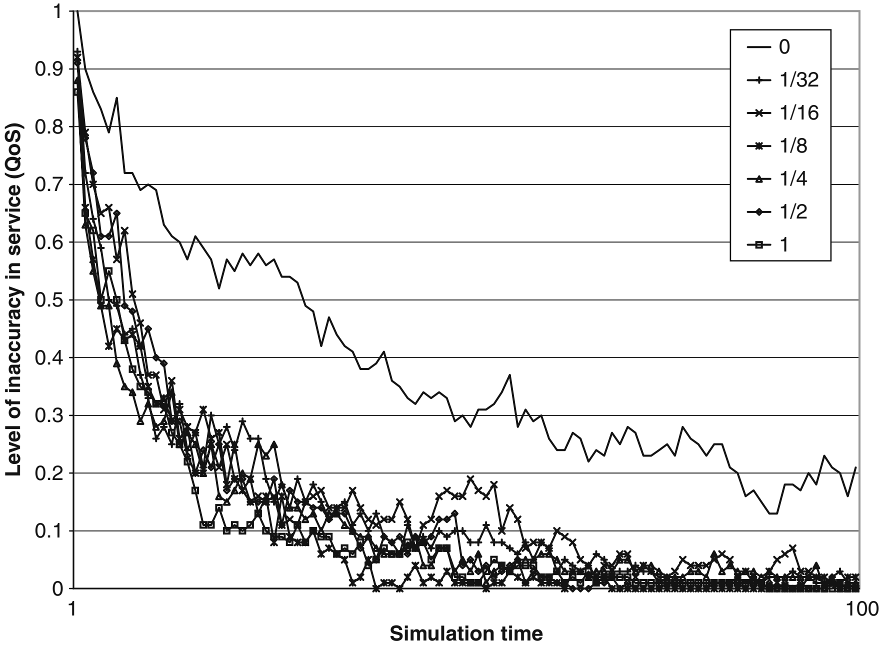



**Fig. 4.9** Effects of communication radius on quality of service for push strategy (cf. Fig. 4.5)

## 4.8 Conclusions and Outlook

This chapter promotes the vision of ambient spatial intelligence (AmSI) for urban environments. Decentralized spatial computing (DeSC) is presented as a key computational strategy enabling AmSI. It is argued that decentralization not only copes with the highly dynamic computing systems underlying AmSI, but even offers additional benefits not easily accessible with comparable centralized solutions. The additional benefit from decentralization addressed in this chapter is safeguarding privacy of LBS users. For a decentralized LBS, the balance of quality of service and level of privacy has been investigated with a comprehensive set of consecutive experiments covering various network communication parameters (one-hop vs. multi-hop communication, variable communication ranges) and motion regime parameters (random vs. goal directed walk). The experiments revealed that for carefully chosen network communication parameters the decentralized approach indeed allowed protection of privacy whilst maintaining reasonable quality of service.

One identified advantage of a decentralized LBS is its inherent affinity for dynamics. The same motivating arguments as presented for “semi-dynamic” POIs, certainly hold for even more dynamic scenarios, when POI change with high turn-over rates. When a quick response to a constantly changing topology is prime, then the intimate relation of proximity and decentralization becomes most obvious. Accepting that nearby changes are probably most relevant to other service users nearby, what could be more efficient and effective than exchanging locally relevant information only with other agents nearby? The chosen example of safeguarding privacy through service decentralization illustrates that accepting the challenge of in-network data processing not only presents additional complexity but potentially also offers benefits. Not only does local information exchange limit the depletion of global network resources, but, as has been shown in the experiments presented in this chapter, it may also safeguard privacy as agents receive the service they want with only local disclosure of their potentially sensitive location information. Clearly, it is exactly this application layer in the otherwise technology driven area of AmSI where geographic information science (GIScience) has to make its contribution, as has been shown with exploiting the spatiotemporal nature of movement for safeguarding privacy.

DeSC and geosensor networks have been widely explored for environmental monitoring, largely focusing on enabling spatial applications under the harsh technological constraints of WSN. Well defined monitoring tasks – such as for example tracking an evolving contamination area – proved to be very useful for inaugural research on DeSC. Such initial work on DeSC addressed well known GIScience challenges, including the complexity of spatial data, uncertainty, and interoperability. Additionally, the new technologies allowing DeSC also revealed the need for in-network data processing and decentralization, which in turn presents a set of new exciting challenges. First, highly dynamic peer-to-peer computing steps alongside client-server architectures. Second, decentralization promotes the notion of local knowledge as opposed to global knowledge. Consequently, decentralization and partial data processing inherently involves incomplete knowledge, requiring in turn the use of heuristics and approximation solutions. In general, these old and new challenges to DeSC also apply for urban applications. There are, however, some more challenges emerging specifically in urban DeSC.

The number of agents in AmSI is potentially much larger than a few thousand nodes deployed in an ecological monitoring network. Just imagine AmSI applications involving RFID tags on retail products. Quite in contrast to environmental applications, the networks used for urban AmSI will rarely be deployed explicitly for a DeSC application and rather make use of existing cyber-infrastructure (e.g. buddy tracking in a cell phone network). Hence, the secondary DeSC application will have much less influence on system configuration and tasking. Hence, urban DeSC applications will rather aim at exploiting given constraints than relying on lofty assumptions. Finally, urban WSNs are expected to be highly heterogeneous, potentially including nodes as different as radio relays, vehicle board-computers, handsets and RFID tags all together. DeSC for urban applications will hence face the difficulty of processing even larger and more heterogeneous data than environmental geosensor applications.

To conclude, we identify four main research and development topics for DeSC in the urban context:

– With respect to decentralized privacy protection in LBS, a further investigation of hybrid push-pull approaches, potentially reaching an optimized QoS and LoP.
– Further exploration of DeSC for safeguarding privacy in mobile communication applications (e.g., buddy tracking, child watch).
– Relaxation of the simplifying assumption of homogeneous single-purpose networks and investigation of DeSC applications for heterogeneous networks.
– The exploration of decentralized spatial data mining techniques for mobile WSN. Decentralized approaches are especially needed for the very large and very heterogeneous systems emerging urban AmSI applications.

**Acknowledgments** Patrick Laube was funded by the Australian Research Council (ARC), Discovery grant DP0662906 and the ARC Research Network on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP). Matt Duckham's research was supported by the Australian Research Council under ARC Discovery Grant DP0662906. The authors would furthermore like to thank Ordnance Survey of Great Britain.

## References

Bettini, C., Wang, X. & Jajodia, S. (2005). Protecting privacy against location-based personal identification. In W. Jonker & M. Petkovic (Eds.), *Secure data management* (pp. 185–199). (Heidelberg: Springer)

Braginsky, D. & Estrin, D. (2002). Rumor routing algorthim for sensor networks. In *Proceedings of the 1st ACM international workshop on wireless sensor networks and applications* (pp. 22–31). (Atlanta, GA: ACM Press)

Cheng, Z., & Heinzelman, W.B. (2005). Flooding strategy for target discovery in wireless networks. *Wireless Networks,* 11, 607–618

Dillenburg, J.F., Wolfson, O. & Nelson, P.C. (2002). The intelligent travel assistant. Paper presented at the IEEE 5th International Conference on Intelligent Transportation Systems, Singapore

Dobson, J.E. & Fisher, P.F. (2003). Geoslavery. *IEEE Technology and Society Magazine,* 22 (1), 47–52

Duckham, M. & Kulik, L. (2005). Simulation of obfuscation and negotiation for location privacy. In A.G. Cohn & D.M. Mark (Eds.), *Spatial information theory: international conference*, COSIT 2005, Ellicottville, NY, USA, September 14–18, 2005, (pp. 31–48). (Heidelberg: Springer)

Duckham, M., Nittel, S. & Worboys, M.F. (2005). *Monitoring dynamic spatial fields using responsive geosensor networks. Paper presented at the ACM GIS*, Bremen, Germany

Estrin, D., Govindan, R. & Heidemann, J. (2000). Embedding the internet – introduction. *Communications of the ACM*, 43(5), (38–41).

Galton, A. (2001). Space, time, and the representation of geographical reality. *Topoi-An International Review of Philosophy*, 20(2), (173–187).

Galton, A. (2003). Desiderata for a spatio-temporal geo-ontology. In W. Kuhn, M.F. Worboys & S. Timpf (Eds.), *Spatial information theory. foundations of geographic information science:*

*international conference*, COSIT 2003, Ittingen, Switzerland, September 24–28 (pp.1-12). (Heidelberg: Springer)

Galton, A. & Worboys, M. (2005). Processes and events in dynamic geo-networks. In M.A. Rodrìguez, I.F. Cruz, M.J. Egenhofer & S. Levashkin (Eds.), *Proceedings of first international conference on geospatial semantics* (pp. 45–59). (Heidelberg: Springer)

Greenfield, A. (2006). *Everyware: the dawning age of ubiquitous computing*. (Berkeley: New Riders Press)

Grenon, P. & Smith, B. (2004). SNAP and SPAN: towards dynamic spatial ontology. *Spatial cognition and computation*, 4(1), 69–103

Kaasinen, E. (2003). User needs for location-aware mobile services. *Personal and Ubiquitous Computing*, 7(1), 70–79

Karp, B. & Kung, H.T. (2000). GPSR: greedy perimeter stateless routing for wireless networks. Paper presented at the Proceedings of the 6th Annual International Conference on Mobile Computing and Networking, Boston, MA

Kido, H., Yanagisawa, Y. & Satoh, T. (2005). An anonymous communication technique using dummies for location-based services. Paper presented at the International Conference on Pervasive Services (ICPS '05)

Kosch, T., Adler, C.J., Eichler, S., Schroth, C. & Strassberger, M. (2006). The scalability problem of vehicular ad hoc networks and how to solve it. *Wireless Communications, IEEE [see also IEEE Personal Communications]*, 13(5), 22–28

Langheinrich, M. (2001). Privacy by design – principles of privacy-aware ubiquitous systems. In G. Abowd, B. Brumitt, & S. Shafer. (Eds.), *Ubicomp 2001: ubiquitous computing* (pp. 273–291). (New York: Springer)

Laube, P. & Duckham, M. (2009). Decentralized spatial data mining for geosensor networks. In H.J. Miller & J. Han (Eds.), *Geographic data mining and knowledge discovery* 2nd (pp. 409–430). (London: CRC Press)

Lynch, N. (1996). *Distributed algorithms*. (San Mateo, CA: Morgan Kaufmann)

Mauve, M., Widmer, A. & Hartenstein, H. (2001). A survey on position-based routing in mobile ad hoc networks. *Network, IEEE*, 15(6), 30–39

Nittel, S., Duckham, M. & Kulik, L. (2004a). Information dissemination in mobile ad-hoc geosensor networks. Paper presented at the Geographic Information Science, GIScience, Heidelberg

Nittel, S., Stefanidis, A., Cruz, I., Egenhofer, M.J., Goldin, D., Howard, A., Labrinidis, A., Madden, S., Voisard, A, & Worboys, M.F. (2004b). Report from the first workshop on geo sensor networks. *ACM SIGMOD Record*, 33(1)

Rule, J., McAdam, D., Stearn, L. & Uglow, D. (1980). Politics of privacy, (New York: New American Library)

Stewart Hornsby, K. & Cole, S. (2007). Modeling moving geospatial objects from an event-based perspective. *Transactions in GIS*, 11(4), 555–573

Verykios, V.S., Damiani, M.L. & Gkoulalas-Divanis, A. (2008). Privacy and security in spatiotemporal data and trajectories. In F. Giannotti & D. Pedreschi (Eds.), *Mobility, data mining and privacy* (pp. 213–240). (Heidelberg: Springer)

Winter, S. & Nittel, S. (2006). Ad hoc shared-ride trip planning by mobile geosensor networks. *International Journal of Geographical Information Science*, 20(8), 899–916

Worboys, M.F. (2001). Modelling changes and events in dynamic spatial systems with reference to socio-economic units. In A.U. Frank, J. Raper & J.P. Cheylan (Eds.), *Life and motion of socio-economic units* (pp. 129–137). (London: Taylor & Francis)

Worboys, M.F. (2005). Event-oriented approaches to geographic phenomena. *International Journal of Geographical Information Science*, 19(1), 1–28

Worboys, M.F. & Duckham, M. (2006). Monitoring qualitative spatiotemporal change for geosensor networks. *International Journal of Geographical Information Science*, 20(10), 1087–1108

Worboys, M.F. & Hornsby, K. (2004). From objects to events: GEM, the geospatial event model. In M.J. Egenhofer, C. Freksa & H. Miller (Eds.), *3rd International conference on geographic information science (GIScience 2004)* (pp. 327–343). (Heidelberg: Springer)

Yu, Y., Govindan, R. & Estrin, D. (2001). *Geographical and Energy Aware Routing: A Recursive Data Dissemination Protocol for Wireless Sensor Networks* (No. UCLA/CSD-TR-01-0023): UCLA Computer Science Department

Zhao, F. & Guibas, L.J. (2004). *Wireless sensor networks – An information processing approach*. (San Francisco, CA: Morgan Kaufmann Publishers)

# Part III
# Modeling Urban Complexity and Hierarchy

# Chapter 5
# Network Cities: A Complexity-Network Approach to Urban Dynamics and Development

**Efrat Blumenfeld-Lieberthal and Juval Portugali**

**Abstract** The aim of this proposed research is to develop an urban simulation model (USM) specifically designed to study the evolution and dynamics of systems of cities. It is innovative in three respects: First, in its structure – the proposed model is built as a superposition of two types of models; the first is an Agent Based Urban Simulation Model (ABUSM) that simulates the movement and interaction of agents in the urban space and the second is a network model that simulates the resultant urban network as it evolves. Secondly, it is innovative in the specific behavior of its agents – urban agents in our model act locally (as usual) but in order to do so, they perceive the city globally, i.e. they "think globally and act locally". In our model, the local activities and interactions of agents give rise to the global urban structure and network that in turn affects the agents' cognition, behavior, movement, and action in the city and so on in circular causality. The third aspect of innovation is connected with the specific urban phenomena it simulates – the vast majority of USM simulate the growth and expansion of urban systems but few simulate the reverse process of re-urbanization and gentrification; our model simultaneously captures the two processes and the interplay between them.

Cities are complex systems by their nature. They have originally emerged, and are still developing, out of the interactions between many agents that are located and move in space and time. These interactions result in many links that create complex networks which form the city. In the last two decades CA and AB USM have provided the main approaches to studying the dynamics of cities as complex self-organizing systems. In the last few years, models based on the new science of networks have been introduced as well. This research direction is new in the conjunction it suggests between traditional network analysis (e.g. graph theory) and complexity theory. We introduce a new dynamic model for city development, based on the evolution and structure of urban networks.

E. Blumenfeld-Lieberthal (✉)
The Yolanda and David Katz Faculty of the Arts, The David Azrieli School of Architecture, Tel-Aviv University, Ramat-Aviv, Tel-Aviv 69978, Israel
e-mail: efrat.lieberthal@googlemail.com

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_5,

**Keywords** Urban simulation model · Urban interactions · Rank size distribution

## 5.1 Introduction

The science of complex networks, whose origins are in graph theory and mechanical statistics, has developed rapidly in the last decade as part of complexity science (Watts and Strogats 1998; Barabasi and Albert 1999; Albert et al. 1999; Newman et al. 2006). It has been applied to various disciplines in order to study the topology of many large networks and to understand their development and robustness (Albert et al. 2000; Barthélémy et al. 2000; Callaway et al. 2000; Gallos et al. 2005; Tanizawa et al. 2006; Paul et al. 2006). Some examples of studied networks are the World Wide Web (Albert et al. 1999), the Internet (Cohen et al. 2000, 2001; Pastor-Satorras et al. 2001), human sexual relations (Liljeros et al. 2003), ecological networks (Solé and Montoya 2001), cellular networks (Jeong et al. 2000; Cohen and Havlin 2003), social networks (Girvan and Newman 2002; Newman et al. 2002; Watts 2002), academic citations (Jeong et al. 2003), neural networks (Grinstein and Linsker 2005) and also cities (Batty 2005).

Cities are typical examples of complex, self organizing systems (Batty and Longley 1994; Portugali 2000; Batty 2005). They have originally emerged, and are still developing, out of the interactions between many agents that are located and move in space and time. These agents are motivated by a variety of forces ranging from cognitive capabilities and needs to economic considerations, political ambitions, etc., with no central planning force that affects their behavior. These interactions result in many links that create complex networks that form the city.

Most if not all complexity theories and models have been applied to the study of cities. As a result, we now have a whole family of urban simulation models that model cities in terms of Prigogine's dissipative structures, Haken's synergetics, Mandelbrot's fractals and more (See review in Portugali 2000, 2009). In the last decade or so cellular automata and Agent Based (CA/AB) have become the main medium to simulate cities as complex systems (Batty et al. 1999; Portugali 1997, 2000; Benenson and Torrens 2004; Couclelis 1997; Silva and Clarke 2005; White and Engelen 1993), while in the last few years we see studies that model cities as networks that are typified by power law distributions (Batty 2005; Andersson et al. 2005, 2006; Porta et al. 2006; Hu et al. 2006; Jiang and Claramunt 2004).

The perception of cities in terms of the power law, commonly known as *Zipf's law* (Zipf 1941), has a long history that has recently been described by Pumain (2004, 2006). In her work, Pumain (2006) offers a thorough review on urban hierarchies and the explanations suggested for this phenomena. In addition, she explores and compares urban dynamics and other types of hierarchies, found in different complex systems. A pioneering study is that of Auerbach (1913), then one can mention Simon (1955) and Berry (1967), while more recently, and in connection with complexity theory, the studies of Mandelbrot (1983); Cameron (1990); Krugman (1996); Gabaix (1999); Blank and Solomon (2000); Batty (2006); Benguigui and Blumenfeld-Lieberthal (2006a, b, 2007). These studies, however, focus mostly on the size distribution of the cities (usually in terms of population) regardless of their spatial distribution. But most importantly, they do not explore the relation between urban networks and the resulted power laws. The approach we propose here ties the evolution of urban areas to the interaction between their components and thus provides a better understanding to city development and spatial evolution.

The aim of this work is to develop an Urban Simulation Model (USM) specifically designed to study the evolution and dynamics of systems of cities. This model is innovative in three aspects. First, it is a superposition of CA/AB and complex network analysis. Secondly, it is a synergetic inter-representation network (SIRN) USM in the sense that the local activities and interactions of agents give rise to the global urban structure and network that in turn affects the agents' cognition, behavior, movement, and action in the city and so on in circular causality (Portugali 2002, 2004). And thirdly, most of the existing USM simulate the growth and expansion of urban systems but few simulate the reverse process of re-urbanization and gentrification; our model will simultaneously capture the two processes and the interplay between them.

Currently, most of the world's population lives in urban areas. Urban growth affects the environment in many ways, which might cause ecological damages due to urban sprawl, for example. We believe that our proposed model is essential to achieve a more realistic picture of urban development, which can lead to developing tools that control and direct urban growth towards a more sustainable reality.

## 5.2 Methodology

Our first hypothesis is that urban systems can be treated as complex networks based on the interactions between the different agents that interact in the urban environment. Our second hypothesis argues that the power law characteristics of urban entities and systems are rooted in the interaction between different agents in urban environments (i.e. urban networks). We suggest that applying this approach, and understanding these networks' topology and evolution, would lead to a better understanding of urban development in both population and physical (e.g. spatial distribution) terms.

In the last decades, computer simulation models have been extensively used to simulate urban phenomena. AB and CA models have been the leading methods in the study of urban environments as complex systems. They are based on the interaction of agents that are located in space and in time, and aim to simulate the city's structure, which emerges from the interaction between its components' agents. In the present work we employ CA/AB models to study urban networks. Our objective is to understand not only the topology of urban networks but their physical characteristics as well. For this purpose we use "NetLogo", which is a cross-platform multi-agent programmable modeling environment. NetLogo is an environment particularly well suited for modeling the dynamics of complex systems. This environment allows programmers to give instructions to numerous agents which operate independently. This enables researchers to explore the connection between the micro-level behavior of individuals and the macro-level patterns that emerge from the interactions of many individuals. An additional advantage of NetLogo is that it enables researchers to combine the two main methods of our work – CA/AB and networks. Thus, NetLogo makes it possible to use the nodes in the network as agents that interact and create the links in a CA platform. In other words, it enables us to develop a model that is a superposition between two modeling methods: CA/AB and Networks.

In this work, we consider the dynamics of urban networks and their different scales. We begin to study the networks at their intra-city level (i.e. within the city itself) and follow their evolution until they exist at the intercity level (i.e. between cities). Using CA/AB models, we can control the definition of the networks and modify them for different scales. By doing so, we are able to fit the networks to realistic aspects that are related to the studied scales. For example, business centers within the city are smaller than regional or national centers. On the other hand, when considering the frequency of operating public transportation (e.g. number of buses per hour), we expect it to be higher at the intra-city level than at the intercity level.

The suggested approach allows us to follow the urban evolution based on many interactions among its components. By studying the network structure of urban entities we remain truthful to the definition of cities as complex entities that emerge from the interaction of their components (Portugali, 2000). Understanding the network's structure (topological and physical) can be used to develop planning tools for many urban aspects (e.g. sprawl, transportation).

## 5.3 The Model

We introduce a new dynamic model for city development based on the evolution and structure of urban networks. The term "Urban Network" might refer to many networks that are sometimes fundamentally different from each other. The nodes, for example, might represent formal entities such as municipalities, but they can also represent activity centers such as employment areas, business districts, or cultural centers. In a similar manner links often represent physical entities such as transportation infrastructure or activities such as commuting or telephone calls. Some examples of such networks are: (1) A commuting network in which the nodes are centers of employment and the links are defined by the volume of commuting, (2) A public transportation network where the nodes are urban centers and the links are defined by the frequency of public transportation (e.g. a minimal number of buses per hour) between them, and (3) The urban cluster networks where the nodes are clusters of built-up areas, while the links are represented by the transportation infrastructure between them.

In this work, we present an USM which is a commuting model simulating a closed urban system (the total population and the number of cities remain constant). We present a new spatial-dynamic model for city development based on the evolution and structure of urban networks. We based our model on the conceptions that large business centers attract more commuters than small ones. Following the above statement, the nodes in the model represent business centers while the links represent commuting rates. The main steps of the model are as follow (Fig. 5.1):

1. An initial number of cities (N) is randomly distributed in space.
2. Each city is given an initial population (P) between 1 and 100. The population is also distributed randomly.
3. At each step a city i is selected at random.
4. A "Probability Value" (PV) is calculated for the selected city that defines its probability to interact with each of the other cities in the model. PV is defined as: $P_i \cdot P_n / D^{\beta}$ where Pi represents the population of the chosen city i, $P_n$ represents the population of the other city n, and D represents the distance between them. This ratio is known as the "gravitation ratio" and it has been widely applied to city interactions (e.g. Reilly 1931; Wilson 1967). The values of β are predefined and can be changed for the different simulations. To allow a comparison between the different values of PV, the original values of PV have been normalized as $0 \leq VP \leq 1$.
5. In each step, if the maximum value of PV ($PV_{max}$) is larger than a predefined limit ($PV_0$), a link between $P_i$ and $P_n$ is made. Otherwise, a new city is selected at random (Step 3). The values of $PV_0$ are the inverse of the probability of a link to be created, as $PV_0$ acts as a barrier. This means that for small values of

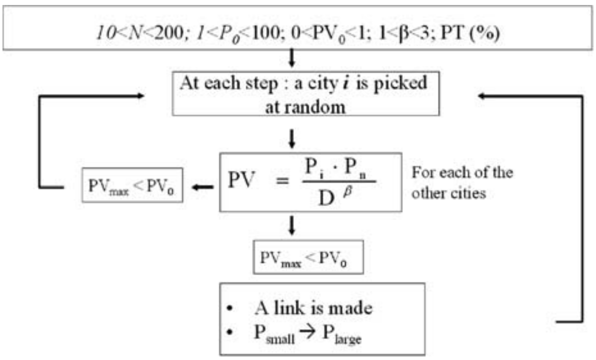



**Fig. 5.1** The steps of the preliminary model

$PV_0$ the probability of a link to be created is high and vice versa – large value of $PV_0$ means low probability for a link.

6. When a link is made there is population transfer from the smaller city to the larger one. The size of the transferred population is determined as a pre-defined percentage (PT) of the population of the smaller city.

## 5.4 Analysis

The model yields a wide variety of urban networks. Some represent highly connected, complex networks in which all nodes are connected to one another (percolation), while others represent simple systems that consist of several unconnected networks (see Fig. 5.2). The results of the model are rather interesting as they suggest some explanation to empirical findings that are observed in reality.

As mentioned earlier, the work on the size distribution of cities has occupied researchers from various disciplines related to urban activities (e.g. geography, economics, urban planning). In his study, Richard and Haggett (1967) suggested three classes of city size distributions of urban systems (see Fig. 5.3): the first refers to urban systems with a primate city, the second class refers to urban systems characterized by the power low distribution, while the third refers to urban systems in which the large cities are rather homogenously distributed. An empirical study (Benguigui and Blumenfeld-Lieberthal 2006a) showed that all the existing systems of cities, which have homogenous properties can indeed be classified into these classes. They defined these classes based on a new positive exponent $\alpha$, which can be larger, equal, or smaller than 1 (see Fig. 5.4).

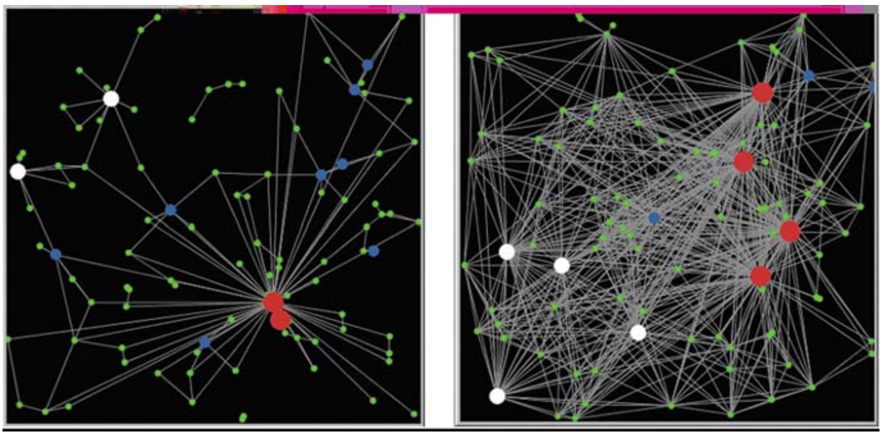


**Fig. 5.2** Two examples of urban commuting networks resulted from the model. The sizes and colors of the nodes represent different sizes of urban economic centers

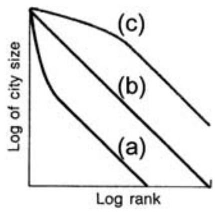



**Fig. 5.3** The three classes of city size distributions according to Richard and Haggett (1967): (*a*) distributions with primate cities, (*b*) power low distributions and (*c*) distributions in which the large cities are homogenously distributed

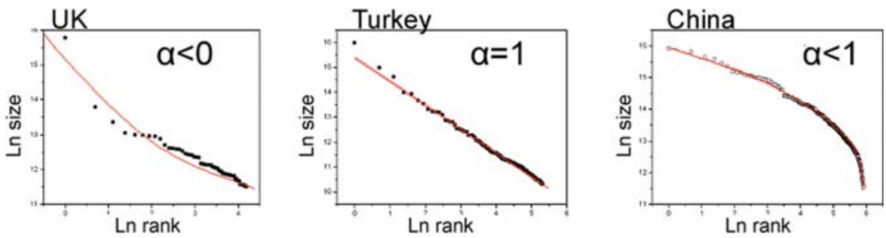



**Fig. 5.4** Real city size distributions of different countries that demonstrates Benguigui and Blumenfeld-Lieberthal (2006) classes

It appears that by changing the values of only three parameters in the basic form of our model ($PV_0$; β, PT) we get all the observed population size distributions (Fig. 5.5). In what follows we study the influence of these parameters ($PV_0$, β, and PT) on the results.

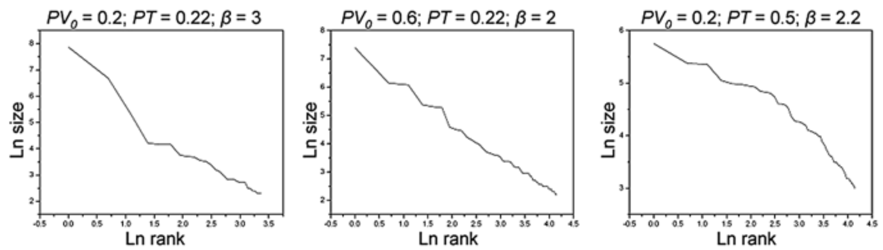



**Fig. 5.5** The three classes of city size distributions, recovered by our model

## *5.4.1 The Parameter* $PV_0$

Real urban systems are sometimes highly connected, i.e. they have many economic interactions between many cities and are thus characterized by high commuting rates between them. On the other hand, other urban systems are poorly connected, in the sense that they have few economic interactions between the cities in general, or that all the economic relations are directed toward a few large cities, and thus these cities attract most of the commuters. In order to help simulate the above different situations, observed as they are in real urban systems, we add the barrier $PV_0$. As mentioned, the values of $PV_0$ are the inverse of the level of the system's connectivity. Therefore, by decreasing $PV_0$ one can simulate urban systems with more developed economic relations.

To examine the influence of $PV_0$ on the resulted networks, in terms of their size distributions, we ran the model with fixed values of the parameters $\beta = 2$ and $PT = 10\%$. Figure 5.6 presents the rank size distributions of the population in urban systems yielded from running the model with different values of $PV_0$. We deleted from the distributions all nodes with a population smaller than 10 (which is 10% of the initial maximum population) as they appeared to correspond to urban noise, i.e. irrelevant to the level of urban activities.

It can be seen that for the small values of $PV_0$ (which represent highly connected networks, i.e. many interactions/links between the urban entities) the population distributions are not homogenous. However, when increasing the values of $PV_0$ the distributions of the population become more homogenous and around the value $PV_0 = 0.5$, they follow Zipf's law (i.e. they obey a power law with an exponent of unity). When the value of $PV_0$ is increased to values larger than 0.5, the distributions diverge from Zip's law in particular and from a power law in general.

These findings are very interesting as they suggest that the level of urban interactions might be related to the size distribution of the urban components. It implies that non homogenous size distributions (i.e. distributions that can be divided into sub-distributions) correspond to systems which are less interactive, and that size distributions that obey Zipf's law are resulted by a highly interactive system.

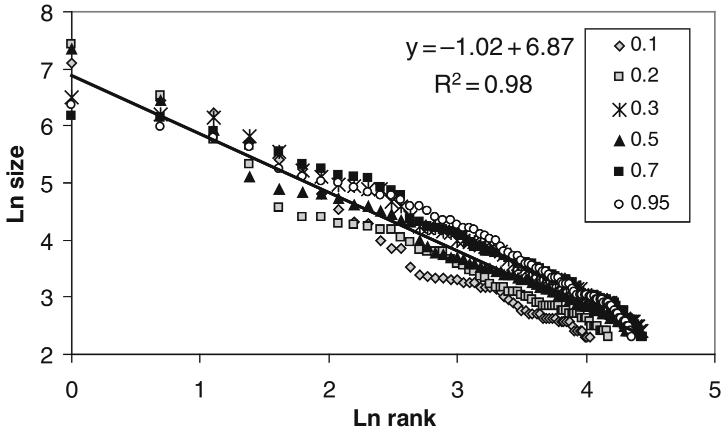



**Fig. 5.6** The rank size distributions yielded from running the model with different values of $PV_0$ and fixed parameters $\beta = 2$ and $PT = 10\%$ (the linear equation refers to the value $PV_0 = 0.5$)

## 5.4.2 *The Parameter* β

The parameter β represents the weight of the physical distance between the cities in PV (the probability of a link to be created). High values of β decrease the values of PV and vice versa: low values of β increase the values of PV. In other words, high values of β generally suggest a lower level of interactivity than lower values of β.

We ran the model with a fixed value of PT = *10%* and two fixed values of $PV_0 = 0.2$ and $PV_0 = 0.8$. Figures 5.7 and 5.8 present the size distribution of the resulted urban systems for $PV_0 = 0.2$ and $PV_0 = 0.8$ correspondingly.

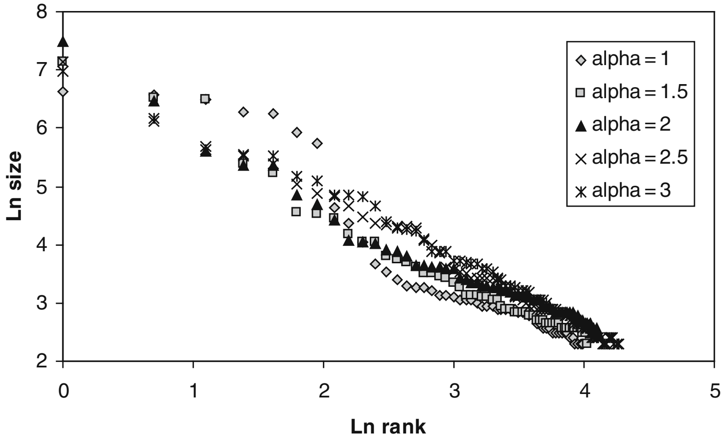



**Fig. 5.7** The size distribution for different values of β with fixed parameters $PV_0 = 0.2$ and PT = 10%

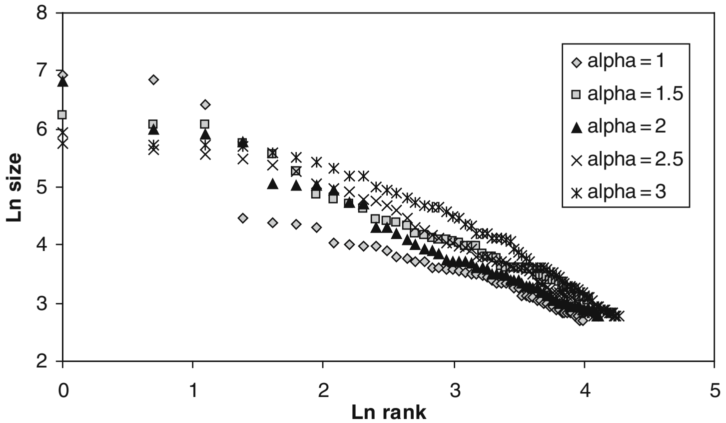



**Fig. 5.8** The size distribution for different values of β with fixed parameters $PV_0 = 0.8$ and PT = 10%

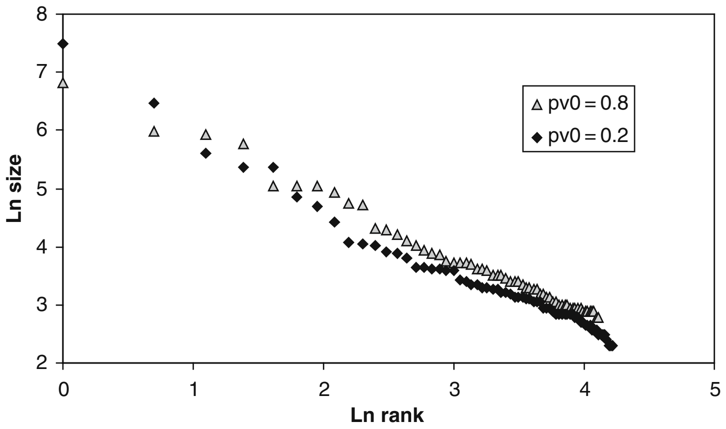



**Fig. 5.9** The size distribution for $\beta = 2$ with $PV_0 = 0.2$ and 0.8

It can be seen (Figs. 5.7 and 5.8) that in both cases of $PV_0$, the distributions are not homogenous for values of $\beta$ smaller than 2. When $\beta = 2$, the distribution follow a power law with an exponent very close to, or equal to, unity. Figure 5.9 presents the rank size distribution of two cases in which $\beta = 2$: one with $PV_0 = 0.2$ (highly connected network) and the other with $PV_0 = 0.8$ (a network with low connectivity). It can be seen that the fit of the data with a linear equation is very good (the coefficient of determination $R^2 = 0.98$ for $PV_0 = 0.2$ and 0.99 for $PV_0 = 0.8$). The exponent of the power law in the case of $PV_0 = 0.2$ is 1.05 while for $PV_0 = 0.8$ the exponent equals 0.99. These results correspond not only to a power law, but also to the specific case of Zipf's law. This result is very interesting as $\beta = 2$ is found in the original Newtonian law of gravitation. Although we can't provide an explanation for this value at this stage we hope to gain more insight on this subject in our future work. For $\beta > 2$, the behavior of the distribution is different for the different values of $PV_0$. For the small value of $PV_0$, the distributions continue to follow Zipf's law (see Fig. 5.9), while for the large value of $PV_0$ the distributions change and correspond to the group of distributions with $\alpha > 1$ (see Fig. 5.3). This suggests that for high levels of interactivity (represented in the model by small $PV_0$) the distance ($\beta$) has limited influence on size distribution that continues to follows Zipf's law even when $\beta \geq 2$.

### *5.4.3 The Parameter* PT

We introduced the parameter PT in order to control the homogeneity of the size distribution of the urban system. PT is defined as a percentage of the smaller city in each interaction that occurs in the model. We assumed that small values of PT will result in rather homogenous size distributions, while large values of this parameter will entail a situation in which most of the population (commuters in our model) concentrates in a small number of large cities.

Figure 5.10 presents the size distribution of urban systems with the fixed parameters $PV_0 = 0.5$ and $\beta = 2$. These parameters were chosen as a result of the previous analyses that indicate that they correspond to Zipf's law. The distributions related to PT = 5% are homogenous and correspond to distributions with $\alpha > 1$. The size distributions related to 10% < PT < 15% follow Zipf's law with good accuracy: the exponent is 0.97 for PT = 10% and 1 for PT = 15% with $R^2 = 0.98$ and 0.99 correspondingly. For PT = 20% the distributions become less homogenous and this trend continues and strengthen when PT is increased to values larger than 20%. For all distributions with PT ≤ 15% the number of small cities (with a population smaller than 10), which have been excluded from the distributions, does not exceed 15% of the total number of cities. On the other hand, for PT = 20%, the number of small cities increases to approximately 50% of the total number of cities; for PT = 30% the number of small cities is approximately 75%; and for PT = 50% the number of small cities is about 95% of the total number of cities. In other words, for PT ≥ 30% most of the commuters (and thus the economic activity) concentrate in a small number of cities. These findings suggest that large volumes of commuting have major influence on the size distribution of the system. While Zipf's law is obtained for low percentages of commuting, systems with large volumes of commuting rates consist of a small number of economic centers alone, and thus characterized by rank size distributions that do not follow Zipf's law.

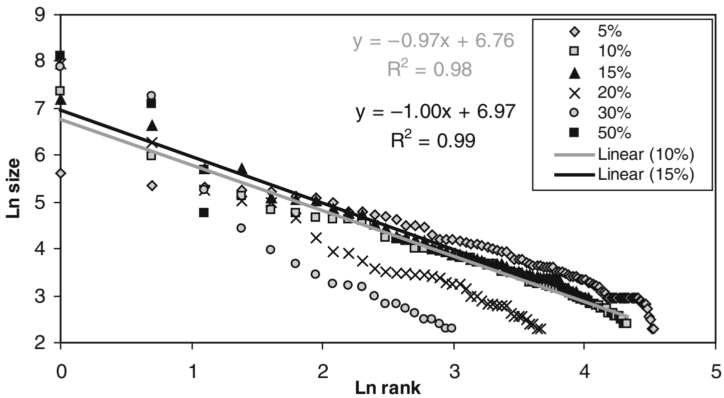



**Fig. 5.10** The size distribution for different values of PT with fixed parameters $PV_0 = 0.5$ and $\beta = 2$. (the linear equation refer to PT = 10 and 15%)

## 5.5 Conclusions

We developed an USM that is innovative in two respects: First, it is a superposition of an ABUSM that simulates the movements and interactions of agents in urban space, and a network model that simulates the resultant urban network. Secondly, its urban agents act locally (as usual) but in order to do so they perceive the city globally. In other words, they "think globally and act locally". In our model, the local activities and interactions of agents give rise to the global urban structure and network that in turn affects the agents' behavior and actions in the system and so on in circular causality.

The strength of our model is in the fact that it is based on spatial interactions and not only on population modifications. Additionally, despite its simplicity (it has a small number of variables that can be measured and controlled), it succeeds in generating the three types of distributions observed in real urban systems, and among them the power law distribution with an exponent of -1 (Zipf's law). In other words, it uses simple tools to achieve complexity and relates to the physical space using Zipf's law.

In this work we focused on the effect of urban interactions (commuting rates) on the size distribution of the entire system of cities. However, the structure and robustness of urban networks have significant importance as they represent many aspects of the city life (e.g. economic, cultural, and political aspects). Some urban networks are also related to the robustness of the city against extreme events such as terrorist attacks and natural disasters. These networks can be related to information distributions, physical evacuation, etc. In order to gain more insights to this subject, in the following stages of this project, we intend to analyze the topology of the networks resulting from the presented model, and to try to relate their characteristics to real urban phenomena.

**Acknowledgments** The authors thank Prof. L. Benguigui for stimulating conversations.

## References

Albert, R., Jeong, H. & Barabási, A.L. (1999). Diameter of the World-Wide Web. *Nature,* 401 (6749), 130–131

Albert, R., Jeong, H. & Barabási, A.L. (2000). Error and attack tolerance of complex networks. *Nature,* 406, 378–382

Andersson, C., Frenken, K. & Hellervik, A. (2006). A complex network approach to urban growth. *Environment and Planning A,* 38, 1941–1964

Andersson, C., Hellervik, A. & Lindgren, K. (2005). A spatial network explanation for a hierarchy of urban power laws. *Physica A,* 345, 227–244

Auerbach, F. (1913). Das gesetz der bevolkerungskoncentration, *Petermanns Geographische Mitteilungen,* 59, 74–76

Barabási, A.L. & Albert, R (1999). Emergence of scaling in random networks. *Science*, 286 (5439), 509–512

Barthélémy, M., Buldyrev, S.V., Havlin, S. & Stanley, H.E. (2000). Multifractal properties of the random resistor network. *Physical Review E* (Rapid Comm.), 61, R3283

Batty, M. (2005). *Cities and complexity*. (Cambridge, MA: MIT Press)

Batty, M. (2006). Rank clocks. *Nature*, 444, 592–596

Batty, M. & Longley, P. (1994). *Fractal cities: a geometry of form and function*. (London: Academic Press)

Batty, M., Xie, Y. & Sun, Z. (1999). Modeling urban dynamics through GIS-based cellular automata. *Computers, Environment and Urban systems*, 23 (3), 205–233

Benenson, I. & Torrens, P. (2004). *Geosimulation: automata-based modeling of urban phenomena*. (West Sussex, England: John Wiley and Sons Ltd)
Benguigui, L. & Blumenfeld-Lieberthal, E. (2006a). A new classification of city size distributions. Computers, *Environment and Urban Systems,* 31 (6), 648–666
Benguigui, L. & Blumenfeld-Lieberthal, E. (2006b). From lognormal distribution to power law: a new classification of the size distributions. *International Journal of Modern Physics C,* 17(10), 1429–1436
Benguigui, L. & Blumenfeld-Lieberthal, E. (2007). A dynamic model for the city size distribution beyond Zipf's Law. *Physica A*, 384, 613–627
Berry, B.J.L. (1967). *Geography of market centers and retail distribution*. (Englewood Cliffs, NJ: Prentice-Hall)
Blank, A. & Solomon, S. (2000). Power laws in cities population, financial markets and internet sites (scaling in systems with a variable number of components). *Physica A,* 287 (1–2), 279–288
Callaway, D.S., Newman, M., Strogatz, S.H. & Watts D.J. (2000). Network robustness and fragility: percolation on random graphs. *Physical Review Letters*, 85, 5468–5471
Cameron, T.A. (1990). One stage structural models to explain city size. *Journal of Urban Economics,* 27, 294–307
Cohen, R., Erez1, K., Ben-Avraham, D. & Havlin, S. (2000). Resilience of the Internet to random breakdowns. *Physical Review Letters*, 85, 4626–4628
Cohen, R., Erez1, K., Ben-Avraham, D. & Havlin, S. (2001). Breakdown of the Internet under intentional attack. *Physical Review Letters*, 86, 3682–3685
Cohen, R. & Havlin, S. (2003). Scale-free networks are ultrasmall. *Physical Review Letters,* 90, 058701
Couclelis, H. (1997). From cellular automata to urban models: new principles for model development and implementation. *Environment and Planning B,* 24(2), 165–174
Gabaix, X. (1999). Zipf's law for cities – An explanation. *Quarterly Journal of Economics,* 114, 739–767
Gallos, L.K., Cohen, R., Argyrakis, P., Bunde, A. & Havlin, S. (2005). Stability and topology of scale-free networks under attack and defense strategies. *Physical Review Letters,* 94, 188701
Girvan, M. & Newman, M. (2002). Community structure in social and biological networks. *PNAS,* 99, 8271–8276
Grinstein, G. & Linsker, R. (2005). Synchronous neural activity in scale-free network models versus random network models. *PNAS,* 102 (28), 9948–9953
Hu, M.B., Wang, W.X., Jiang, R., Wu, Q.S., Wang, B. H. & Wu, Y.H. (2006). Urban traffic dynamics: a scale-free network perspective. arXiv:physics/0606086v1 [physics.soc-ph]
Jeong, H., Neda, Z. & Barabási, A.L. (2003). Measuring preferential attachment in evolving networks. *Europhysics Letters,* 61, 567–572
Jeong, H., Tombora, B., Albert, R., Oltvai Z.N. & Barabási, A.L. (2000). The large-scale organization of metabolic networks. *Nature,* 407, 651–654
Jiang, B. & Claramunt, C. (2004). Topological analysis of urban street networks. *Environment and Planning B,* 31(1), 151–162
Krugman, P. (1996). *The self-organizing economy*. (Cambridge: Wiley-Blackwell)
Liljerosa, F., Edling, C.R. & Nunes Amarald, L.A. (2003). Sexual networks: implications for the transmission of sexually transmitted infections. *Microbes and Infection,* 5(2), 189–196
Mandelbrot, B.B., (1983). *The fractal geometry of nature*. (New York: W.H. Freeman)
Newman, M., Barabási, A.L. & Watts, D.J. (2006). *The structure and dynamics of networks*. (Princeton, NJ: Princeton University Press)
Newman, M., Watts, D.J. & Strogatz, S.H. (2002). Random graph models of social networks. *PNAS,* 99, 2566–2572
Pastor-Satorras, R., Väzquez, A. & Vespignani, A. (2001). Dynamical and correlation properties of the Internet. *Physical Review Letters,* 87, 258701

Paul, G., Sreenivasan, S., Havlin, S. & Stanley, H.E. (2006). Optimization of network robustness to random breakdowns. *Physica A,* 370, 854–862

Porta, S., Crucittib, P. & Latora, V. (2006). The network analysis of urban streets: a dual approach. *Physica A,* 369, 853–866

Portugali, J. (1997). Self-organized cities. *Futures,* 29(4/5), 353–380

Portugali, J. (2000). *Self-Organization and the City*. (Berlin: Springer)

Portugali, J. (2002). The seven basic propositions of SIRN (Synergetic Inter-Representation Networks). *Nonlinear Phenomena in Complex Systems,* 5, 428–444

Portugali, J. (2004). Toward a cognitive approach to urban dynamics. *Environment and Planning B,* 31(4), 589–613

Portugali, J. (2009). Self-organization and the city. In R.A. Meyers (Ed.), *Encyclopedia of complexity and systems* (pp. 759–761). (Berlin: Springer)

Pumain, D. (2004). Scaling laws and urban systems, Working papers of the Santa Fe Institute, [electronic version]. Retrieved from http://www.santafe.edu/research/publications/working-papers/04-02-002.pdf

Pumain, D. (2006). Alternative explanations of hierarchical differentiation in urban systems. In D. Punain (Ed.), *Hierarchy in natural and social sciences* (pp. 169–222). (Berlin: Springer)

Reilly, W.J. (1931). *The law of retail gravitation*. New York: Knickerbocker Press

Richard, J. & Haggett, P. (1967). *Models in geography: the second Madingley lectures for 1965*. (London: Methuen & Co. Ltd.)

Silva, E.A. & Clarke, K.C. (2005). Complexity, emergence and cellular urban models: lessons learned from applying SLEUTH to two Portuguese metropolitan areas. *European Planning Studies,* 13(1), 93–115

Simon, H. (1955). On a class of Skew distributions. *Biometrika*, 42, 425–440

Solé, R.V. & Montoya, J.M. (2001). Complexity and fragility in ecological networks. *Proceedings of the Royal Society B- Biological Sciences,* 268(1480), 2039–2045

Tanizawa, T., Paul, G., Havlin, S. & Stanley, H.E. (2006). Optimization of the robustness of multimodal networks. *Physical Review E,* 74, 016125

Watts, D.J., Dodds, P.S. & Newman, M. (2002). Identity and search in social networks. *Science,* 296(5571), 1302–1305

Watts, D.J. & Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. *Nature,* 393 (6684), 409–410

White, R. & Engelen, G. (1993). Cellular automata and fractal urban form: a cellular modeling approach to the evolution of urban land-use patterns, *Environment and Planning A,* 25, 1175–1199

Wilson, A.G. (1967). A statistical theory of spatial distribution models. *Transportation Research*, 1, 253–269

Zipf, G.K. (1941). *National unity and disunity*. (Bsloomington Indiana: The Principia Press)

# Chapter 6
# Scaling Analysis of the Cascade Structure of the Hierarchy of Cities

**Yanguang Chen**

**Abstract** The scaling relations indicating fractal nature can be employed to associate a hierarchy and a network of cities. The cascade structure of an urban hierarchy follows the $2^n$ rule of size class, which can be formulated as a set of exponential functions. From a pair of exponential laws, we can derive a power law indicating the scaling relation between city number and city size of different classes. The scaling exponent is the fractal dimension of the city-size distribution. Owing to the relationship of mathematical transformation between a hierarchy and a network, the scaling analysis of hierarchical structure can be used to enrich geographical spatial analysis such as spatial interaction studies. The interaction among cities from different classes has locality property, i.e., in theory, cities of one class act merely on cities of the immediate adjacent classes. Not all cities have significant impact on other cities where hierarchy is concerned. Only the interaction between adjacent cities, both in space and size, is strong enough to have a significant effect on the structure of urban systems.

**Keywords** Cascade structure · Fractal dimension · Hierarchy of cities · Network of cities · Scaling law

## 6.1 Introduction

A network of cities and towns can be regarded as a system with special structure and function, since it depends on the movement flows of labor, goods and services, ideas, and capital through the network. The movement flows suggest some kind of interaction. The fundamental notion for understanding urban systems is "interaction", which precedes the concept of "urban system". There would be no systems of cities without interaction. Action and interaction

Y. Chen (✉)
Department of Geography, College of Urban and Environmental Sciences,
Peking University, Beijing 100871, China
e-mail: chenyg@pku.edu.cn

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_6,

among cities in the evolution of an urban system can be interpreted in three aspects. The first is the temporal aspect, that is, the past state of a city affects the present state of a city and the present state will affect the future state. This identifies the autocorrelation process of urban evolution. The second is the spatial aspect, namely, one city acts on another. This defines spatial interaction and spatial autocorrelation of cities (Haggett et al. 1977). If we consider both aspects of time and space, it will involve the spatial cross-correlation process (Chen 2008). The third is the hierarchy aspect, that is, cities of one level affect cities of another level. Thus, our problems are linked to the hierarchy of cities and its cascade structure. A system of cities is in fact a cascading system with scaling invariance. Comparatively speaking, the problems of ordinal hierarchy aspect are easier to be dealt with than those in other two aspects, namely time and space, at the present time.

The cascade structure of hierarchy of cities is related to the well-known Zipf's law, which is equivalent to the Pareto distribution in theory. The Zipf's law was usually employed to describe the scaling relations between ranks and sizes of cities, but the geographical information can be obtained from the rank-size patterns is limited. The concepts from fractals provide new ways of looking at Zipf's law indicative of rank-size distribution of cities (see e.g., Chen and Zhou 2004; Frankhauser 1990; Nicolis et al. 1989; Wong and Fotheringham 1990). Using the idea from fractals, we can reconstruct the rank-size distribution model and the mathematical expression of Zipf's law. A set of exponential laws are advanced to characterize the cascade structure of urban hierarchies, and a set of power laws based on the exponential laws are proposed to reveal the scaling relations of the hierarchical structure (Chen and Zhou 2003; Chen and Zhou 2006). At the large scale, the size distributions of all cities in America, all cities in China, and all the large cities in the world, follow the scaling laws (Chen 2008). We can model the cascade structure of hierarchy of cities with scaling relations and thus develop a theory on the interaction between different classes of cities. The models and principles can be employed to aid spatial interaction analysis of systems of cities. The reason is that a hierarchy and a network represent different sides of the same coin (Batty and Longley 1994).

Recent years, in a consulting project for the comprehensive planning of Hebi Pefecture, the author has studied the spatial distribution and hierarchical structure of Henan's cities and Hebi's towns. Hebi is a medium-sized city in the Henan Province of China (geographical coordinates: N35.54, E114.11), with an urban population of 396,753 according to the Fifth National Census in 2000 (Zhou and Yu 2004b). By 2006, the non-agricultural population was around 474,629, and the urban population was about 638,220 (Appendix "The Concept of 'City' in China at present"). Of course, we are not concerned about these individual population figures, but we are interested in the status of Hebi in the hierarchy of Henan's cities, and the cascade structure of the town system within the prefecture of Hebi. Therefore, we investigate the hierarchy of cities in the Henan Province and the hierarchy of towns in the Hebi Prefecture. The results show that both the system of cities in Henan and the system of towns in

Hebi have fractal cascade structure. However, hierarchical structures based on different scale levels have some subtle differences. The theory and models on urban hierarchies is beneficial for us to conduct spatial analysis of urban evolution. In the following parts, we take the Henan province, as well as Hebi, as examples to illustrate how to bring to light the cascade structure of a hierarchy of cities, and how to apply the scaling relation to aid spatial analysis.

## 6.2 Models and Principles

### 6.2.1 *Mathematical Models*

A large number of studies show that the hierarchy of cities in a region complies with certain mathematical rules, for example, the rank-size rule, or more generally, Zipf's law (Zipf 1949). In fact, the two-parameter Zipf model is too simple to fully reveal the elaborate structure of city-size distributions. We hope to find some more accurate, deeper, and more exquisite methods. Batty and Longley (1994, pp. 140–141) once pointed out: "City size distributions display regular properties which are consistent with subdivision of a national or regional space into market areas whose decreasing size reflects the frequency of spatial dependence and the rarity value of spatial goods. Idealized size distributions can be developed by taking a primate city and its national market area, generating two next order cities, then four, and eight, and so on. . . ." The recursive subdivision of a national or regional space and the cascade structure mentioned above are the methods used in central place theory (Christaller 1966). Davis (1978) once formulated such rules for the worldwide cities at large scale. Davis's discovery can be extended to a more general mathematical form. Suppose that the cities in a region are divided into $M$ classes ordered in a top–down way. The hierarchical structure of an urban system can be described by the following exponential equations

$$f_m = f_1 r_f^{m-1}, \tag{1}$$

$$P_m = P_1 r_p^{1-m}, \tag{2}$$

where $m = 1, 2,\ldots, M$ are the order of urban size-classes, $f_m$ refers to the city number of the $m$th order class, and $P_m$ to the average size of the $f_m$ cities of order $m$, $f_1$ is the number of the first order cities (generally $f_1 = 1$), $P_1$ is the average size of the first $f_1$ cities (when $f_1 = 1$, $P_1$ is just the size of the largest city). Parameter $r_f = f_{m+1}/f_m$ is the inter-class number ratio of cities, and $r_p = P_m/P_{m+1}$ the corresponding size ratio ($r_f \geq 1$, $r_p \geq 1$). Equations (1) and (2) are derived from Davis's $2^n$ rule, which states that "the number of cities is inversely proportional to their size" (Davis 1978, p. 96). Equation (2) can also be deduced from Beckmann's (1958) urban size-class model originating from central place theory

(Liang 1999). Thereby, this couple of exponential functions can be termed as the generalized Beckmann-Davis model. Because Equations (1) and (2) have invariance under translational transform, they belong to the scaling law of translational symmetry (Williams 1997) in contrast to the scaling relation of dilation symmetry (Batty and Longley 1994).

The scaling relations between the city number $f_m$ and the average city size $P_m$ of the $f_m$ cities in the $m$th class can be derived from Equations (1) and (2). The result is

$$f_m = \mu P_m^{-D}, \tag{3}$$

in which $\mu = f_1 {P_1}^D$ is the proportionality coefficient, and the power exponent $D = \ln r_f / \ln r_p$ is the fractal dimension of the hierarchy of cities. Equation (3) has invariance under dilation transform, so it represents a scaling law of dilation symmetry. Equation (3) can be decomposed to Equations (1) and (2) through an inverse transform process. That is, a scaling law of dilation symmetry can be divided into two scaling laws of translational symmetry. The scaling law mentioned nowadays generally implies the scaling relation based on dilation transform, and it is always expressed with a power function. However, few has noticed that theoretical deduction and empirical studies will be easier if we decompose a power law indicating dilation symmetry into two exponential laws indicating translational symmetry.

Changing the order of class ranked in a top–down way to the order in a bottom-up way, we can derive the three-parameter Zipf law from Equations (1) and (2) through mathematical transformation (Chen and Zhou 2003), and the model is in the form

$$P(\rho) = C(\rho - \varsigma)^{-d_z}, \tag{4}$$

where $\rho$ is the city rank, and $P(\rho)$ is the population size of the $\rho$th city. The parameters are defined by

$$C = P_1 \left( \frac{r_f}{r_f - 1} \right)^{\ln r_p / \ln r_f}, \varsigma = \frac{1}{1 - r_f}, d_z = \frac{\ln r_p}{\ln r_f},$$

where $C$ is the proportionality coefficient ($P_1$ is the average city size of the top class), $\varsigma$ is the fine adjustment parameter, and power exponent $d_z$ can be called as Zipf's dimension, which is actually the reciprocal of the fractal dimension, namely

$$d_z = \frac{\ln r_p}{\ln r_f} = \frac{1}{D}. \tag{5}$$

In Equation (5), $D$ is the fractal dimension of the city size distribution, which is also the dimension of the hierarchy of cities. In fact, Equation (4) is theoretically equivalent to Equation (3). Mandelbrot (1982) and Winiwarter (1983) once proposed the three-parameter Zipf's law based on linguistics, biology, and economics. The underlying rationale of the Equations (1) and (2), and thus Equations (3) and (4), is the principle of entropy-maximization (Chen 2008).

The relationships between the rank-size rule, Zipf's law, and the three-parameter Zipf model should be explained in a few words here. Since the parameter $\varsigma$ is very small in general, it can be ignored in practice, thus Equation (4) can be approximately simplified to a two-parameter model

$$P(\rho) = C\rho^{-d_z}, \tag{6}$$

This is just the so-called Zipf's law, which, as indicated above, is equivalent to the Pareto distribution in theory. A great amount of studies have been made about this law associated with the rank-size distributions (Brakman et al. 1999; Carroll 1982; Gabaix and Ioannides 2004). In equation (6), $C = P(1)$ is usually regarded as the size of the largest city. Further, if $d_z = 1$ as given, then we have the rank-size rule. The one-parameter rank-size rule is a special case or an approximation of Zipf's law and the two-parameter Zipf law is a special case or an approximation of the three-parameter Zipf model.

The above models work well when describing all cities around the world with a population greater than or equal to 10,000 (Chen and Zhou 2003), or the 500 largest cities in China (Chen and Zhou 2008), or the 500 largest cities in the United States (Chen 2008). Now, what concerns us is the theoretical meanings and applicable realm of these models.

### 6.2.2 Basic Principles

A set of basic principles about the hierarchy of cities and its cascade structure can be propounded through theoretical researches and empirical analyses based on the above models, from Equations (1) to (6). The main points of the theory are summarized as follows (Chen 2008; Chen and Zhou 2003, 2004, 2008).

First, for large-scale regions such as China and the United States, and middle-scale regions such as the Henan Province of China, the cities can be classified according to the $2^n$ rule. If the number of cities in the first order class of the hierarchy is one, the number of the second order will be two, and in the third class it will be four, then eight in the fourth class, 16 in the fifth class, and so on, with the number increasing in terms of a geometric series (Fig. 6.1). The cascade distribution suggests a fractal structure of the hierarchy of cities, which is brought to light in detail by Chen and Zhou (2004). For a simple fractal, the $2^n$ rule is mathematically equivalent to the rank-size rule with exponent -1 (See Appendix "The Equivalence Relation Between the $2^n$ Rule and the Rank-Size

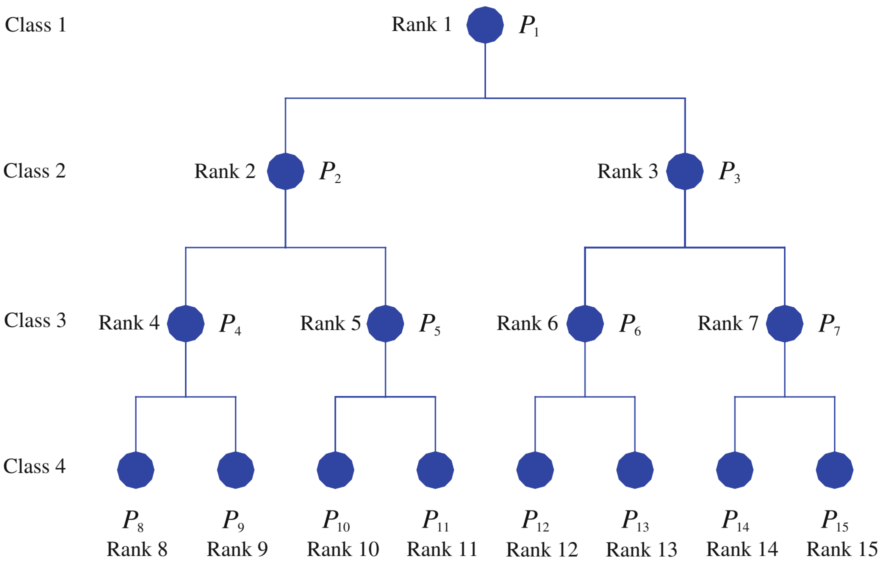



**Fig. 6.1** The sketch map of the first four classes of hierarchy of cities based on the $2^n$ principle

Rule with an Exponent-1"). Suppose that the size of largest city in the top class is one unit. The average city size of the second class will be 1/2 unit, then the third 1/4, the fourth 1/8, and the rest may be deduced by analogy, with the average size decreasing in the manner of geometric progression.

Second, the aforementioned $2^n$ principle of hierarchy of cities reflects a special cascade structure. The cascade structure is in essence a kind of self-similar organization, indicating the fractality of urban hierarchies. The $2^n$ rule is equivalent in theory to the rank-size rule associated with Zipf's law. If urban settlements in a region are ranked in order of their size, the rank-size rule states that the population of a city ranked $r$ will be $1/r$th of the size of the largest city (Knox and Marston 1998, p. 427). For a standard cascade distribution, the inter-class size ratio and the number ratio are both 2, and the fractal dimension will be $D = \ln 2/\ln 2 = 1$. The actual ratios ranges from 1 to 3, and in many cases, the values of the ratios come between 1.5 and 2.5. Therefore the fractal dimension varies from 0.5 to 1.5 approximately.

Third, the city classification based on the $2^n$ rule can be made based both on the city sizes and the number of the cities in different classes. For the size-based classification, we first determine a certain city size ratio $r_p$, and then calculate the number ratio $r_f$ by taking the average of several interclass ratios. In contrast, for the number-based classification, we first give the fixed city number ratio $r_f$, then calculate the average size ratio $r_p$. These two classification methods are theoretically equivalent to each other, but there are some subtle differences in practice. The extent to which these two classification results are similar to one another reflects the symmetrical extent of the urban hierarchy. The development of the cascade structure is better the more symmetric the hierarchy of cities becomes.

Fourth, from a long-term and evolutive point of view, the average inter-class ratios, including number ratio and size ratio, approach 2, thus the fractal dimension of the urban rank-size distribution approaches 1. Concretely speaking, if the cities are classified into cascade distribution by taking a certain number ratio $r_f = 2$, the average size ratio $r_p$ will approach 2. On the other hand, if the cities are classified by taking an ad hoc size ratio $r_p = 2$, the average city number ratio $r_f$ will approach 2. For example, according to Davis (1978, p. 96), "Among the 153 urbanized areas of the United States in 1960, the successive classes from 100,000–200,000 to 6,400,000–12,800,000 had an average inter-class ratio of 1.97." That is to say, if the size ratio $r_p = 2$ is given, then the average number ratio is $r_f = 1.97$. So, the fractal dimension of the hierarchy of the 153 US cities in 1960 is estimated as $D = \ln(1.97)/\ln(2) = 0.978$.

Fifth, without considering the effect of space, we can prove with Fourier transform and spectral analysis that the interactions between cities of different classes have a property of locality (Chen and Zhou 2008). In physics, the *principle of locality* coming from Einstein (1948), states that the distant objects cannot have direct influence on each another. In other words, an object is influenced directly only by its immediate surroundings. Regarding urban hierarchies, the cities of order $m$ only act on the cities of the same order and order $(m \pm 1)$. There is little interactions between the cities in the $m$th order class and those in the $(m \pm 2)$th or more order classes. In practice, we can use quasi-locality analysis to replace locality analysis.

Theoretically, Equations (1) and (2) can be derived by means of the entropy maximizing method, which suggests that the physical basis of the fractal structure of urban hierarchies is the entropy maximization (Chen 2008). For human geographical systems, entropy maximization implies optimization: cost least, but product most, or benefit maximum. In this sense, fractal means optimization of structure and function. Actually, not only human systems, some studies on the physical systems also support the following viewpoint: "optimality of the system as a whole explains the dynamic origin of fractal forms in nature."(Rigon et al. 1998, p. 3181) According to the above ideas, we can develop an analytical method on the rank-size distribution of cities in a region as follows. (1) Whether or not the hierarchy of cities has significant fractal properties; (2) Whether or not the fractal dimension approaches 1; (3) Whether or not the cascade structure based on city size classification is identical to that based on city number classification, namely, whether or not the two modes of classification results is symmetric. The clearer the fractal structure is, the more closely the fractal dimension approaches 1, and the more symmetrical the results of the two classifications are, the better the development of the cascade structure of the hierarchy of cities is. Furthermore, we can analyze the interactions among cities of different classes in the whole system of cities with a focus on one objective city. In the following section, we will use two examples to show that such analytical approach can be applied to aiding spatial interaction analysis.

## 6.3 Data Processing Methods and Empirical Evidences

### *6.3.1 City Classification Based on Urban Numbers*

At large geographical scales, if city-size distributions satisfy the rank-size rule, the corresponding hierarchy of cities can be approximately described with Equation (6). In this instance, the hierarchy can be characterized by a couple of exponential laws such as Equations (1) and (2), or power laws such as Equations (3) and (4). In other words, the exponential laws and the power laws are equivalent to one another in describing the cascade structures of the hierarchy of cities (Chen and Zhou 2003). Then, do the cities in the region of the middle scale, such as a province in China, or a region of small scale, such as a prefecture of a city in a province, follow the above laws? Is there any difference for the rules abided by the cities at different scales of regions? Taking the Henan Province of China as an example, we can illuminate these questions by empirical analyses. Owing to the Fifth National Census of China in 2000, we have population data of relatively high quality. After the data related to urban population were processed by Zhou and Yu (2004a, b), the urban sizes of China's cities become more credible and more comparable.

The first mode of classification is to classify the cities into different levels by taking a certain inter-class number ratio such as $r_f = 2$. Through this approach, the city numbers in different classes increase geometrically by 2, and thus the hierarchy bifurcates from top to bottom (Fig. 6.1). In other words, the city number in $(m + 1)$th class is twice the city number in the $m$th class. Then the average city sizes of different orders can be calculated in light of the classification result. The ratio of the average size of cities in one class to that in the upper order class is expected to be 2. Some values of size ratios may deviate from 2, but the mean value of all the size ratios should be close to $r_p = 2$. If so, we can judge whether the hierarchy of cities follows the $2^n$ rule of city numbers. According to Equation (5), the fractal dimension of city-size distribution is about $D = 1$.

The detailed classifying process is as follows. In the top class, the city number is 1. The largest city is in the first-order class. For the Henan Province, the top-level city is the provincial capital city, Zhengzhou. In the second-order class, the city number is 2. The cities are the 2nd and the 3rd largest cities by rank, Luoyang and Pingdingshan. In the third order, there are 4 cities, which rank from 4 to 7. Then 8 cities rank from 8 to 15 in the fourth order, 16 cities rank from 16 to 31 in the fifth order (Table 6.1). In the sixth order, the final class, the city number is expected to be 32. However, there are altogether only 38 cities in Henan. Therefore, the number of the lowest class is incomplete and the last order becomes a lame-duck class (LDC).

The next step is to calculate the total population and then the average population of cities in each class. Theoretically, the total population in different orders tends to be equal to one another, and the average population in one order is half of that in the upper order, and twice as much as the average size in the

**Table 6.1** Classification result of the cities in Henan Province based on the $2^n$ rule of urban number (2000)

| Order ($m$) | City number ($f_m$) | City | Population | City | Population | Total population | Average population ($P_m$) | Size ratio ($r_p$) |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | Zhengzhou | 2490838 | | | 2490838 | 2490838 | |
| 2 | 2 | Luoyang | 1206516 | Pingdingshan | 871274 | 2077790 | 1038895 | 2.398 |
| 3 | 4 | Xinxiang | 775941 | Jiaozuo | 634237 | 2770527 | 692632 | 1.500 |
| | | Anyang | 768992 | Nanyang | 591357 | | | |
| 4 | 8 | Kaifeng | 578635 | Xinyang | 393821 | 3229377 | 403672 | 1.716 |
| | | Puyang | 448290 | Xuchang | 373387 | | | |
| | | Shangqiu | 410648 | Zhoukou | 323738 | | | |
| | | Hebi | 396753 | Luohe | 304105 | | | |
| 5 | 16 | Zhumadian | 274023 | Xinmi | 150367 | 2596580 | 162286 | 2.487 |
| | | Sanmenxia | 227128 | Yuzhou | 148887 | | | |
| | | Linzhou | 195791 | Lingbao | 146993 | | | |
| | | Xinzheng | 194299 | Yima | 136543 | | | |
| | | Dengzhou | 174783 | Xingyang | 131489 | | | |
| | | Jiyuan | 153043 | Gongyi | 130084 | | | |
| | | Xiangcheng | 151870 | Yanshi | 118823 | | | |
| | | Huixian | 150806 | Qinyang | 111651 | | | |
| 6 (LDC) | 7 | Dengfeng | 109327 | Wugang | 97397 | 639425 | 91346 | 1.777 |
| | | Changge | 105923 | Yongcheng | 72936 | | | |
| | | Ruzhou | 101455 | Mengzhou | 52241 | | | |
| | | Weihui | 100146 | | | | | |

*Note:* The original data comes from the Fifth National Census of China in 2000. The data are processed and the population sizes of cities are calibrated by Zhou and Yu (2004). The average size ratio $r_p = 1.975$, approximates to 2.

lower order. However, urban systems are evolving systems, and due to the imperfections of the cascade structure, the hierarchy of cities in the real world usually shows the following characteristics. (1) The total urban population in the top class and that in the bottom class are often less than what they are expected to be, and only the total population of cities in the intermediate order classes tends to be constant. (2) The mean value of the urban size ratios of different adjacent orders approaches 2, but an individual ratio may be greater than 2 or less than 2. To sum up, the above formulations are all indicative of the rules in the sense of statistical average.

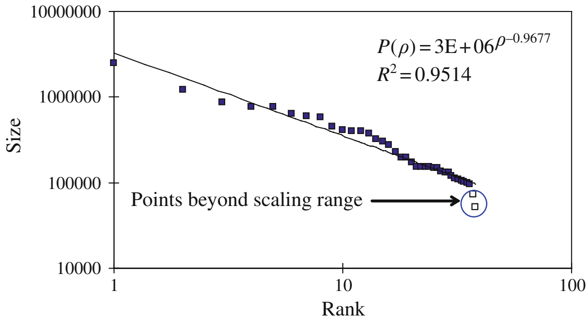
The largest city in the system, i.e., the top-class city, is usually less developed and thus to has a size smaller than expected (Zhou 1995). Sometimes the top-class city is over developed, and thus leading to a primate distribution of cities. Mostly, the city number in the last class is incomplete by the $2^n$ principle. As a result, the cascade structure of the hierarchy of cities will form a scaling range of sizes in a double logarithmic plot. Only the size distribution of the cities falling into the scaling range follows the self-similar rule really. By removing the top-class city (cities) and the last class cities, the remaining cities usually conform well to the above scaling laws.

Where the Henan Province is concerned, the city in the top class is not an exception, but the cities in the bottom class developed incompletely. Without considering the LDC, the average total population of the first five classes is about 2633022. The urban size of the first-order city is 2.398 times the average size of the second-order cities, and the average size of the second order is 1.5 times that of the third class, …, the average size of the fifth-order cities is 1.777 times that of the sixth class (Table 6.1). The average size ratio is around 1.975. Since we classified the cities according to the city number in each order, the city number in one class is rigidly twice as much as that in the upper class. That is, the number ratio is $r_f = 2$ and the size ratio $r_p = 1.975$. Thus, the fractal dimension of the hierarchy of cities in the Henan Province can be estimated as $D = \ln(2)/\ln(1.975) = 1.018$.

Theoretically, the $2^n$ rule of hierarchy of cities is equivalent to Zipf's law of city size distribution with exponent -1 (Appendix "The Equivalence Relation Between the $2^n$ Rule and the Rank-Size Rule with an Exponent-1"). The size distribution of cities in the Henan Province complies with Zipf 's law: when plotted on a double logarithmic plot, the graph of city size ($P$) and rank ($\rho$) emerge as a straight line in general. Only two data points representing the two smallest cities, Yongcheng and Mengzhou (Table 6.1), are exceptions, which are beyond the scaling range because of less development (Fig. 6.2). Since we only have two exceptional points, we can use all the data of city size and rank to estimate the model parameters. For simplicity, we fit the data to the two-parameter Zipf model, and a least squares computation yields

$$P(\rho) = 3286767.402\rho^{-0.968}$$

**Fig. 6.2** The rank-size pattern of cities in Henan Province (2000)




The goodness of fit is $R^2 = 0.951$. The estimated fractal dimension is about $D = 1/0.968 = 1.033$, close to the estimated value 1.018 given above. If the two exceptional data points are eliminated, we will have $d_z = 0.933$ with a goodness of fit of $R^2 = 0.966$. Based on the classes within scale-free range, the scaling relation between the average population of each class ($P_m$) and the corresponding number ($f_m = 1, 2, 4, 8, 16$ ) can be formulated in the following form

$$f_m = 5995360.261 P_m^{-1.062}.$$

The goodness of fit $R^2 = 0.9819$, and the fractal dimension of the hierarchy of cities is $D = 1.062$, close to the values 1.018 and 1.033 coming from other two approaches.

The advantages of city classification based on the $2^n$ rule of city number are as follows. First, the implementation procedure is simple. Generally, the city number of each order is fixed to be $2^n$ ($n = 0, 1, 2, \ldots$): 1 for the first class, 2 for the second, 4 for the third, and so on. This method is very suitable for the well-developed cascade structure. Second, the average effect is good. If not considering the local fluctuation of data, the classification result as a whole is usually close to the ideal cascade structure based on the $2^n$ principle. For the hierarchy of cities with incomplete cascade structure in the real world, the average result can still reflect the self-similarity of the network of cities.

However, there is an obvious disadvantage, that is, in the case of imperfect cascade structure, the order of a city can be artificially lifted or lowered sometimes. For example, in Table 6.1, the urban population of Qinyang is 111,651, and the population of Dengfeng is 109,327. The sizes of these two cities have no significant difference. However, because of the constant inter-class number ratio of cities, Qinyang is placed into the fifth class, while Dengdeng is put into the sixth class.

### 6.3.2 City Classification Based on Population Sizes

The second mode of classification is based on the scale of population sizes. By setting the size scales of different classes using a geometric sequence with a common ratio of 2, the number of cities in each class is determined and the result is expected to be another geometric progression. Theoretically, the size scales correspond to the average population sizes of different orders. However, for simplicity, we use the size scales as the lower bounds of urban population in different classes. In this instance, the average population size or the lower limit of size in one class is twice as much as that in the next class. That is, if the city size ratio is fixed as $r_p = 2$, then the average number ratio of different classes can be figured out. In fact, Davis's (1978) initial classification of cities is based on the urban population, and the formulation is termed as the $2^n$ rule of city sizes. Assuming that the lower limit of size in the top class is $a_1$, the lower limit of the second class is $a_2 = a_1/2$, the third one $a_3 = a_1/4$, and the rest can be determined by analogy. Generally, the lower limit of the $n$th class is $a_n = a_{n\text{-}1}/2 = a_1/2^{n-1}$ (Davis 1978, p. 96). The cities with a population greater than or equal to $a_1$ are grouped into the first class, the ones with a size greater than or equal to $a_1/2$ are grouped into the second class, and so on. In short, the cities greater than or equal to $a_1/2^{n-1}$ are grouped into the $n$th size class.

A question is what is the best value of the first lower limit $a_1$, which is the largest scale of measurement instead of the largest size of cities. The most objective and simplest approach is to let $a_1 = P_1$, where $P_1$ is the population of the largest city. The primate city in Henan is Zhengzhou with $P_1 = 2{,}490{,}838$. Therefore, the lower limit of the second class is $a_2 = P_1/2 = 1{,}245{,}419$, and so on. However, if so, no cities fall into the second order because the population of the second largest city, Luoyang, is 1,206,516. In order to reveal the cascade structure of Henan's hierarchy of cities in a more proper way, let $a_1 = 1{,}700{,}000$, which approaches the sum of the average city size of order 2 and that of order 3,1,731,527. Thus we have $a_2 = 1{,}700{,}000/2 = 850{,}000$, $a_3 = 850{,}000/2 = 425{,}000$, and so forth. In this way, the first-order city is Zhengzhou, the city number is 1; the cities in the second class are Luoyang and Pingdingshan, the city number is 2. In short, there are 6 cities in the third class, 8 in the fourth, 15 in the fifth, 5 in the sixth, and 1 in the seventh. What we expect is that there is 1 city in the first class, 2 in the second, 4 in the third,. . ., 32 in the sixth, and 64 in the seventh. However, there are only 38 cities in the study area. Both the sixth and seventh class cannot be filled up. Consequently, two LDC appear in this scheme (Table 6.2).

In this result of the classification, generally speaking, the first and the last class are usually exceptions, and can be regarded as the classes beyond the scaling range. On the one hand, theoretically, if and only if the order $m$ of an urban hierarchy is large enough, the scaling relation will become stable (Chen and Zhou 2003). Therefore, the first or even the second class is often an exceptional value empirically. On the other hand, because of undergrowth of the small cities

**Table 6.2** Classification result of the cities in Henan Province based on the $2^n$-principle of urban size (2000)

| Magnitude ($m$) | Lower limit ($a_m$) | City | Population | City | Population | City number ($f_m$) | Number ratio ($r_f$) |
|---|---|---|---|---|---|---|---|
| 1 | 1700000 | Zhengzhou | 2490838 | | | 1 | |
| 2 | 850000 | Luoyang | 1206516 | Pingdingshan | 871274 | 2 | 2 |
| 3 | 425000 | Xinxiang | 775941 | Nanyang | 591357 | 6 | 3 |
| | | Anyang | 768992 | Kaifeng | 578635 | | |
| | | Jiaozuo | 634237 | Puyang | 448290 | | |
| 4 | 212500 | Shangqiu | 410648 | Zhoukou | 323738 | 8 | 1.333 |
| | | Hebi | 396753 | Luohe | 304105 | | |
| | | Xinyang | 393821 | Zhumadian | 274023 | | |
| | | Xuchang | 373387 | Sanmenxia | 227128 | | |
| 5 | 106250 | Linzhou | 195791 | Lingbao | 146993 | 15 | 1.875 |
| | | Xinzheng | 194299 | Yima | 136543 | | |
| | | Dengzhou | 174783 | Xingyang | 131489 | | |
| | | Jiyuan | 153043 | Gongyi | 130084 | | |
| | | Xiangcheng | 151870 | Yanshi | 118823 | | |
| | | Huixian | 150806 | Qinyang | 111651 | | |
| | | Xinmi | 150367 | Dengfeng | 109327 | | |
| | | Yuzhou | 148887 | | | | |
| 6 (LDC) | 53125 | Changge | 105923 | Wugang | 97397 | 5 | 0.333 |
| | | Ruzhou | 101455 | Yongcheng | 72936 | | |
| | | Weihui | 100146 | | | | |
| 7 (LDC) | 26562.5 | Mengzhou | 52241 | | | 1 | 0.2 |

and towns, the last class or even the last two classes are often outliers. In fact, both the power law and the exponential law are empirically valid in certain ranges of scales (see e.g. Bak 1996; Chen and Zhou 2006; Clark 1951; Davis 1978). Clark (1951) once treated the central density of a city as the exceptional value of the negative exponential model, and the largest city population of Zipf's law is usually an exceptional data point (Chen 2008; Zhou 1995). After discussing "power laws and criticality", Bak (1996, p. 27) pointed out: "Of course, this (power-law relation) must eventually break down at small and large scales."

However, for cities in Henan Province, the exceptional classes are the sixth and the seventh order ones. The first class can be taken into account. Leaving out the last two classes, the first five orders conform to the $2^n$ rule of city sizes, and the average city number ratio is $r_f = 2.052$. Thus, the fractal dimension of the city size distribution can be estimated as $D = \ln(2.052)/\ln(2) = 1.037$. This value is close to the result 1.033 coming from the two-parameter Zipf law. Based on the classes within the scaling range, the inverse power relation between the average urban size of each order ($P_m$) and the corresponding city number ($f_m = 1, 2, 6, 8, 15$) can be built through a least squares calculation as follows

$$f_m = 2273750.600 P_m^{-0.990}$$

The goodness of fit is $R^2 = 0.953$. This suggests that the fractal dimension of the hierarchy of cities is $D = 0.990$, with some disagreement with the above results. If the positions of the independent and the dependent variables are exchanged, we will have $P_m = 2455312.574 f_m^{-0.963}$. In this instance, the dimension is estimated as $D = 1/0.963 = 1.038$, close to the results coming from the first approach.

The advantage of this mode of classification is that it can make the city numbers of different classes more objective if treated properly. In principle, the city number in each class is determined by means of the geometric progression based on 2. But due to the incomplete development of urban hierarchy, the cascade structure is not so standard. The realistic situation can be reflected by the classification based on city size rather than that based on city number. However, it doesn't mean that the cities of different orders can be effectively distinguished by the $2^n$ rule of city size. For example, in Table 6.2, the population of Dengfeng and Changge is respectively 109,327 and 105,923, which show no significant difference of size. However, Dengfeng is of the fifth order class because of a little larger population, while Changge is of the sixth order one. The difficulty to differentiate the cities of two adjacent orders is still inevitable.

The disadvantages of this mode of classification are as follows. (1) It is difficult to determine the first lower limit of size. As mentioned above, the simplest way is to let $a_1 = P_1$, but the effect is not satisfying in many cases. One possible way is to find an appropriate value between the size of the largest city, $P_1$, and that of the second city by rank, $P_2$. However, the selection of this value is subjective. (2) The implementation process is cumbersome. Because it is not easy to find out a proper value of $a_1$ between $P_1$ and $P_2$, sometimes it is necessary to test many times to obtain a satisfying result. In short, the operation of size based classification is more difficult than the number based classification.

For selecting $a_1$, several principles should be made clear. The first is the convenience principle. The numerical values should be simple and specific. The second is the integer-getting principle. Since it is a classification based on a multiple of 2, it is better to choose a value which can be divided exactly by $2^n$, where $n$ is an integer. For example, Davis (1978) selected the following two values as $a_1$: one is 12,800,000, which indicates the lower limit in the eighth order is 10,000,000; the other is 8,000,000, which suggests the lower limit in the eighth order is 625,000. The third is the best effect principle. The final classification result should show the cascade structure as standard as possible. The last principle is the most important one, and the difficulty of implementation just lies in this principle.

## 6.4 Applications and Discussions

### *6.4.1 Symmetry Analysis of the Two Classification Results for Regional Studies*

As far as the research on the system of cities is concerned, the above two classifications are different, but each has special use. The number-based classification can be easily used to estimate the fractal dimension of the hierarchy of cities, and the size-based classification can be used to judge the development degree of the cascade structure. The comparison between the two classifications can be employed to bring to light the symmetry of the hierarchical structure, and then evaluate the optimization degree of the urban system.

The more similar the results of the two classification results are to one another, the more symmetrical the urban system is. Under the absolute ideal condition, the first classification result is identical to the second one. The better the result of the size-based classification conforms to the $2^n$ rule; the better is the development of the fractal structure of the urban hierarchy. The better the development of the fractal structure is, the more symmetrical the urban system is. For a hierarchy of cities, symmetry implies optimization of structure. Weyl (1952/1980, p. 5) said: "Symmetry, as wide or as narrow as you may define its meaning, is one idea by which man through the ages has tried to comprehend and create order, beauty, and perfection." From the angle of view of the cascade structure of the hierarchy of cities measured by population, there are some differences between the two kinds of classification results of the cities in Henan Province. Even so, it is easy to find that they are generally similar to each other by comparing Table 6.1 with Table 6.2. Of course, the class orders of some cities in Table 6.1 are different from those in Table 6.2.

The cascade structure of the hierarchy of cities is quite stable as a whole, but the class order of individual city is variable. Some cities may transit from the lower class to the upper class and others may transit to the lower class from the upper class. This phenomenon bears an analogy to the city rank changes in the evolution of city size distributions. In fact, the former is related to the latter. Though the size distribution of cities is very steady at the macro level (Knox and Marston 1998; Madden 1956; Pumain 1997), at the micro level, the ranks of some cities change with time. Precisely because of this, the concept of the"rank clock" comes into being (Batty 2006).

### 6.4.2 Analysis of Spatial Interactions in Urban Studies

The analyses on the cascade structure of the hierarchy of cities and the analyses on the spatial structure of urban systems supplement each other. Analyses on the spatial interaction of cities involve the formulation of the gravity model. If the impedance function of the gravity model is an inverse power function, the spatial effect of cities will be an action at a distance; but if the impedance function is a negative exponential function, the spatial effect will be of locality. There is a parameter $r_0$ representing the characteristic length in the negative exponential function $f(r) = f_0\exp(-r/r_0)$, and the spatial action $f(r)$ decreases rapidly to zero when the distance $r$ is greater than the length $r_0$. In contrast, without any parameter indicating characteristic length, the inverse power function is scale-free (Fig. 6.3). It is still pending whether the impedance function of the spatial interaction model takes the form of the inverse power function or the negative exponential equation. Whatever the impedance function is, a city has action on its coordination cities.

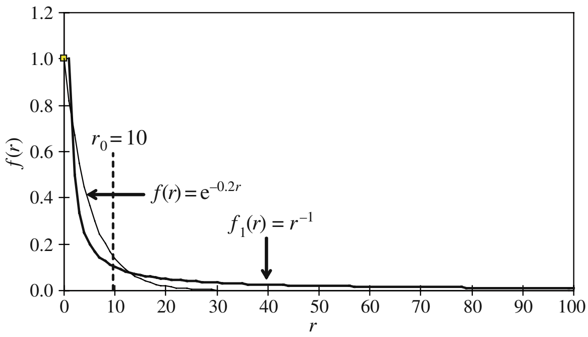



**Fig. 6.3** The distance–decay patterns based on negative exponential and inverse power functions

In theoretical geography, the coordination cities are the adjacent cities of the subject city, and they have similar size (Chen 2008; Ye et al. 2001). A group of cities with a great disparity in size cannot act as coordination cities of the subject city. On a map, the linking lines between these coordination cities around the objective city form a convex polygon (Figs. 6.4 and 6.5), termed as coordination

**Fig. 6.4** The city of Hebi and its coordination cities based on spatial relationships

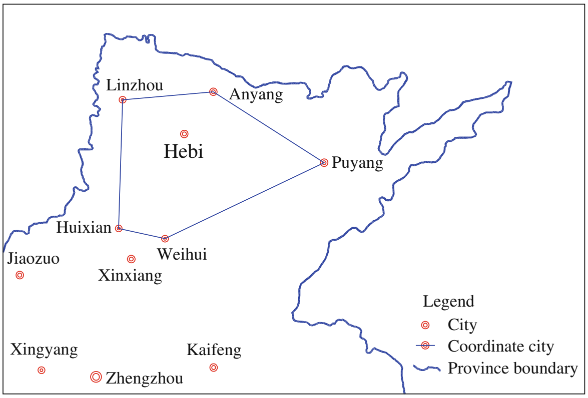
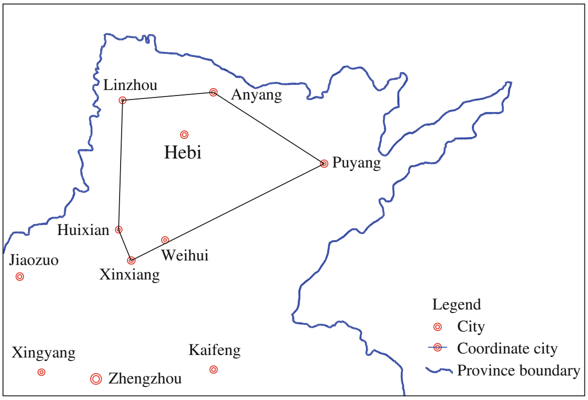



**Fig. 6.5** The revised coordination polygon of Hebi based on spatial and hierarchical contiguity




polygon of cities, and the number of the coordination cities can be called the urban coordination number. According to the central place theory, the expected coordination number of a city is 6 under ideal conditions, that is, there are 6 cities with similar size distributing immediately around a city.

Because the hierarchy of cities follows (negative) exponential law, the interaction between cities from different classes has the property of locality or quasi-locality (Chen and Zhou, 2008). In other words, a city can only act directly on the cities in the same and next immediate adjacent classes. The interaction between the cities not in immediate adjacent classes can be ignored in theory. For example, in Tables 6.1 and 6.2, Zhengzhou is in the first class and Hebi in the fourth one. According to the concept of locality, the action of Zhengzhou on Hebi can be ignored in the sense of hierarchy. Zhengzhou mainly acts on the cities in the second class, and it has very weak effect on the cities in the third class, to say nothing of the action on the cities in the fourth, fifth, and sixth class.

All in all, the interaction between cities in a region can be examined from two angles of views: the spatial view and the hierarchical view. The first view is the spatial interaction of cities. The closer the cities are, the stronger the interaction is. Therefore, the coordination cities can be influenced directly by the subject city. The second view is the interaction between cities from different classes. Cities in the same and adjacent size classes have significant interaction with each other. If two cities are immediately adjacent in both space and class, both the spatial interaction and the hierarchical interaction are significantly large; the overlap of these two interactions will lead to a strong interaction between two cities. If there is no direct effect in both space and class, the interaction between cities will be relatively weak.

The classification of the cities in Henan in Section 6.3 is not to play with numbers, but to investigate the status of the subject city, Hebi, in the hierarchy of cities in Henan. We can analyze the interaction between Hebi and the surrounding cities from the spatial and the hierarchical aspects. As viewed from size classes and spatial contiguity, the coordination cities of Hebi consist of Anyang, Linzhou, Huixian, Weihui, and Puyang, five cities altogether, which form a convex polygon: Anyang – Puyang – Weihui – Huixian – Linzhou – Anyang (Fig. 6.4). Regarding the administrative relationships, Linzhou belongs to Anyang, and Huixian and Weihui belong to Xinxiang. As viewed from the pure geometrical shape, Xinxiang should not belong to coordination cities of Hebi since the connecting lines change to a concave polygon with the addition of Xinxiang. However, if the effect of Beijing-Guangzhou railway is taken into consideration, Xinxiang should be introduced and included in the polygon.

Among these coordination cities, Anyang and Xinxiang have the closest relation to Hebi. Anyang, Hebi, and Xinxiang are connected by the Beijing–Guangzhou Railway and National Highway No.107 (i.e., G107 Line) from Beijing to Shenzhen. To a great extent, the relations between Hebi and Xinxiang are mainly owing to the transportation connection between Hebi and the provincial capital Zhengzhou, which leads to the southern places out of Henan. The relations between Hebi and Anyang are mainly owing to the connection of Hebi with cities in other provinces, especially Hebei and Shandong. The external transportation destination of Hebi includes Beijing, Qingdao, Jinan, Tianjin, Handan and some cities in Shanxi Province ("Jin" for short). The most convenient transportation from Hebi to Beijing, Tianjin, and Handan goes through Anyang. Presently, the Hebi train station is small-sized, low-graded, and with only a few number of train runs. People traveling northward usually go through the Anyang railway station, and those traveling southward go through the Xinxiang railway station to travel to cities south to Henan. This shows the geographical importance of Anyang and Xinxiang for the connection of Hebi with the outside.

The 38 cities in Henan can be grouped into 6 or 7 classes according to the $2^n$ principle, including one or two LDCs (Tables 6.1 and 6.2). The city of Hebi is in the fourth class using either the city number based classification (Table 6.1) or the city size based classification (Table 6.2). Anyang and Xinxiang are in Hebi's upper class, Linzhou and Huixian in the lower class and Weihui in the next class of Hebi's lower class. The provincial capital Zhengzhou is in the top class. Zhengzhou and Hebi are separated by two classes. In the first classification scheme, Puyang is in the same class as Hebi; while in the second classification, Puyang is in the upper class. As viewed from the population size, the cities closely related to Hebi are Anyang, Xinxiang, Puyang, Linzhou, and Huixian. On the one hand, they are adjacent to Hebi in space; on the other hand, they are in the same or immediate classes. With the overlap effects in space and hierarchy, the city planning of Hebi should consider the interaction with its coordination cities except Weihui. As Weihui is not in the same or in any of the immediate adjacent classes to Hebi, its action on Hebi is much weaker than other coordination cities. According to traffic conditions and hierarchical structure, Weihui should be removed from the coordination cities, and Xinxiang should be introduced to make the coordination polygon. The provincial capital Zhengzhou is neither adjacent to Hebi in space nor in class, the action of Zhengzhou on Hebi can be ignored. Considering spatial relationships, hierarchical contiguity, and traffic connection, we can revise the coordination polygon of Hebi as follows: Anyang – Puyang – Xinxiang – Huixian – Linzhou – Anyang (Fig. 6.5).

### 6.4.3 *Town Classification in the Small-Scaled Region*

The fractal cascade structure of systems of cities in Henan has been brought to light above. Now we come to the size distribution of towns within the administrative region of Hebi, which used to be called the Hebi Prefecture. Theoretically, the hierarchy of towns in the Hebi area complies with the $2^n$ rule. As stated above, the most convenient approach to judging the existence of the cascade structure is to test the rank-size distribution. If the size distribution of towns follows the Zipf's law, it has a cascade structure indicative of the $2^n$ principle; if not, there is no tangible cascade structure. There are two variables available for measuring the urban and rural settlements of China: one is the total population, the other the non-agricultural population. The non-agricultural population is a better measure than the total population to characterize the size distribution of cities in China (Appendix "The Concept of 'City' in China at present"). The reason is that the non-agricultural population of a city is generally consistent with its urbanized area, while the total population is in correspondence with the administrative area.

The rank-size pattern based on the non-agricultural population clearly shows that the size distribution of cities and towns in Hebi generally follows Zipf's law. However, in the double logarithmic plot, the distribution of some data points deviates from the straight trend line significantly. There are altogether 5 cities and 24 towns inside Hebi's administrative scope. The 5 cities are in fact 5 central places of 3 districts and 2 counties. The largest settlement is the Shancheng District, with a non-agricultural population of 154,405. In the second order, the population of Qibin District and Heshan District is 85,971 and 68,402 respectively, the sum of which is 154,373. The sum is close to the non-agricultural population of Shancheng District, indicating the characteristic and attributes of the $2^n$ rule. In the third order, the non-agricultural population of Zhaoge Town, Xun's Chengguan Town, Lulou Town, and Hebiji Town is 40,384, 39,706, 23,594, and 12,919 respectively, and the sum is 116,603. The result can be considered to agree with the $2^n$ rule approximately, only Lulou Town and Hebiji Town are somewhat small. The rest are all small towns, which disagree with the $2^n$ principle. Therefore, the cities and towns in Hebi Prefecture cannot be measured with the same scaling relation and it is necessary to analyze them separately.

If the five cities are not considered, the size distribution of the remaining 24 towns abide by Zipf's law (Fig. 6.6). The sizes of the last four towns are so small that the data points show a "drooping tail" in the rank-size pattern. Such is the common character of the rank-size distribution – Zipf's law holds only in certain scaling range. With the four towns beyond the scale-free range removed, we can estimate the fractal dimension of the size distribution of towns in Hebi as $D = 1/0.9818 = 1.019$. In a sense, the rank-size curve can be divided into two scaling ranges, and the point of the intersection of the two scaling ranges is rank $\rho = 20$ (Fig. 6.6). Thus, the rank-size distribution of Hebi's town can be described by a step function such as

$$P(\rho) = \begin{cases} 22612.384\rho^{-0.9818}, & \rho \leq 20 \text{ ( the 1st scaling range)} \\ 289676032473456.000\ \rho^{-\ 8.8206}, & \rho \geq 20 \text{ (the 2nd scaling range)} \end{cases}$$

A step or staircase function is in fact a piecewise constant function having finitely many pieces. However, it is hard for us to explain the parameter value of the second scaling range. In fact, the second range is modeled according to theory. Empirically, the second scaling range should be characterized by a negative exponential function rather than an inverse power function, and the equation is in the form

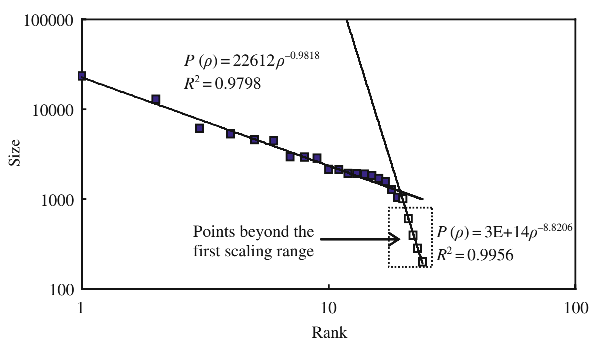



**Fig. 6.6** The rank-size pattern of the non-agricultural population of towns in Hebi Prefecture (2005)

$$P(\rho) = 2899634.074e^{-0.4012\rho}$$

The goodness of fit $R^2 = 0.9916$, and the rank $\rho$ come between 20 and 24 (the degree of freedom is low). This model seems to lend further support to the argument that the rank-size distribution for rural settlements follows the negative exponential law (Grossman and Sonis 1989; Sonis and Grossman 1984).

Further investigation suggests that the non-agricultural population of towns can be fitted to the $3^n$ rule: there is one town in the first class, three in the second class and nine in the third class. In brief, the inter-class number ratio of towns is $r_f = 3$. There should be 27 towns in the fourth class, but we have only 11 ones in reality, indicating a LDS (Table 6.3). In this instance, the calculated average size ratio is 2.802, close to the expected value $r_p = 3$. The fractal dimension of the hierarchy of towns is $D = \ln(3)/\ln(2.802) = 1.066$, close to the result estimated by Zipf's law, i.e, 1.019. Without considering the lame-duck class, the size-number scaling relation of towns can be given as $f_m = 36938.284{P_m}^{-1.045}$, and then $D = 1.045$, also close to the above results.

If the lower limit of the size in the first class is taken as $a_1 = 16{,}000$, the lower limit of the second one is $a_2 = a_1/3 = 5{,}333$, and then $a_3 = a_2/3 = 1{,}778$ for the third one. Hereby, according to the size based classification, the town number in the first class is $f_1 = 1$, then $f_2 = 3$ in the second one, and $f_3 = 11$ in the third one. The fourth order is a lame-duck class, the number of towns is $f_4 = 9$. Generally, the data can be fitted to the $3^n$ rule of size, although not perfectly symmetric to the $3^n$ rule of number based on the first mode of classification. The fractal dimension is estimated as $D = 1.096$. This suggests that the hierarchy of towns in Hebi is evolving into the cascade structure through self-organization. The difference between the system of cities and the system of towns is that the former follows the $2^n$ rule while the latter follows the $3^n$ rule. There are some relations between the $3^n$ rule and the central place system under the market principle ($k = 3$). Of course, the hierarchical structure of towns in Hebi is analyzed in the sense of statistical average, which differs from the standard central place model (Christaller 1966).

## 6.5 Concluding Remarks

The hierarchy of cities has a cascade structure at the large, middle, and the small scale, and this structure can be described by a set of exponential laws or power laws. Converting one equation for the power law of hierarchy of cities into two equations for the exponential law of cascade structure, we can get more geographical information on the network of cities and towns. An intriguing discovery is that the hierarchy of cities in the large or middle scale region complies with the $2^n$ rule, while the network of towns in the small scale region follows the $3^n$ rule. Empirically, both the $2^n$ rule and $3^n$ rule are equivalent to the rank-size rule with exponent -1. No matter whether it is a large, middle, or small scale,

**Table 6.3** The cascade structure of hierarchy of towns in the Hebi Prefecture indicating the $3^n$ principle based on the non-agricultural population (2005)

| Order ($m$) | Town | Population | Town | Population | Town | Population | Total population | Average population ($P_m$) | Size ratio ($r_p$) |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Lulou | 23,594 | | | | | 23,594 | 23,594 | |
| 2 | Hebiji | 12,919 | Baisi | 6,173 | Qiaomeng | 5,340 | 24,432 | 8,144 | 2.897 |
| 3 | Liyang | 4,586 | Xinzhen | 2,922 | Shantang | 2,133 | 25,935 | 2,882 | 2.826 |
| | Tiexiqu | 4,454 | Gaocun | 2,875 | Weixian | 1,938 | | | |
| | Xigang | 2,959 | Xiaohe | 2,154 | Tunzi | 1,914 | | | |
| 4 (LDC) | Wang-zhuang | 1,899 | Pangcun | 1,276 | Huang-dong | 398 | 11,816 | 1,074 | 2.683 |
| | Beiyang | 1,831 | Shilin | 1,044 | Jijiashan | 284 | | | |
| | Miaokou | 1,693 | Dahejian | 1,017 | Dalaidian | 200 | | | |
| | Juqiao | 1,567 | Shangyu | 607 | | | | | |

*Data source*: The *Hebi Statistical Yearbook* published by Hebi Bureau of Statistics in 2006.

the fractal dimension of urban hierarchies is around $D = 1$. The $2^n$ rule is different from the central place principle, but the $3^n$ rule seems to be statistically corresponding to the market principle dominating the central place system with $k = 3$. Two basic problems remain to be solved in the future. One is to find an effective approach to determining the form of impedance function of spatial interaction (negative exponential function or inverse power function), and the other is to find the empirical evidences to back up the locality of hierarchies of cities. Finally, the main points of the analytical process in this paper can be summarized as follows.

First, the urban settlements in a region are ranked by order of size to examine whether or not they approximately satisfy the rule of rank-size distribution. If the size distribution of cities or towns follows Zipf's law, then there exists a cascade structure. Otherwise if the cities show a primate size distribution rather than a rank-size distribution, the study area should be expanded or the largest city should be neglected as an exception. The $2^n$ rule should be tried first, then the $3^n$ rule, or even the $4^n$ rule, until arriving at the best classification. With the classification results, the fractal dimension of the hierarchy of cities can be estimated from three aspects: the first is based on the size ratio and the number ratio; the second based on the scaling relations between the average sizes and the numbers of cities in different classes; the third on Zipf's law. The three results are theoretically equivalent and empirically similar to one another.

Second, group cities in the region into different classes by using two methods: one is based on city number and the other, based on city size. Thus we will get two kinds of cascade structure of the hierarchy of cities. The similarities and differences between the two structures can be used to evaluate the symmetry of the urban system. The more similar the two results are, the better the symmetry of the cascade structure is, and thus the better the development of the fractal structure of the urban hierarchy is. Fractality is the optimization structure of nature, and the fractal body can fill up space most efficiently. The development degree of the fractal structure implies the evolution level and the robustness of urban systems. If the two cascade structures are asymmetric, the hierarchy of cities should be optimized by giving priority to the development of some cities so that all the cities follow the rank-size distribution.

Third, examine the strength of the interaction between one city and the other cities by integrating the two classification results. The hierarchy of cities has the property of locality: cities of one class mainly act on cities in the same or the immediate adjacent classes. The interaction between cities in one class and those in the non-immediate adjacent classes is too weak to count in theory. According to the idea of locality and hierarchy, we can find the cities that have close relationship with the subject city easily. Finally, we can use the interaction between cities of different classes to enrich the spatial analysis of cities. Two cities have strong interaction when they either have strong spatial or hierarchical interaction or both; otherwise, they have weak interaction or nearly no significant interaction.

**Acknowledgments** This research was supported financially by the National Natural Science Foundation of China (Grant no.40771061) and the Ministry of Science and Technology of China (Grant no. 2007FY140800). The author wishes to thank Dr. Shiguo Jiang at Ohio State University for his editorial help, and the anonymous reviewer for his/her insightful comments which have improved the quality of this work.

# Appendix

## *The Concept of "City" in China at Present*

In China, there has been no canonical and accepted definition of "city" so far. The concept of a city is very confusing and complicated. A city in China, which consists of rural and urban areas, is not a non-agricultural entity, but a large region with a system of cities and towns and their hinterlands. In the sense of its administration, a city is composed of two parts: counties and districts under the jurisdiction of the city. The districts under the jurisdiction of a city can be divided into two parts: urbanized area and non-urbanized area. In other words, the urbanized area of a city is inside the region comprising several districts. A county is always a vast rural region, within which there are always a capital town (*Chengguanzhen* in Chinese) and many other small towns. The capital town is usually the largest town in the county independent of the main body of the city. Therefore, the population in a "city" can be classified in three fashions: (1) population of counties (under the jurisdiction of the city) and population of districts (under the jurisdiction of the city); (2) agricultural population and non-agricultural population; (3) rural population and urban population. Taking Hebi as an example, we can tabulate the three sorts of population classification systems as follow (Table 6.4). According to the *Hebi Statistical Yearbook*, the so-called urban population of Hebi (638,220) includes three components: agricultural population in districts (171,880), non-agricultural population in districts (346,490), and (a great majority of) non-agricultural population in counties (128,139). The sum is $171{,}880 + 346{,}490 + 128{,}139 = 646{,}509 > 638{,}220$, and the difference is merely 8,289. Among all these kinds of populations, only the non-agricultural population corresponds to the urban entity.

**Table 6.4** The three kinds of classification systems of Hebi's population (2006)

| The first and second classification systems | | | | The third classification system | | |
|---|---|---|---|---|---|---|
| Type | Agricultural population | Non-agricultural population | Total | Type | Population | Proportion (%) |
| Districts | 171,880 | 346,490 | 518,370 | Urban | 638,220 | 44.18 |
| Counties | 798,082 | 128,139 | 926,221 | Rural | 806,371 | 55.82 |
| Total | 969,962 | 474,629 | 144,4591 | Total | 144,4591 | 100.00 |

*Data source*: The *Hebi Statistical Yearbook* published by the Hebi Bureau of Statistics in 2007.

## *The Equivalence Relation Between the $2^n$ Rule and the Rank-Size Rule with an Exponent -1*

According to the *Oxford Dictionary of Geography*, the rank-size rule states that, if the population of a town is multiplied by its rank, the sum will equal the population of the highest ranked city. For simplicity, let's take the size of the largest city $P_1 = 1$, then the size of the $k$th town by rank $P_k$ will be $1/k$. In order to show the equivalence relation between the $2^n$ rule and the rank-size rule with an exponent -1, we can make a simple mathematical experiment. (1) Produce a harmonic sequences based on the rank-size rule, $\{1/k\}$, where $k = 1, 2, 3, \ldots$. (2) Create a hierarchy by dividing the harmonic sequences into $M$ classes in a top-down way according to the $2^n$ rule: [1]; [1/2, 1/3]; [1/4, 1/5, 1/6, 1/7]; ...; $[1/(2^{m-1}), 1/(2^{m-1}+1), \ldots, 1/(2^m-1)]$;. . ., and so on. In this instance, the interclass number ratio is $r_f = f_{m+1}/f_m = 2$, where $m = 1, 2, \ldots, M$, and $M$ is a large number. (3) Examine the cascade structure of the hierarchy. The total population of the first class is 1, the second class $1/2 + 1/3 = 0.8333$, the third class $1/4 + 1/5 + 1/6 + 1/7 = 0.7595$, and so forth (Table 6.5). If $m$ is large enough, we will have total population in $m$th class of $f_m P_m = \ln(2) = 0.6931$ in theory. By limit analysis, the size ratio is

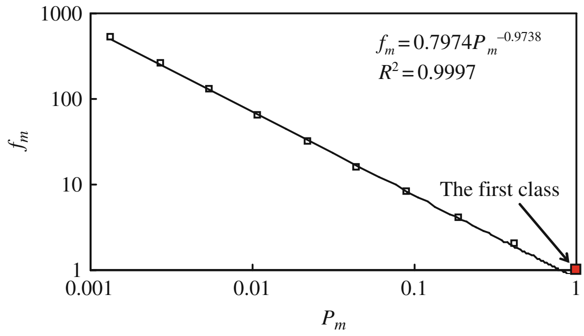
$$r_p = \lim_{m\to\infty} \frac{P_m}{P_{m+1}} = \frac{f_{m+1}}{f_m} = 2 = r_f,$$

thus the fractal dimension is $D = \ln r_f/\ln r_p \to 1$. The result of the mathematical experiment is shown in Table 6.5. Without considering the first value of size ratio 2.4 (because the corresponding standardized value $2.352 > 2\sigma = 2$, where $\sigma = 1$ is standard error), we have an average common ratio of $\bar{r}_p = 2.047$. Accordingly, the fractal dimension can be estimated as $D = \ln r_f/\ln \bar{r}_p = 0.967$.

**Table 6.5** The result of mathematical experimentation by converting the harmonic sequences based on the rank-size rule into geometric sequences based on the $2^n$ rule ($r_f = 2$)

| Order ($m$) | City number ($f_m$) | Total city population size ($f_m P_m$) | Average city size ($P_m$) | Size ratio ($r_p$) |
|---|---|---|---|---|
| 1 | 1 | 1.0000 | 1.0000 | |
| 2 | 2 | 0.8333 | 0.4167 | 2.400 |
| 3 | 4 | 0.7595 | 0.1899 | 2.194 |
| 4 | 8 | 0.7254 | 0.0907 | 2.094 |
| 5 | 16 | 0.7090 | 0.0443 | 2.046 |
| 6 | 32 | 0.7010 | 0.0219 | 2.023 |
| 7 | 64 | 0.6971 | 0.0109 | 2.011 |
| 8 | 128 | 0.6951 | 0.0054 | 2.006 |
| 9 | 256 | 0.6941 | 0.0027 | 2.003 |
| 10 | 512 | 0.6936 | 0.0014 | 2.001 |
| . . . | . . . | . . . | . . . | . . . |
| $M$ | $2^{M-1}$ | ln(2) | $\ln(2)/(2^{M-1})$ | 2 |

**Fig. 6.7** The scaling relation between city numbers and average sizes of different classes




The scaling relation between the city number $f_m$ and the average size $P_m$ is illustrated by Fig. 6.7.

The $2^n$ rule, $3^n$ rule, and even $4^n$ rule can be derived mathematically from the rank-size rule with exponent -1. Despite this, theoretically, the first several classes depart from the scaling range to some extent (Fig. 6.7, Table 6.5). As for the empirical data, the last class always drops out of the scaling range due to the underdevelopment of small cities and towns. Therefore, the exponential laws and the power laws of the hierarchy of cities are always invalid at the extreme scales, i.e. the very large and small scales.

## References


Bak, P. (1996). *How nature works: the science of self-organized criticality*. (New York: Springer)

Batty, M. (2006). Rank clocks. *Nature*, 444, 592–596

Batty, M. & Longley, P.A. (1994). *Fractal cities: a geometry of form and function*. (London: Academic Press)

Beckmann, M.J. (1958). City hierarchies and distribution of city sizes. *Economic Development and Cultural Change*, 6, 243–248

Brakman, S., Garretsen, H., Van Marrewijk, C. & Van Den Berg, M. (1999). The return of Zipf: Towards a further understanding of the rank-size distribution. *Journal of Regional Science*, 39(1), 183–213

Carroll, C. (1982). National city-size distributions: what do we know after 67 years of research? *Progress in Human Geography*, 6(1), 1–43

Chen, Y.G. (2008). *Fractal urban systems: scaling, symmetry, and spatial complexity*. (Beijing: Scientific Press) (in Chinese)

Chen, Y.G. & Zhou, Y.X. (2003). The rank-size rule and fractal hierarchies of cities: mathematical models and empirical analyses. *Environment and Planning B: Planning and Design*, 30, 799–818

Chen, Y.G. & Zhou, Y.X. (2004). Multifractal measures of city-size distributions based on the three-parameter Zipf model. *Chaos, Soliton and Fractals*, 22, 793–805

Chen, Y.G. & Zhou, Y.X. (2006). Reinterpreting central place networks using ideas from fractals and self-organized criticality. *Environment and Planning B: Planning and Design*, 33(3), 345–364

Chen, Y.G. & Zhou, Y.X. (2008). Scaling laws and indications of self-organized criticality in urban systems. *Chaos, Soliton and Fractals*, 35(1), 85–98

Christaller, W. (1966). *Central places in Southern Germany*. (New Jersey: Prentice Hall)
Clark, C. (1951). Urban population densities. *Journal of Royal Statistical Society*, 114, 490–496
Davis, K. (1978). World urbanization: 1950–1970. In I.S. Bourne & J.W. Simons (Eds.), *Systems of cities* (pp.92–100). (New York: Oxford University Press)
Einstein, A. (1948). Quanten-mechanik und wirklichkeit (Quantum mechanics and reality) (In German). *Dialectica*, 2, 320–324
Frankhauser, P. (1990). Aspects fractals des structures urbaines. *L'Espace Géographique*, 19(1), 45–69
Gabaix, X. & Ioannides, Y.M. (2004). The evolution of city size distributions. In J.V. Henderson & J.F. Thisse (Eds.), *Handbook of urban and regional economics* (pp. 2341–2378). (Amsterdam: North-Holland Publishing Company)
Grossman, D. & Sonis, M. (1989). Reinterpretation of the rural rank-size: examples from England and the land of Israel. *Geographical Research Forum*, 9, 67–108
Haggett, P., Cliff, A.D. & Frey, A. (1977). *Locational analysis in human geography*. (London: Edward Arnold Ltd)
Knox, P.L. & Marston, S.A. (1998). *Human geography: places and regions in global context. Upper Saddle River*. (New Jersey: Prentice Hall)
Liang, J.S. (1999). Beckmann urban class-size model in inverse order and its power to explain rank-rule. *Journal of Beijing Normal University* (*Natural Science*), 35(1), 132–135 (in Chinese)
Madden, C.H. (1956). On some indications of stability in the growth of cities in the United States. *Economic Development and Culture Change*, 4, 236–252
Mandelbrot, B.B. (1982). *The fractal geometry of nature*. (New York: W. H. Freeman and Company)
Winiwarter, P. (1983). The genesis model – Part II: frequency distributions of elements in selforganized systems. *Speculations in Science and Technology*, 6(2), 103–112
Nicolis, G., Nicolis, C. & Nicolis, J. (1989). Chaotic dynamics, Markov partitions, and Zipf's law. *Journal of Statistical Physics,* 54, 915–924
Pumain, D. (1997). City-size dynamics in urban systems. In S.E. Van der Leeuw & J. McGlade (Eds.), *Time, process and structured transformation in archaeology* (pp. 97–117). (London and New York: Routledge)
Rigon, R., Rodriguez-Iturbe, I. & Rinaldo, A. (1998). Feasible optimality implies Hack's law. *Water Resources Research,* 34(11), 3181–3189
Sonis, M., & Grossman, D. (1984). Rank-size rule for rural settlements. *Socio-Economic Planning Sciences,* 18, 373–380
Weyl, H. (1952). *Symmetry*. (Princeton, NJ: Princeton University Press (renewed 1980))
Williams, G.P. (1997). Chaos theory tamed. (Washington, D.C.: Joseph Henry Press)
Wong, D. & Fotheringham, A.S. (1990). Urban systems as examples of bounded chaos: exploring the relationship between fractal dimension, rank-size, and rural to urban migration. *Geografiska Annaler B*, 72, 89–99
Ye, D.N., He, W., Xu, W.D. & Li, Z. (2001). Symmetry distribution of cities in China. *Science in China* (*Series D*), 44(8), 716–725
Zhou, Y.X. (1995). *Urban geography*. (Beijing: Shangwu Publishing House) (in Chinese)
Zhou, Y.X. & Yu, H.B. (2004a). Reconstruction city population size hierarchy of China based on the fifth population census (I). *Urban Planning,* 28(6), 49–55 (in Chinese)
Zhou, Y.X. & Yu, H.B. (2004b). Reconstruction city population size hierarchy of China based on the fifth population census (II). *Urban Planning,* 28(8), 38–42 (in Chinese)
Zipf, G.K. (1949). *Human behavior and the principle of least effort* (pp. 417–441). (Reading, MA: Addison-Wesley)

# Part IV
# Simulating and Modeling Urban Transportation Systems

# Chapter 7
# The Dilemma of On-Street Parking Policy: Exploring Cruising for Parking Using an Agent-Based Model

**Karel Martens, Itzhak Benenson and Nadav Levy**

**Abstract** Virtually all major cities around the world face severe parking problems in their centers. While existing models of parking search and choice behavior do provide insight into the basic dynamics of parking in cities, as well as into the phenomenon of drivers cruising for on-street parking, virtually all models discussed in the literature ignore a number of key factors that influence parking behavior and parking dynamics. This paper makes a first step in this direction, by proposing a non-spatial model of parking search and an explicit geosimulation model of the parking process, termed PARKAGENT, which accounts for street network, drivers' decisions and their destination. We employ both models to analyze the phenomena of cruising for parking and compare the models' outcomes, focusing on the impact of space on parking dynamics. We estimate the main characteristics of these dynamics, and specify the conditions under which spatial effects are, or are not, important for analyzing parking. In particular, we demonstrate that traffic engineers' recommendation that about 15% of all on-street parking places should remain vacant to ensure easy ingress and egress and prevent cruising for parking can be decreased to 7% and even less, especially in case of relatively low parking turnover levels. The paper ends with a short discussion, in which we explore the implications of our model study for establishing urban parking policies.

**Keywords** Agent-based modeling · Transportation modeling · Parking modeling · Parking search · Cruising for parking

## 7.1 Introduction

Virtually all major cities around the world face severe parking problems in their centers. In most cases, on-street parking is used to capacity throughout the day, and often even exceeds capacity as a result of illegal parking and double parking

---

I. Benenson (✉)
Geosimulation and Spatial Analysis Lab, Department of Geography and Human Environment, Tel-Aviv University, Ramat Aviv, Tel Aviv, Israel, 69978
e-mail: bennya@post.tau.ac.il

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_7,

(Arnott 2006). Different user groups, such as commuters, often occupy a large share of available parking space, to the detriment of other groups like visitors or residents. Moreover, the tension between demand and supply, in combination with cheap on-street parking, often results in an underutilization of available off-street parking facilities during large parts of the day and to high levels of cruising for cheap on-street parking (Shoup 2004; Shoup 2006).

While existing models of parking search and choice behavior do provide insight into the basic dynamics of parking in cities, as well as in the cruising phenomenon, virtually all models discussed in the literature ignore a number of key factors that influence parking behavior and parking dynamics. Arnott (2006, p. 469) stresses that the new generation of parking models should be able to deal with "the complications caused by heterogeneity among travelers (especially with respect to the value of time, parking duration, and distance traveled), the inhomogeneity of downtown space, congestion interaction between cars and mass transit, parking by downtown residents and delivery vans, through traffic, and the endogeneity of parking duration".

This paper makes a first step in this direction, by comparing the aggregate and non-spatial view of parking search with an explicit geosimulation view (Benenson and Torrens 2004) of the parking process. The geosimulation model accounts for three major system components: two representing urban infrastructure (street network and destinations) and one representing urban agents (drivers searching for parking). We focus on the impact of space on parking dynamics, estimate the main characteristics of these dynamics, and specify the conditions under which spatial effects are or are not important for analyzing parking. Geosimulation of the parking process is done with an agent-based model of parking behavior (Benenson et al. 2008; Benenson and Martens 2008; Martens and Benenson 2008). The impact of space is explored for the case of so-called commercial parking: short-term (an hour or so) parking in a city center by customers and business relations close to services and shops.

The paper is organized as follows. First, we provide a brief overview of existing approaches to modeling parking in the city and assess to what extent these approaches are able to deal with (i) the inherently spatial nature of the parking process, and (ii) driver's reaction to the changing local situation during parking search (Section 7.2). We then briefly present PARKAGENT, an agent-based model to simulate parking search and choice in the city (Section 7.3). In Section 7.4, we present a non-spatial (point) model of parking. In the point model, space plays no role at all in that drivers do not have to search for parking, but can directly occupy a free parking place, if it exists. This model serves as the base line against which we compare the PARKAGENT results. In Sections 7.5, 7.6 and 7.7, we analyze searching and cruising for commercial parking in an abstract rectangular two-way street network, using PARKAGENT. We demonstrate the effects of space that cannot be recognized in an aggregate model and show that these effects diminish in case the arrival rate is more than $\sim$1.1 times the egress rate. The paper ends with a short discussion, in which we explore the implications of the obtained results for establishing urban parking policies.

## 7.2 Parking Models

Various types of models have been developed to simulate and analyze drivers' parking behavior in urban settings. An elaborate review can be found in Young et al. (1991) and Young (2000). For our purposes, the models of parking can be distinguished in terms of their level of aggregation, as well as in their theoretical or applied orientation. In terms of aggregation, a distinction can be made between models that consider groups of drivers and those which explicitly consider individual drivers, and between those considering space implicitly in the stage of model formulation only and those explicitly simulating drivers' movements in space.

One side of the parking modeling spectrum represents spatially implicit and aggregate models. *Dynamic* models of this kind are mostly associated with the economic view of the parking processes (e.g. Arnott and Rowse 1999; Arnott 2006; Shoup 2006; Verhoef et al. 1995). These models provide deep insights into the persistence of urban "parking pattern" dynamics and are the necessary "litmus tests" for parking management in the real city.

The most important input of economic models lies in the systematic analysis of the interrelationship between parking conditions and parking policy. This results in the identification of sets of conditions and policies that optimize parking utilization based on peak hour traffic flows, departure time, modal split and so on (Arnott 2006; Anderson and de Palma 2004; Calthrop and Proost 2006; Zhang et al. 2007; D'Acierno et al. 2006; Petiot 2004; Wang et al. 2004). Necessary for the analytical investigation, the regular economic assumptions of perfectly rational and utility maximizing behavior limits extrapolation of the models' conclusions towards real-world situations. E.g. Shoup's (2006) model does not include space as it eliminates walking distance to the destination. Hence, Shoup can conclude that if prices of on-street and off-street parking are the same, the equilibrium cruising time is zero. However, if off-street parking is relatively sparsely scattered, and destinations are scattered over space, the decision to cruise depends highly on the walking distance between the closest off-street parking facility (assuming it is always available) and the destination.

The other side of the modeling spectrum – that of spatially explicit simulations of drivers' parking search and choice – has started in the 1990s and is still in its infancy. The models we are aware of deal with intentionally restricted situations of search and choice, e.g. parking search within an off-street parking lot (Harris and Dessouky 1997) or several adjacent street segments (Saltzman 1997). These explicit simulations consider parking behavior of drivers as a set of sequential events, in which drivers respond to the actual traffic situation. In principle, these dynamic models are capable of capturing the self-organization of the cruising phenomenon with a changing balance between parking demand and supply (see Shoup 2006 for a spectacular presentation of the problem), but the latter demands a substantial extension of the spatial dimensions of the models and an essential generalization of driver's behavioral rules.

The only attempt in this direction we are aware of is presented in a paper by Thompson and Richardson (1998). They consider driver's parking search and choice between on-street and off-street alternatives within a small (twenty street segments of about 50 m length), but realistic, grid network of two-way streets. The model is developed to follow one driver searching for parking within a fixed parking environment (i.e. no other drivers park or egress during this search). Nonetheless, the paper clearly demonstrates that optimal parking search behavior is hardly possible. Namely, the information available to the driver during parking search and choice is local in nature and long term experience does not necessarily lead to better choices. The outcome of parking processes in a spatially explicit model may thus be removed quite far from those obtained in economic models growing from perfectly rational drivers.

The above brief review leads to the conclusion that, while providing deep insight, most of the existing models cannot be used to formally assess the spatial effects of parking search. The models that potentially do provide these opportunities – spatially explicit and disaggregate models – are still underdeveloped and none of them can be employed to systematically assess the real-world situation of many drivers simultaneously searching for on-street and off-street parking, and simultaneously entering and leaving parking places in a realistic urban environment. This paper aims to start filling the gap by employing a recently developed, agent-based, GIS-based, model of parking in the city.

## 7.3 The PARKAGENT Model

PARKAGENT is a spatially explicit, agent-based model of parking search and choice in the city. The model links a geosimulation approach to a full-fledged GIS database, which are in use for an increasing number of cities around the world. In this way, PARKAGENT enables a representation of driver's parking behavior in a real-life city and in-depth analysis of the global consequences of driver's inherently local view of the parking situation.

### *7.3.1 Infrastructure GIS*

Four components of the model GIS are either directly obtained from, or constructed on, the available infrastructure GIS of a city (Fig. 7.1).

*Road network* contains information on traffic directions, turn permissions, parking permissions, fees, and probability and size of parking fines, for each street segment.

*Destinations* are associated with the features of two layers: Buildings and Open Spaces. The features of these layers can simultaneously have several uses, e.g. a building can be used for dwelling and for offices. Each use is characterized by capacity, which reflects the number of drivers that can use this feature as a


(a) (b)

**Fig. 7.1** View of the PARKAGENT map window for (**a**) a real city; and (**b**) the abstract grid network. Blue points represent road cells, small single black points represent on-street parking places and larger attached black points represent parked cars

destination. For example, a building's dwelling capacity of ten and workplace capacity of three means that ten residents can choose it as a destination when driving home after a working day, while three workers can choose it as a destination when driving in the morning to the workplace. Destinations' attractiveness for different groups of drivers is estimated based on the number of apartments in a building, or type and size of the enterprise, public place, or open space (small, medium, large). Open space destinations (parks and gardens) are also characterized by their capacity in respect to the number of drivers that can choose it as a destination.

*Off-Street Parking Places* are associated with houses and parking lots. The number of off-street residential private parking places is an attribute of the building. Public parking lots are organized as a separate GIS layer and are characterized by capacity and price.

*Road cells and on-street parking cells* are employed for driving and parking, respectively. Road cells are constructed by dividing the street segment centerline into fragments with the length of an average parking place (in Tel Aviv, according to the field survey, the average length of a parking place is 4 m) and employed for representing driving. One (for one-way street) or two (for two-way street) "parking cells" are set parallel to the road at a given distance of the centerline (Fig. 7.1), depending on the physical possibility to park at one or both sides of the street. The layers of the road cells and on-street parking cells are built by PARKAGENT based on the layer of streets; the attributes of the road cells and on-street parking places are transferred from the street layer. These are traffic direction, turn restrictions, parking permission (including "parking not allowed"), fees, and the probability of a fine for illegal parking per hour.

PARKAGENT is a generic model and can be applied to any real (Fig. 7.1a) or imaginary (Fig. 7.1b) city given that the layers above are available. It also contains tools for constructing artificial street networks, as has been done for the purpose of this paper and shown in Fig. 7.1b.

### 7.3.2 Driver Agents and Their Behavior

PARKAGENT is an agent-based model. This means that every driver in the system is assigned a specific origin, destination and form of behavior. A full description of drivers' behavior should include: (1) driving towards the destination at a large distance from the destination (before searching for parking actually commences); (2) driving in proximity to the destination, while searching for parking; (3) parking choice; (4) parking; and (5) driving out. PARKAGENT simplifies the first stage of driver's behavior and focuses on the other ones.

The initiation of a driver in the model begins with assigning the destination and desired parking duration. Then a network service area of 300 m radius is constructed for the driver's destination, consisting of a set of road cells at the boundary of this area, from which the destination can be reached after a 300 m drive. The driver enters the model by "landing" randomly at one of these cells, that is, we ignore driving towards the parking search area. Then the car drives towards the destination while searching for parking and, if succeeding to find a parking place, parks for the time interval assigned during initiation. We erase the driver from the system directly after the parking duration is completed.

Based on Carrese et al. (2004), and our own observations while driving with the drivers and recording their activities, we assume that the driving speed during the parking search stage is 12 km/h, no matter what the speed was before. At each road junction, a driver chooses the street (from those permitted) whose next junction is closest to the destination (Fig. 7.2). Following this rule, the model driver usually takes the shortest path from the road cell they "land" on to her destination.

On-street parking in Tel-Aviv is substantially cheaper than off-street and in this paper we assume that every driver tries to find an on-street parking place before parking in a for-pay off-street parking facility.

### 7.3.3 Driver's Decision to Park on the Way to Destination

Driving towards the destination from the place of "landing", a driver registers free and occupied parking places at both sides of the road. Then, depending on her estimate of the distance to the destination, the driver estimates the expected number, F, of free parking places *on the remaining route* to the destination. Based on F, when passing a free parking place, the driver decides whether to park or continue driving towards the destination. The decision depends on the value of F as follows (Fig. 7.3): The driver

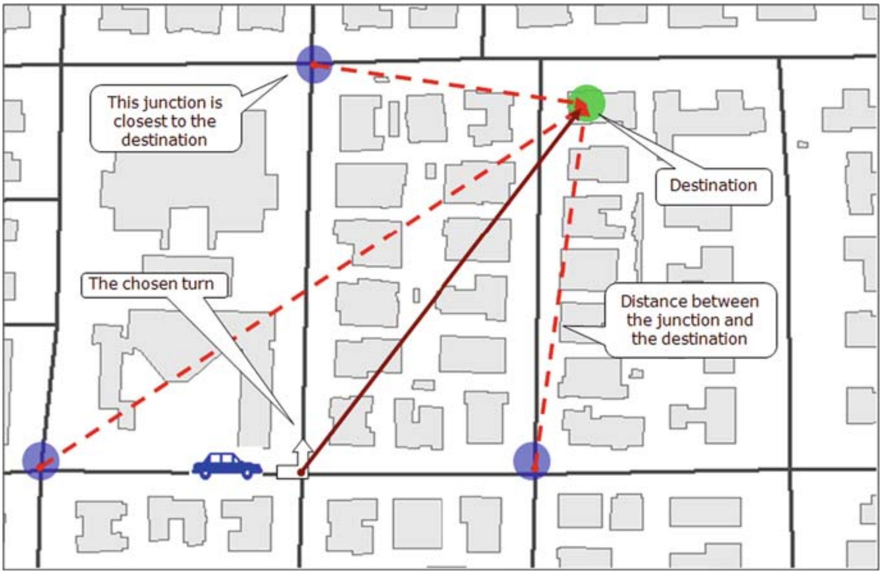



**Fig. 7.2** Schematic representation of the driver's choice of the street segment (Benenson et al. 2008)

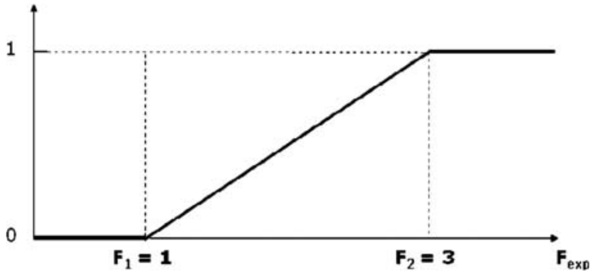



**Fig. 7.3** The probability to continue driving when passing a free parking place as a function of the expected number of free parking places between driver's location and destination (Benenson et al. 2008)

- Continues driving towards the destination if $F > F_2 = 3$,
- Parks immediately if $F < F_1 = 1$,
- Continues driving with the probability $p = (F - F_1)/(F_2 - F_1)$ if $F_1 \leq F \leq F_2$ .

Besides instantaneous re-estimating of F, the driver "remembers" several of the latest street links she has passed during the parking search (in the current version of the model, the driver remembers two recent links). The driver tries to avoid using these links when arriving to a junction and deciding which street to turn to; if impossible, she prefers the link visited least recently. We admit that the knowledge of the local road network and the parking experience in the area can differ between drivers, but we do not account for drivers' long-term memory in the current version of the model.

When a driver parks *on the way to the destination*, we consider cruising time to be zero ($T_{\text{cruising}} = 0$). Obviously, in a real-world setting, drivers slowing down to search for parking *before* reaching their destination may also be perceived to be cruising for parking. Certainly when parking reaches saturation level, drivers may actually start searching for parking at a substantial distance from their destination, slowing down traffic and causing unsafe traffic situations. However, in the current analysis, we do not include this part of the parking search under the cruising definition.

### *7.3.4 Cruising for Parking*

The model driver who *has passed her destination without finding a parking place* is considered to be cruising for parking. During the cruising stage, the driver changes the decision rule shown in Fig. 7.3, and is ready to park anywhere as long as it is not *too far from the destination*. We assume that after passing the destination the driver is ready to park at an aerial distance of 100 m or less from it. During cruising, this maximal parking distance grows at a constant rate of 30 m/min until reaching 400 m, when the expansion of the area stops. The turn rule (Fig. 7.2) at this stage is updated in order to account for the growth of the appropriate parking area: at each junction, only turns for which the next junction is within the (growing) parking area are considered.

We also assume that each driver has a maximal cruising time, after which she will park her car at an off-street parking facility against a fee, and, in this paper, we set this time equal to 10 min for all drivers.

We characterize parking search and cruising at a system level by the following indicators: the fraction of cars who find a parking place on the way to destination; the distribution of cruising time (the time spent searching for parking after failing to park on the way to destination by the drivers who succeeded to park); the average cruising time for drivers that parked during cruising; the fraction of cars that did not find a parking place within the maximal search time of 10 min; and the fraction of cars that did not find a parking place within 1/4th and 1/2nd of this time, i.e. 2.5 and 5 min.

### *7.3.5 Algorithm of Car Following*

The simulation runs at a time resolution of 1 s. Each time-step, the driving car can advance one or more road cells ahead, pass the junction while deciding on the turn, or occupy a free parking cell. We employ sequential updating and consider all moving cars in a random order, established anew at every time-step. All roads are considered as one-lane and before advancing, the driver checks if a cell ahead is not occupied by another car. If yes, the driver does not advance during the time step. To represent an advance, let us note that at a speed of 12 km/h the car passes 3.33 m during 1 s, and this distance is shorter than the length of a parking place, which is 4 m. To relate between the car speed and the length of the parking place, we assume that the driver advances one road cell with probability $p = 3.33/4 = \sim 0.833$, and stays at the current road cell with probability $1 - p = \sim 0.167$.

### 7.3.6 Technical Characteristics of the Model

The model is implemented as a C#.NET ArcGIS™ application and its performance remains high for several thousands of drivers simultaneously searching for parking. The latter is sufficient for theoretical and practical implementations.

## 7.4 Non-spatial (Point) Model of Parking

The goal of this paper is to explore the spatial effects of parking dynamics. In order to assess these effects, we compare PARKAGENT to a theoretical "point" model of parking. The model, first described in Benenson and Martens (2008), describes a non-spatial situation in which drivers are able to immediately assess the situation of all parking spaces and hence know immediately whether a free parking place is available or not. If the number of free parking places is higher than the number of cars searching for parking, then all of them park instantaneously. If there are more searching cars than free parking places, then those who park are selected randomly and "cruising" for parking occurs among the remaining cars. The easiest way to imitate this situation is to assume that the number of cars entering the system is systematically higher than the number of cars leaving the system (i.e. access/egress ratio > 1). In such a case, after a certain time period, all parking places will be occupied and newly entering cars will have to wait until parked cars leave the system and vacate a parking space. We assume that the ability of a driver to occupy a vacated parking place does not depend on the time the driver is searching for parking. This means that the queue system (see e.g. Cohen 1969) we investigate does not work on a standard first-come first-served basis.

The point model is formalized as follows. Let us assume that cars arrive to a specific area at rate $a(t)$ (cars/$\Delta t$) in order to find a parking place, and that cars already parked in the area leave at an egress rate $e(t)$ (cars/$\Delta t$). Let us also assume that the driver's maximum search time is $n*\Delta t$, and that the driver leaves the area if failing to find a parking place during this time. We ignore the spatial dimensions of the area, that is, the distance between the driver's location and the free parking place, and assume that all parking places are occupied at $t = 0$ and that $a(t) \geq e(t)$.

Let us denote $C(t)$ as the overall number of cars in the system, $N(t, t - k*\Delta t)$ as the number of cars that entered the system at $t - k*\Delta t$ and are still searching for a parking place at $t$. Let $p(t)$ be the fraction of cars that fail to find a parking place between $t$ and $t + \Delta t$, and $F(t)$ be the number of cars that have already searched for a parking place starting from the moment $t - n^*\Delta t$ and thus have to leave the system just after $t$ that is, $F(t) = N(t, t - n^*\Delta t)$ and $p(t) = 1 - e(t)/C(t)$.

The dynamics of $N(t, t - k^*\Delta t)$, and $C(t)$ can be represented by the following system of equations:

$$
\begin{aligned}
&C(t+\Delta t) = C(t) - F(t) + (a(t) - e(t))^* \Delta t \\
&N(t+\Delta t, t) = a(t)^* p(t), \\
&N(t+\Delta t, t-\Delta t) = N(t, t-\Delta t)^* p(t), \\
&N(t+\Delta t, t-2\Delta t) = N(t, t-2^*\Delta t)^* p(t), \\
&\ldots \\
&N(t+\Delta t, t-(n-1)^*\Delta t) = N(t, t-2(n-1)^*\Delta t)^* p(t)
\end{aligned}
\tag{1}
$$

where $F(t) = N(t, t - n^*\Delta t)$, $p(t) = 1 - e(t)/C(t)$.

Below, we will use this point model as the base line against which we will compare the results of PARKAGENT, for cases with a 100% occupancy rate for on-street parking and an arrival rate equal to, or higher than, the egress rate.

## 7.5 Cruising for Commercial Parking

Cruising (or searching) for parking in central city areas is a common phenomenon. Drivers prefer to park close to their destinations and pay as little for parking as possible. Hence, if off-street parking is expensive in comparison to on-street parking, or located far away from the destination, and the supply of on-street parking is insufficient, drivers tend to search for a vacant parking space for a while before deciding to park far from the destination or in a for-pay parking lot or garage.

People with different travel motives may cruise for parking. Typically, three types of cruising drivers are distinguished: commercial parkers, work-related parkers (commuter parkers), and residential parkers. The case of commercial parking, which we are investigating in this paper, refers to parking for travel motives like shopping, leisure, or business. Commercial parking differs in three respects from parking by commuters and local residents. First, commuters and residents are long-term parkers (typically, commuters park for 8 h during daytime and residents for 10–12 h overnight), while commercial parkers tend to park for a short period of time (ranging from a few minutes to a few hours). Second, commuters and residents park close to the same destination on a day-to-day basis and usually at the same time (commuters typically during early morning, residents typically at the end of the working day), while commercial parkers tend to visit a central city area on an irregular basis and during different periods of the day (typically starting later in the morning than commuters and lasting till closing time of facilities). Third, a large share of commuters and residents tend to have dedicated, and often free, parking close to their working place, while commercial parkers are dependent on public parking spaces, either on-street or off-street. It may therefore come as no surprise that the majority of the studies on cruising for parking focuses on commercial parking (see e.g. Anderson and de Palma 2004; Arnott and Inci 2006; Shoup 2006). Note, however, that cruising behavior is not limited to commercial parkers but may also occur among commuters and residents (for the latter, see the discussion in Martens and Benenson 2008).

Here, we explore the case of commercial parking. As mentioned in Section 7.2, cruising for commercial parking has not yet been modeled explicitly for the case in which a driver is aware of the parking availability along her search path only. In order to be able to assess the effects of space, we study in details an abstract situation, which, however, is sufficient to reflect the main phenomenon. In the analysis, we assume that drivers travel to a central city area in order to do shopping, visit a café, or for a business meeting, and want to park for a short period of time, up to an hour. We assume that on-street parking is for free, and drivers try to avoid for-pay off-street parking facilities in the area. Hence, drivers will have a tendency to cruise to find a vacant on-street parking place. We assume that all drivers behave in the same way (i.e. no driver heterogeneity) and are willing to search for a maximum amount of time to find an on-street parking place. In case they fail, drivers refer to an off-street parking facility in the area, which is assumed to be immediately available (as in the case of Shoup 2006). We further assume that the drivers' destinations are spread evenly over a rather large central city area covered by a square two-way street network (Fig. 7.1b). Drivers' parking search and choice behavior is guided by relatively simple rules, as described in Section 7.3. As mentioned above, we compare the outcomes of PARKAGENT to the results obtained in the "point" model.

We analyze parking dynamics for an area with a parking capacity $K$ of 7,000 parking places ($K = 7{,}000$), which is sufficient to ignore boundary effects and corresponds to an area of about 1.5 km$^2$ within the center of Tel Aviv. We run each scenario for 1 or 2 h, depending on the scenario. In case of a 2-h run, the first hour is used to establish a stable model regime and the second for observing the model dynamics. One-hour runs are performed in cases when the period of initial convergence to a stable regime is unnecessary. The results of the model analysis are presented by several parameters:

- $P_0$, which represents the fraction of cars that succeeded to park without cruising, i.e. before reaching the destination, as a share of the total number of cars that arrived into the modeled area during the scenario period;
- $T_{\text{cruising}}$, which represents the average search time for the cars that found a parking place while cruising;
- $P_t$, which represents the share of cars that search for more than $t$ seconds. We present $P_{300}$, and $P_{600}$, i.e. the percentage of cars which search for more than 5 and 10 min, respectively. Note that $P_{600}$ represents the percentage of cars

arriving during the scenario period but failing to find a parking place during the maximum possible search time (termed "parking failure" below).

We present the results of two explorations: an analysis of the so-called cruising threshold and an analysis of the differences between the point and spatial model of parking. Let us note that at least some of the results presented below of the spatial modeling can be obtained theoretically; here we present the simulation results.

## 7.6 Cruising Threshold

Traffic engineers usually recommend that about 15% of all on-street parking places – one space in every seven – should remain vacant to ensure easy ingress and egress (in Shoup 2005, p. 297). In line with this recommendation, Shoup (2006) provides an excellent analysis of the cruising phenomenon and claims that cruising can be eliminated if prices for on-street parking are set in such a way that only 85% of all on-street parking spaces are occupied (see Shoup 2005, Chapter 12 and 13). Let us call the occupancy rate that results in close-to-zero cruising for parking the "cruising threshold" $f_{\text{cruising}}$ and explore what is its value.

To obtain an estimate of $f_{\text{cruising}}$, we assume that the arrival and egress rates of cars in the study area, a and e, are constant in time and equal, i.e. $a(t) = a = e(t) = e$. For convenience, we will consider a and e as per hour rates.

To estimate $f_{\text{cruising}}$ and its possible dependence on the arrival rate $a$, we consider a number of scenarios that differ in terms of (1) the initial occupancy rate, defined as fraction $f$ of the occupied parking places in the area , and (2) the number of cars that enter and leave the area ($a$ and $e$). We assume uniform distribution of car departures (egress) and car arrivals and investigate cases of low, average and high turnover rates: $a = e = 1{,}000$ cars/h, $a = e = 2{,}000$ cars/h, $a = e = 3{,}500$ cars/h and $a = e = 5{,}000$ cars/h, i.e. average turnover rates close to 15, 30, 50, and 70%. We search for the cruising threshold $f_{\text{cruising}}$ by varying initial percentages of occupied parking places $f$, increasing it stepwise from 80 to 100%.

Figures 7.4 and 7.5 show the relationship between the share of cars parking without cruising $P_0$, the average search time $T_{\text{cruising}}$ for cars that find a parking place after cruising, and the percentage of cars searching for parking longer than a given time, for each of the four parking turnover levels defined above.

According to Figs. 7.4 and 7.5, and comparing the results to the initial view of Shoup, the notion of the cruising threshold $f_{\text{cruising}}$ should be further specified, as the results essentially depend on the turnover rate. For a 15% turnover rate, the model outputs essentially depend in a non-linear way on the parking occupancy rate $f$. For $f$ up to 90%, the fraction $P_{600}$ of cars that fail to find a parking place within 10 min remains below 0.5% (Fig. 7.4a), $T_{\text{cruising}}$ remains below 40 s (Fig. 7.5a), and $P_{300} < 1.5\%$ (Fig. 7.4b). For $f > 90\%$ all these figures quickly grow with the increase in $f$, but yet for $f = 95\%$, $T_{\text{cruising}}$ remains below

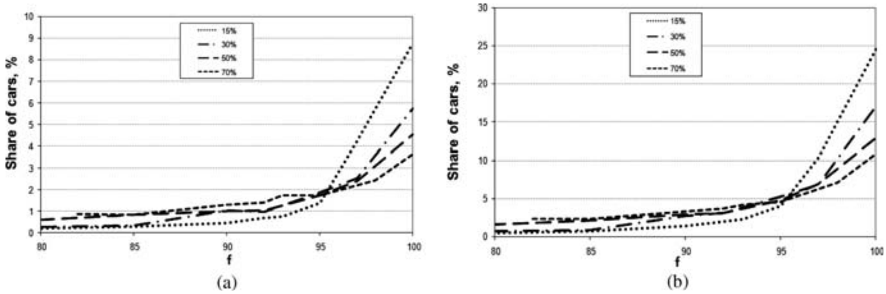



**Fig. 7.4** Share of cars that search for parking longer than (**a**) 600 s ($P_{600}$) and (**b**) 300 s ($P_{300}$), for average turnover rates of 15, 30, 50, and 70%

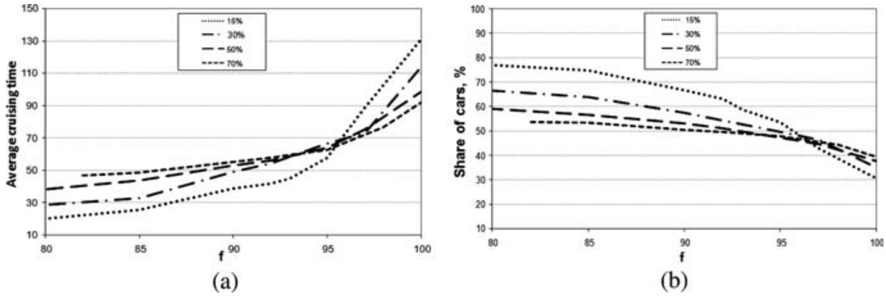



**Fig. 7.5** (**a**) Average cruising time ($T_{cruising}$) in seconds; and (**b**) Share of cars parking without cruising ($P_0$) for average turnover rates of 15, 30, 50, and 70%

1 min and $P_{600}$ below 3%, i.e. one parking failure per hour. That is, for the relatively low turnover rate, $f_{cruising}$ should be set at 90% at least, which is essentially higher than the figure proposed by traffic engineers and strongly advocated by Shoup.

For higher turnover rates, the dependencies of the model outputs on $f$ are very close to linear up to $f \sim 95\%$. To preserve $P_{600} < 0.5\%$, $T_{cruising} < 40$ s and $P_{300} < 1.5\%$, which is characteristic of $f_{cruising}$ for the turnover rate of 15%, $f_{cruising}$ should be set below 80%, substantially *lower* than 85% as advocated by Shoup. Note that for high turnover levels, even for $f_{cruising,}$ some drivers will be searching for a parking place longer than 10 min.

## 7.7 Effects of Space on Cruising

The goal of the second series of model runs is to explore and estimate the effect of space on parking search dynamics. Note that spatial effects become insignificant when large numbers of cars are searching for a free parking place. In that case, a cruising car is highly likely to immediately see and occupy a parking place when it is vacated by another driver, just as it happens in the point model. These conditions become realistic if the instantaneous number of cruising cars (1) is high enough to have a car searching for parking in close proximity to *every* parking place, and (2) is essentially higher than the instantaneous number of free parking places.

To quantitatively estimate the effects of space, we consider two series of runs, both for the situation when all parking places are initially occupied ($f = 100\%$). The first series investigates the boundary case in which arrival/egress rates remain equal. This case results in zero cruising time in a point model and we compare it to the PARKAGENT results for a range of values of $a = e$, starting from 700 cars per hour (i.e. turnover rate $a/K = 10\%$ per hour), both increasing stepwise by 700 till reaching a level of 7,000 (i.e. turnover rate $a/K = 100\%$ per hour).

The second series investigates the case of more arrivals than egresses, $a > e$. Both in the PARKAGENT and point models there are now *many* arriving cars which compete for a limited number of free parking places. Assuming a 100% initial occupancy rate, we investigate the differences between the outcomes of the spatial and point model for $e = 1{,}000$ and $e = 5{,}000$ cars per hour, while $a$ increases stepwise, starting from $e$, while $e$ itself remains constant. The case of $a > e$ demands time to stabilize, and we add a 30 min period of initial stabilization (i.e. a period three times longer than the maximal search time) before recording the model outcomes.

### 7.7.1 Equal Arrival and Egress Rates

Figures 7.6 and 7.7 show the dependence of the percentages of cars searching for parking longer than a specified time, the share of cars parking without cruising ($P_0$), the average search time for cars that find a parking place after cruising ($T_{cruising}$), and the share of cars which are in search more than 150, 300, and 600 s $P_{150}$, $P_{300}$, and $P_{600}$.

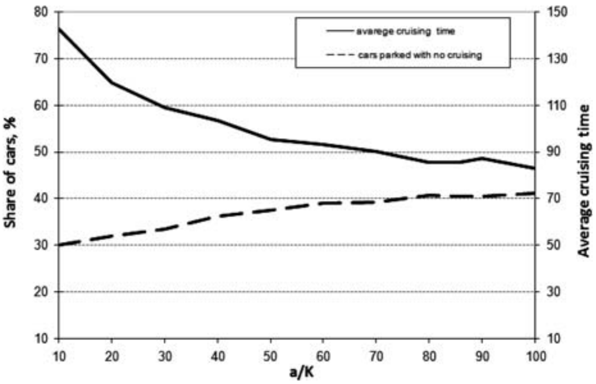



**Fig. 7.6** Share of cars parking without cruising ($P_0$) and average cruising time in seconds ($T_{cruising}$) as dependent on the average turnover rate

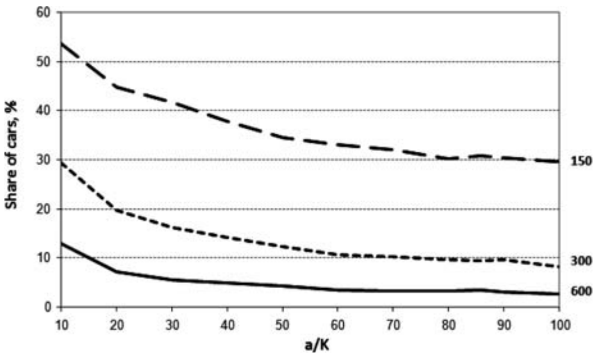



**Fig. 7.7** Share of cars that search for more than 150, 300, and 600 s before finding a parking place ($P_{150}$, $P_{300}$ and $P_{600}$)

The results presented in Figs. 7.6 and 7.7 confirm the necessity of a spatial view when constructing a model of parking and cruising. Remember that in a point model of this scenario, all cars find a parking place immediately after entering the system, i.e. average search time is zero, there are no free parking places, and none of the cars is searching. The spatial model demonstrates that this is never so in the real-world, and stresses the importance of the turnover rate. Low turnover rates of 10–20% result in a less than 35% chance to park on the way to the destination and a more than 2-min average cruising time as a result (Fig. 7.6). The chance of parking failure is close to 10%, while about 25% of all cars search for parking for more than 5 min (Fig. 7.7).

A high turnover rate (90–100%) is better for the driver: the chance to park on the way to the destination increases, average cruising time goes down, and the chance of parking failure decreases to less than 5%. However, even for a 100% turnover rate, the estimates of all the parameters are significantly far from "zeros", as characteristic of the point model.

## 7.7.2 *When the Point Approximation Becomes Sufficient?*

The goal of the third series of model runs is to verify qualitatively the outcomes of the PARKAGENT and point models in case $a > e$, when searching cars accumulate and the conditions of "close to point" dynamics become realistic for the spatial model.

The runs of the first series of this kind start with an initial occupancy rate of 100% and $a = e = 1{,}000$ cars/h (implying a 15% turnover rate). Then we increase $a$ by 500 cars/h, going up to $a = 4{,}000$ cars/h. The runs of the second series starts with $a = e = 5{,}000$ (implying a 70% turnover rate) and $a$ goes up by 500 to $a = 8{,}000$. Note, that under these circumstances the turnover rate is determined by the egress rate and remains 15% in the first series and 70% in the second series of runs. Note also, that in these runs we compare the *search time* obtained in the point model to the entire search time $T_{\text{search}}$ in the PARKAGENT model. The search time elapses from the moment the car enters the model environment (and not, as the cruising time, from the moment the car passes the destination) and includes driving towards the destination.

Figures 7.8 and 7.9 present the standard output of the PARKAGENT model together with the estimates obtained with the point model.

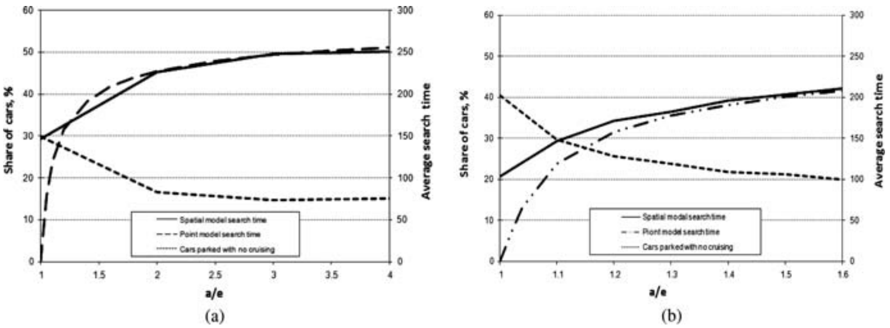



**Fig. 7.8** Share of cars parking without cruising ($P_0$) and average search time ($T_{search}$): (**a**) $e = 1{,}000$ cars/h, $a = 1{,}000 \div 4{,}000$ cars/h, (**b**) $e = 5{,}000$, $a = 5{,}000 \div 8{,}000$ cars/h

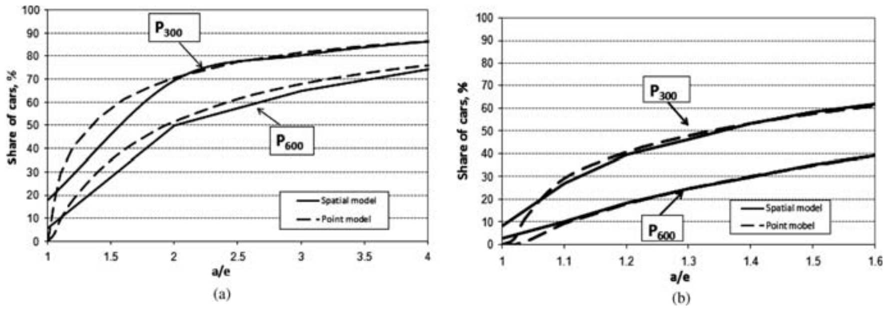



**Fig. 7.9** Share of cars that search for parking longer than 300 and 600 s ($P_{300}$ and $P_{600}$): (**a**) $e = 1{,}000$ cars/h, $a = 1{,}000 \div 4{,}000$ cars/h, (**b**) $e = 5{,}000$, $a = 5{,}000 \div 8{,}000$ cars/h

As should be expected and can be seen in Figs. 7.8 and 7.9, average search time, the number of cruising cars and the chance of failure are higher in PARKAGENT than in the point model. However, the differences between both models become insignificant when the arrival rate exceeds the egress rate by ~10% (i.e. $a/e > 1.10$).

## 7.8 Conclusions

Despite the fact that we are only in the beginning of exploring the cruising for parking phenomenon, several preliminary conclusions can already be drawn.

First, the findings so far show that for low levels of parking turnover (at around 15% per hour), the cruising threshold (i.e. parking occupancy level) can be increased from the value of 85%, as advocated by Shoup, to about 90% or even 95%. However, for high levels of parking turnover (more than 50% per hour) the notion of cruising threshold does not work. Under such circumstances, the average cruising time and the share of long searchers (more than 300 s) increases almost linearly to the increase in the occupancy level. Even for a parking occupancy level of 80%, the majority of drivers have to cruise in order to find a free parking place and some of the drivers will fail to find a parking place within 10 min of search time. This suggests that policy makers, especially those responsible for parking areas with high levels of parking turnover, will be faced with a policy dilemma. Either, they have to be moderate in terms of parking fees and accept high occupancy rates and the related cruising for parking among a substantial share of drivers. Or they will have to set relatively high prices for on-street parking to force parkers to off-street parking facilities, so that the average occupancy rate will fall well below 80% and cruising is largely avoided. The former option will inevitably lead to complaints from drivers about a lack of parking places. The latter option, however, would lead to complaints about overpricing of on-street parking places. After all, in this case, more than 2 out of 10 parking places will be free on average, which makes it practically impossible for policy makers to make a case for high parking fees.

Second, the explorations presented in the paper suggest that spatial effects are unimportant for understanding the cruising phenomenon in a situation when the ratio between arrival and egress rates is above $\sim$1.1. This may occur, for instance, in cases of major events like a soccer game or a rock concert, or, more commonly, in the evening hours in high density cities when residents search for overnight parking (as discussed in Benenson et al. 2008). When such a high (temporary) shortage of on-street parking occurs, many cars are continuously searching for parking in an area, and, consequently, will virtually always occupy a parking place directly after it is being vacated. This happens when the arrival rate is essentially higher than the egress rate and does not depend on the level of parking turnover. In other words, in such circumstances, parking search may be described with a simple mathematical model.

The question remains which of these findings hold true in the real-world, when parking space is heterogeneous (e.g. when parking fees differ between streets or streets differ in terms of parking permissions), destinations are not uniformly distributed over space (e.g. attractions are strongly clustered within a larger area or off-street parking facilities are scarce), and drivers behave heterogeneously (e.g. differ in terms of their willingness to search for on-street parking). It seems likely that an explicit spatial view of parking is necessary in these cases, even under the circumstances of a strong (temporary) imbalance between demand and supply for on-street parking facilities and high parking turnover levels.

The PARKAGENT model is not only a suitable tool to explore these theoretical issues, but can also support decision-makers to set policies that reduce cruising for parking as much as possible.

## References

Anderson, S.P. & de Palma, A. (2004). The economics of pricing parking. *Journal of Urban Economics*, 55(1), 1–20

Arnott, R. (2006). Spatial competition between parking garages and downtown parking policy. *Transport Policy – Special Issue on Parking*, 13(6), 458–469

Arnott, R. & Inci, E. (2006). An integrated model of downtown parking and traffic congestion. *Journal of Urban Economics*, 60(3), 418–442

Arnott, R. & Rowse, J. (1999). Modeling parking. *Journal of Urban Economics*, 45(1), 97–124

Benenson, I. and Martens, K. (2008). From modeling parking search to establishing urban parking policy. *Zeitschrift Künstliche Intelligenz*, 08(3), 8–13

Benenson, I., Martens, K. & Birfir, S. (2008). PARKAGENT: an agent-based model of parking in the city. *Computers, Environment and Urban Systems*, 32(6), 431–439

Benenson, I. & Torrens, P.M. (2004). Geosimulation: automata-based modeling of urban phenomena. (London, Wiley)

Calthrop, E. & Proost, S. (2006). Regulating on-street parking. *Regional Science and Urban Economics*, 36, 29–48

Carrese, S., Negrenti, A. & Belles, B.B. (2004). Simulation of the parking phase for urban traffic emission models. *TRISTAN V – Triennial Symposium on Transportation Analysis*. Guadeloupe

Cohen, J.W. (1969). *The single server queue*. (Amsterdam: North-Holland)

D'Acierno, L., Gallo, M. & Montella, B. (2006). Optimisation models for the urban parking pricing problem. *Transport Policy*, 13(1), 34–48

Harris, J.M. & Dessouky, Y. (1997). A simulation approach for analyzing parking space availability at a major university. *Proceedings of the 1997 Winter Simulation Conference,* Atlanta, GA

Martens, K. & Benenson, I. (2008). Evaluating urban parking policies with agent-based model of driver parking behavior. *Transportation Research Record: Journal of the Transportation Research Board*, 2046, 37–44

Petiot, R. (2004). Parking enforcement and travel demand management. *Transport Policy,* 11(4), 399–411

Saltzman, R.M. (1997). An animated simulation model for analyzing on-street parking issues. *Simulation*, 69(2), 79–90

Shoup, D.C. (2004). The ideal source of local public revenue. *Regional Science and Urban Economics*, 34(6), 753–784

Shoup, D.C. (2005). *The High cost of free parking*. (Chicago: Planners Press)

Shoup, D.C. (2006). Cruising for parking. *Transport Policy – Special Issue on Parking,* 13(6), 479–486

Thompson, R.G. & Richardson, A.J. (1998). A parking search model. *Transportation Research Part A: Policy and Practice*, 32(3), 159–170

Verhoef, E.T., Nijkamp, P. & Rietveld, P. (1995). The economics of regulatory parking policies: the (im)possibilities of parking policies in traffic regulation. *Transportation Research Part A: General*, 29(2), 141–156

Wang, J.Y.T., Yang, H. & Lindsey, R. (2004). Locating and pricing park-and-ride facilities in a linear monocentric city with deterministic mode choice. *Transportation Research Part B: Methodological*, 38(8), 709–731

Young, W. (2000). Modeling parking. In D.A. Hensher & K.J. Button (Eds.), *Handbook of transport modeling*, (pp. 409–420). (Amsterdam: Elsevier)

Young, W., Thompson, R.G. & Taylor, M.A.P. (1991). A review of urban car parking models. *Transport Reviews*, 11(1), 63–84

Zhang, X., Huang, H.-J. & Zhang, H.M. (2007). Integrated daily commuting patterns and optimal road tolls and parking fees in a linear city. *Transportation Research Part B: Methodological*, 42(1), 38–56

# Chapter 8
# Multiscale Modeling of Virtual Urban Environments and Associated Populations

**Walid Chaker, Bernard Moulin, and Marius Thériault**

**Abstract** This chapter introduces a novel approach to model Virtual Urban Environments (VUE) by taking into account: (1) the multiscale issue, (2) activities' locations, multi-modal transport networks and their relationships, and (3) the associated synthetic population. This new modeling approach is based on a scale-free design pattern, which introduces the notion of *Service* as a way of representing locations' accesses as well as modal shift opportunities using the transportation network. An example of creating such a VUE is given for Quebec-city's case study. Simulating daily travel activities of an entire population can be performed with our TransNetSim software at the meso-scale. In order to forecast the travel demand we use a weighted bipartite matching algorithm that assigns activities' locations to agents representing people, according to their predicted travel times. Using a precompiled routing matrix, TransNetSim is able to simulate 1 million trips on the Quebec-city's network (around 32,000 nodes and 81,000 links) in less than a minute. TransNetSim is also coupled with the commercial traffic microsimulator Vissim, which provides an environment that is compatible with our micro-VUE model. This work is a first step toward a long-term objective of modeling VUEs as a framework in which many multiscale urban phenomena might be integrated.

**Keywords** Multiscale modeling · Virtual urban environment · Synthetic population · Design pattern · Traffic assignment · Traffic simulation · Shortest path

W. Chaker (✉)
Department of Computer Science and Software Engineering, Laval University,
Quebec G1V 0A6, Canada; Center de Recherche en Aménagement et Dévelopment,
Laval University, Canada
e-mail: walid.chaker@ift.ulaval.ca

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_8,

## 8.1 Introduction

"Virtual Urban Environment" (VUE) was originally a term used by the Virtual Reality (VR) and Geographic Information System (GIS) communities to mean a limited built environment in which one can simulate the movement of pedestrians and vehicles (Donikian 1997; Jiang 1999). The dramatic improvements of computing facilities, as well as the availability of a variety of geospatial data (particularly for road networks and buildings) that occurred during the last decade, makes it now possible to create VUEs representing large spaces such as whole cities. Consequently, a wide range of applications can take advantage of VUEs, including the dynamics of urban systems (Albeverio et al. 2008) and transportation, which is particularly targeted in our investigation. A VUE is not only a virtual built environment useful to simulate pedestrians' and vehicles' behaviours, but it can also be thought of as a framework in which many urban phenomena can be reproduced. In this spirit, Michael Batty suggested an integration approach based on a generic reaction/diffusion behavioural model and explained how it may be applied to three kinds of simulations at very different spatial scales: "pedestrian movement at the building scale, the evolution of systems of cities at a regional scale, and urban growth at the city scale" (Batty 2005).

However, in order to create such large VUEs we need to address new modeling requirements. First, since it is well known that urban phenomena are scale-dependent, we suggest that a multiscale representation of the VUE is most desirable. In addition, we need to model multi-modal transportation networks as well as places (or locations) because they are basic elements of urban forms. Consequently, modal shift opportunities and places' accessibilities also become important features to include in a VUE. Finally, since a VUE may represent a large city, simulating travel activities requires the collection and processing of a large quantity of data about the population living in it. We suggest that modeling VUEs should jointly be performed with modeling its associated synthetic population. Let us give more details about these new requirements for VUEs:

- *Requirement 1: The multiscale issue should be considered when modeling the VUE.*

Even if detailed geospatial data are increasingly accessible, we still need abstract representations in order to deal with existing behavioural modeling and spatial reasoning techniques such as path finding, navigation and perception processes. In addition, a VUE is intended to be a generic framework that can support different simulations while spatial scales may vary from one simulation to another. In traffic simulations, for example, a driver's model and a flow-based model require different representations (or abstractions) of the road network. GIS researchers have already investigated the multiscale issue, commonly using cartographic generalisation approaches (Mackaness et al. 2007). Their purpose is to process GIS data to create maps at different scales of visualisation. Modeling a VUE at different scales is a new problem since we need to make the multiscale representation of the VUE available for simulations, and not only for map visualization.

- *Requirement 2: A VUE should include multi-modal transport networks, locations and their relationships.*

Multi-modality and accessibility of activity locations are two important features that need to be considered when modeling a VUE dedicated to applications like transport simulations. Some GIS modeling approaches in the transportation domain known as GIS-T (Miller and Shaw 2001) have investigated a subset of this requirement, but there is still a need to link simultaneously different modal transportation networks together and each transportation network to its accessible locations through service points.

- *Requirement 3: Population that will be simulated in the VUE should be modeled and associated with the VUE.*

As soon as VUEs could cover large spaces such as cities, and could support the simulation of people's travel living in these cities, there is a need to model synthetic populations in order to be able to plausibly define the inputs of these simulations. When a VUE covers a limited space, agents (i.e. the simulated entities such as vehicles and pedestrians) can be generated inside the simulation (Moulin et al. 2003). Otherwise, it is more promising to follow a bottom-up process by modeling one by one individuals and households where their spatial and non-spatial attributes are estimated by analysing sample data, which characterize the real population and are provided by censuses or Origin Destination surveys.

For these reasons we propose in this chapter a multiscale modeling approach of VUEs (including multi-modal transportation networks and activity locations) combined with a synthetic population creation approach. This chapter is structured as follows. In Section 8.2, we introduce the VUE design pattern, which provides the foundations for our modeling approach in accordance to requirements 1 and 2. Then, we report our experiment using the creation of Quebec-city's VUE (Section 8.3) and of its synthetic population (Section 8.4) in accordance with requirement 3. In Section 8.5 we present our meso-simulator, TransNetSim, and show how it is able to plausibly simulate daily work and school travel activities for the synthetic population.

## 8.2 Our VUE Design Pattern

A design pattern in the software engineering domain as initially introduced by architect Christopher Alexander (1977), is an initial description or a template specifying how to solve a problem. A design pattern, once defined, may be used in many different situations or at different scales, as in our case. As a starting

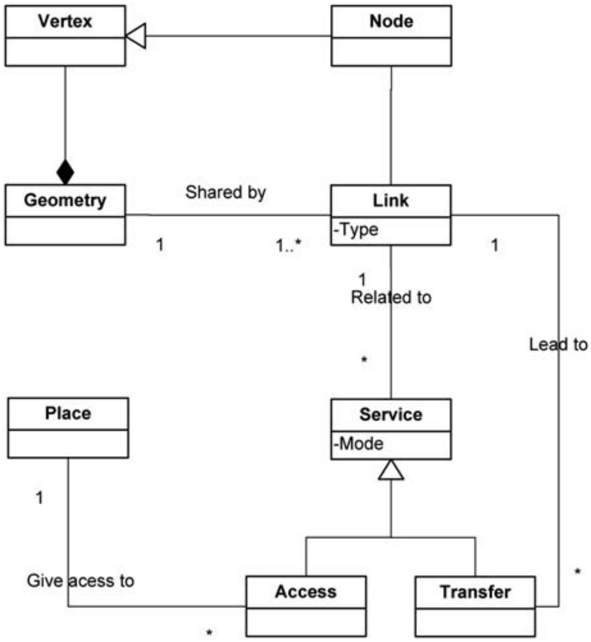



**Fig. 8.1** Our VUE design pattern (UML class diagram notation)

point for the creation of a multiscale VUE, we similarly propose to use a VUE design pattern. This pattern can be considered as a meta-model to be used to instantiate a VUE at any scale. Whatever the scale, the VUE is mainly characterized by two spatial components: *Locations* and the *Transport Network*. As the VUE groups concepts related to *Locations* and the *Transport Network*, it must also be used to model their relationships. The VUE design pattern of Fig. 8.1 is the result of our investigation and it presents a set of abstract classes and their relationships. Next subsections comment these abstract classes according to (1) *Locations*; (2) *Transport Network* and (3) *Locations/Transport Network relationships* as the three main components that characterize our VUE.

## *8.2.1 Locations*

We define a *Place* (a location such as a parcel, a block, a building, or a room) as a generic concept that may have different semantics depending on the scale. In our model, a *Place* does not have a spatially explicit definition. This may seem unusual compared to traditional GIS approaches where visualisation and spatial data handling are fundamental elements. The simplification that we propose in this pattern provides more flexibility when dealing with uncertainty and scale variations without loosing consistency, because reaching a place using our definition, amounts to reaching one of its accesses. This notion of *Access* is spatially defined and it is part of the Location / Transport Network relationship (Section 8.2.3).

### *8.2.2 Transport Network*

In the literature, a transport network is usually represented by a graph made up of nodes and links (also called edges). But nodes and links are not interpreted in the same way at different scales. A link may represent an entire road or a section of a road, or even a single lane that is part of a road section. Therefore, we define *Links* and *Nodes* (abstract classes *Link* and *Node* in Fig. 8.1) as basic elements of the transport network in the VUE design pattern. It is important to notice that even if the design pattern indicates that the network is always structured as a graph, it does not define precisely the relationship between nodes and links (no cardinality characterizes the *Link/Node* association). In Section 8.3, we will show how the network is more precisely structured depending on the considered scale.

Unlike *Places*, *Links* and *Nodes* are spatially defined in our model. Each link has a polyline geometry (*Geometry* class) which is a set of vertices (aggregation of *Vertex Class*) and each *Node* inherits spatial coordinates from a Vertex. A given geometry component can be shared by several links (for example, a road and its adjacent sidewalk share the same geometry in a large scale representation).

### *8.2.3 Places/Transport Network Relationships*

The notion of *Service* (*Service* abstract class in Fig. 8.1) conceptualizes the relationships between *Places* and the *Transport Network*. A service is also considered as a generic concept playing two main roles: (1) to give *Places* access to the transport network (thanks to the *Access* class) and (2) to link several transport networks together in order to form a multimodal transport network (thanks to the *Transfer* subclass).

Being considered as a *Service* (inheritance relation between *Access* and *Service* in the design pattern) an access is always positioned on a link (modeled with the *Related to* association between *Service* and *Link*). We also emphasize that the distinction between the *Link Type* (attribute *Type* in the *Link* abstract class) and *Service Mode* (attribute *Mode* in the *Service* abstract class) arises, for example, from the distinction between the road system and transport modes using this road system (travel by car, by bus, etc). A *service* is always related to one and only one *Link* and hence the attribute *Mode* in the *Service* class specifies the transport mode used to carry out this service. A bus station can then be specified as a *Transfer* in a given model. In this case, the *Transfer* is associated with the mode “bus”. It is related to one *Link* of type “road” and it leads to one link of type "sidewalk". If a *Service* is an access to a *Place* such as a garage adjacent to a house, then the *Access* mode is "car". It relates one *Link* of type "road" and it gives access to the *Place* "house".

Since a service is positioned on a specific position on a link, our model offers more flexibility compared to traditional approaches where the origin and destination of trips are specified as nodes of the transportation network. Dealing with graph-based algorithms is certainly more obvious when origins and destinations are nodes but it does not faithfully reflect reality, especially when the multiscale issue is taken into account.

### 8.2.4 Illustration

Starting from the previous VUE design pattern (Fig. 8.1), let us show how some concepts may have different representations. Let us consider a train station, a bus stop, and a parking lot. Figure 8.2 shows three different modeling alternatives for each of them.

Since in alternative 1, the train station is represented by a *Place* with two *Accesses*, one from a road *Link* and another from a train *Link*, alternatives 2 and 3 model it as *Transfer(s)*. Alternative 2 defines a single *Transfer* between a road *Link* and a train *Link*, while alternative 3, which is the most detailed one, considers the train station as a set of *Transfers*: (1) between the road *Link* and the Walk *Link* (parking lots) (2) between the walk *Link* and the train *Link*. The last-mentioned *Transfer* represents the platform of the train station and it originates from a segment (or a section) within the walk *Link* contrarily to other cases, this is why it is graphically different from the other *Transfers* in Fig. 8.2.

Considering the bus stop case in Fig. 8.2, alternatives 1 and 2 are similar to the first example. We put two *Accesses* from the road *Link* to the bus stop *Place* in alternative 1 because one *Access* is of *Mode* "Bus" and another one is of *Mode* "Car". This solution considers that the bus stop is adjacent to a park-and-ride commuter lot. Alternative 3 models the bus stop as a set of *Accesses* to all *Places* reachable from this bus stop. Finally, for the parking example, alternative 1 considers it as an *Access* to a *Place* from a road *Link* (this is mostly the case of an indoor parking lot) while alternatives 2 and 3 model it as *Transfer(s)*. When the scale is small enough to be able to distinguish separate parking slots, we put one *Transfer* for each parking slot (alternative 3), otherwise one *Transfer* can have a capacity attribute and then correspond to an entire parking lot (alternative 2).

To sum up, examples of Fig. 8.2 emphasize the expressivity of our VUE design pattern as well as its capability to model many spatial scenes related to transportation activities. Our concept of *Service* plays a central role in this model. It plays a similar role as a *Verb* which organizes a sentence: it indicates a state, a change, an action. In our case, a *Service* organizes a portion of a spatial scene because it indicates how elements are connected; how they form a navigable space and how to use or change transportation mode within this area.

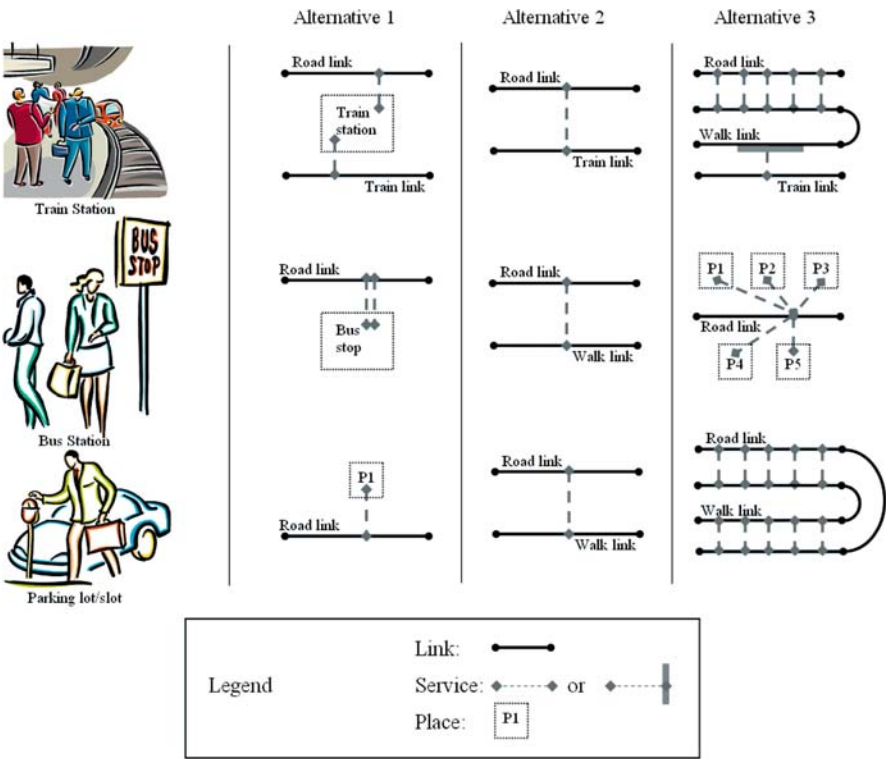



**Fig. 8.2** Three examples showing different alternatives to model the same concept using the VUE design pattern: a train station (*at the top*), bus stop and a parking lot (*at the bottom*). NOTE: While only *Links* and *Nodes* are spatially defined, we normally can not draw *Places* and *Services,* which are abstract concepts as mentioned before. Therefore, what we show in in *dotted lines* are fictive lines for illustrative purposes only

Recent works on road-based topological analysis consider *Connectivity Graphs* (Jiang and Liu 2007); also called dual modeling approach for the street network (Crucitti et al. 2006) or an information city network (Rosvall et al. 2005). It is interesting to remark that the nodes of a *connectivity graph* (where vertices represent named streets and edges represent intersections between streets) can simply be considered as *Services* in our model.

## 8.3 Example of Creating a Multiscale VUE: The Case of Quebec–City

Here is an outline of the method that we followed to create Quebec-city's multiscale VUE: Starting from our design pattern (Fig. 8.1), we instantiated one conceptual model for each desired scale. We chose to start with three different scales called *Macro*, *Meso* and *Micro* scales. Then, we used appropriate GIS data to feed each conceptual model, in order to create distinct single-scale environments. In addition, a synthetic population has been generated and linked to the meso-VUE by estimating people's activity locations as we will see in Section 8.4. Let us detail here the modeling and creation stages of a multiscale VUE. In order to define *Macro*, *Meso,* and *Micro* scales, we need to provide an interpretation of a *Place* for each scale. We noticed that if places are specified relative to the scale, the definition of all the remaining concepts (especially links) follow systematically. Table 8.1 shows the interpretation of places and links chosen for our *Macro*, *Meso*, and *Micro* scales.

Once *Places* and *Links* are defined at each desired scale, it is easy to apply our design pattern (Fig. 8.1) in order to obtain a conceptual model of the environment at each scale. Figure 8.3 shows the application of the design pattern in relation to the scale specification given in Table 8.1, as well as GIS cartographic data that we used to generate VUEs. Only changes of the generic design pattern are discussed here. In Fig. 8.3, only the scale-varying concepts are represented in each VUE model. Please, look at the design pattern of Fig. 8.1 for a more complete view. The *Transfer*, *Access*, and *Place* classes of the design pattern remain exactly the same as they were introduced in Fig. 8.1.

**Table 8.1** Scale specification

| | Place | Link |
|---|---|---|
| Macro | Region: An entire city or at least a district | Road section |
| Meso | Parcel or block | One way road section |
| Micro | Use area (Building, parking lot, park, ...) | Lane |

At the macro-scale, a road section is a bidirectional *macro-link* which can represent a section of a highway, a railroad, or a seaway (as the Saint-Lawrence River in our case) that links several cities. One *macro-node* is associated with several *macro-links* that materialize the crossroad branches. A *Service* is located on a *macro-link*. This is why the "related to" relationship has an attribute "position", which is a vertex of a *link*. The scope of this macro-VUE comprises entire multi-county metropolitan regions. As shown in Fig. 8.3, we cover in our case the metropolitan region of Quebec-city and its surrounding areas. Any commercial regional map is suited to feed such a macro-VUE.

Since a *Link* at the *meso-scale* represents a road segment (Table 8.1), it is more appropriate to consider these *meso-links* as one-way links. Thus, a highway is represented by a series of two *meso-link*s. Each *meso-link* is associated with one starting *meso-node* and one ending *meso-node*. *Services* are still located on *meso-links,* as in the macro-model. Maps where we can distinguish buildings (at least blocks) are suited to generate a meso-VUE. A complete road network is required but details such as numbers of lanes and traffic lights are not necessary since *meso-links* and *meso-nodes* are usually associated with traffic impedance values.

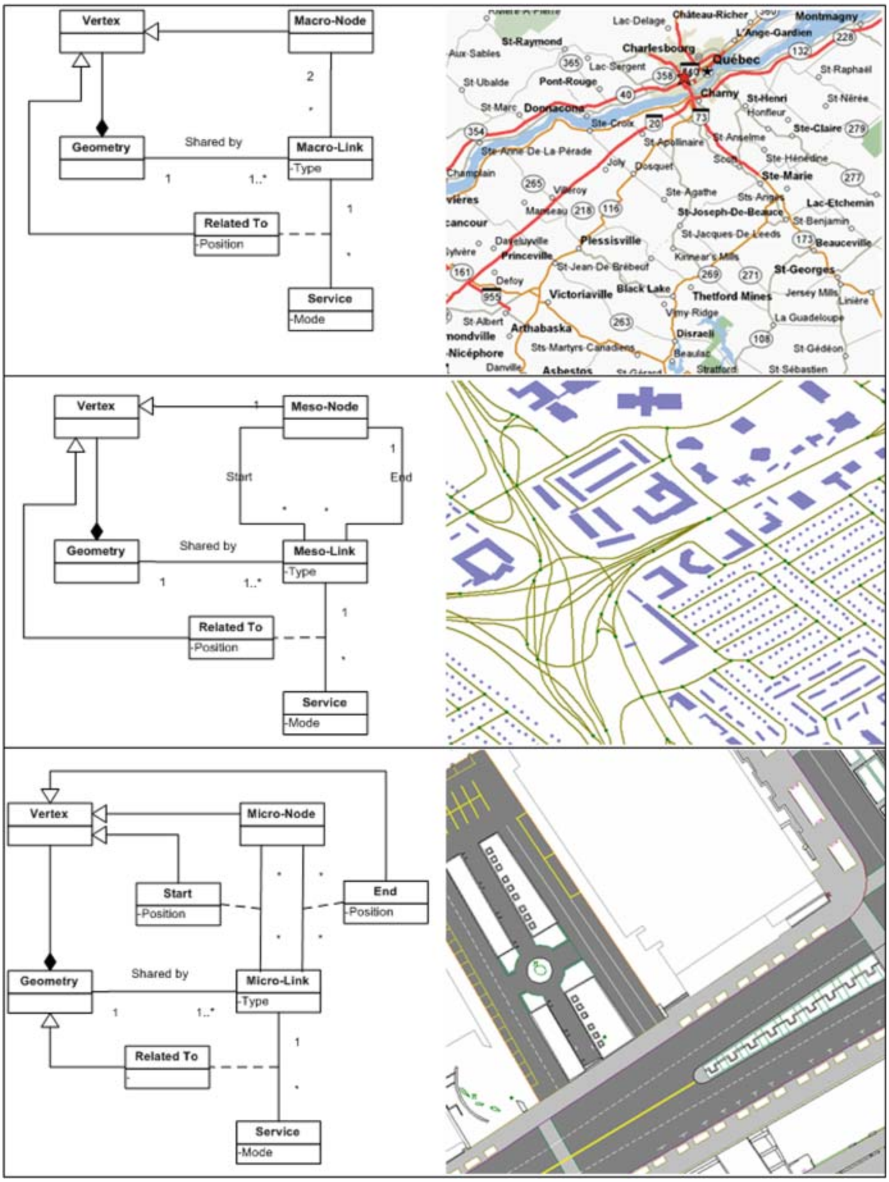



**Fig. 8.3** Macro-model (*top pattern*), Meso-model (*middle pattern*), and Micro-model (*bottom pattern*) of Quebec-city VUE. Maps on the *right hand* side are not VUEs but only GIS cartographic data used as input to generate VUEs

At the micro-scale, the network is a lane-based representation that we can find in traffic micro-simulation software such as Vissim (PTV 2008). A *micro-link* represents a lane and a *micro-node* represents a "connector". A connector allows us to model possible turns at a crossroad for example. *Micro-nodes* remain vertices and they can be onramps or off-ramps. One *micro-node* connects two segments that have the same number of lanes. Unlike the macro- and meso-models, the "related to" relationship between *Service* and *micro-link* inherits from the *Geometry* class. Let us recall that a *Service* means a transfer from one *Link* to another, or an *Access* from one *Link* to a *Place*. The "from" here can be materialized as a single position at macro- and meso-scales. Nonetheless, when considering a micro-scale, a bus station, a parking lot, or an entrance of a building is so detailed that it should be related to a portion of a *Link* rather than to a *Vertex*. The last representation of the train station in Fig. 8.2 shows such a *Service* where the source is a link portion and not a link position (the transfer between the train link and the walk link).

In our case study we are particularly interested in creating the meso- and micro-VUEs. The CRAD Research Center provided us with a topological road network of Quebec-city generated from 1:20,000-scale topographical maps (Thériault et al. 1999). This road network is fully compatible with our VUE meso-model. By considering other GIS data sources such as the municipal assessment roles, we were able to determine *Places* and *Services*. For the micro scale, Quebec-city's Geomatics services provided us with a high resolution map of the entire city. Because this map was a CAD drawing rather than a GIS well-structured and layered map, automatic generation of the micro-VUE was a demanding task. Since at the micro-scale we usually deal with limited spaces, we chose to manually configure the micro-VUE using the micro-scale map as background. Figure 8.4 shows an example of such a generated micro-VUE for an area in the old part of Quebec-city.

## 8.4 The Synthetic Population of Quebec-city

In accordance with *Requirement 1* (Section 8.1), the Quebec-city multiscale VUEs created in Section 8.3 need to be populated in order to carry out simulations of travel activities. This is the travel demand forecasting part of a transport problem where the VUE creation represents an estimation of the supply part. Contrarily to the VUE creation, the generation of a synthetic population is less objective and requires more approximation techniques because available data (mostly in the form of surveys) represent samples and can not cover the entire population.

A number of approaches exist for estimating synthetic populations, including stratified sampling, geo-demographic profiling, data fusion, data merging, iterative proportional fitting, re-weighting, synthetic reconstruction, and their combinations (Huang and Williamson 2001). The synthetic reconstruction approach involves the use of Monte Carlo sampling from a series of conditional probabilities, derived from census surveys, to create synthetic data (Williamson et al. 1998). For example, a combination of iterative proportional fitting and

**Fig. 8.4** A portion of Quebec-city's micro-scale map (*top*) and its associated generated micro-VUE (*Bottom*)

synthetic reconstruction has been used in the TRANSIMS Project (Beckman 1996). Given its flexibility to handle different data sources and its capability to deal with the issue of confidentiality, we also apply a synthetic reconstruction approach that we couple with some Operational Research techniques (the assignment problem) and statistical methods (logistic regression).

Individuals of our synthetic population belong to households. We have different kinds of households (one person, two persons, with or without children, etc.). We consider all relevant information or variables that may be taken into account when choosing a travel destination, as well as trips' frequencies and durations. In order to assign origins and destinations of trips, frequently visited places such as workplaces (and when appropriate, school places) have been assigned to each person of the synthetic population. In addition, several non-spatial attributes such as age, gender, and profession should be considered to allocate workplaces to people because they are predictive variables in the assignment model. The following subsections give more details about that and summarize how we estimated non-spatial (Section 8.4.1) and spatial attributes (Section 8.4.2) of households and of persons.

### 8.4.1 Estimation of Non-spatial Attributes

Our synthetic population is built in conformity to an Origin Destination (OD) survey. Conducted by phone during the fall of 2001 by MTQ (Ministère des Transport du Québec) and RTC (Réseau de Transport de la Capitale) in 2001, this OD survey gives personal and household information for approximately 8% of the total population of the Quebec Metropolitan Area. Respondents accepted to describe trips made during the previous weekday (Monday to Friday) disclosing purpose as well as transport mode and specifying origin and destination places.

Figure 8.5 summarizes the process that we followed to generate our Quebec-city's synthetic population starting from the OD sample population. First, cross tables are compiled from the sample to show various dependencies between

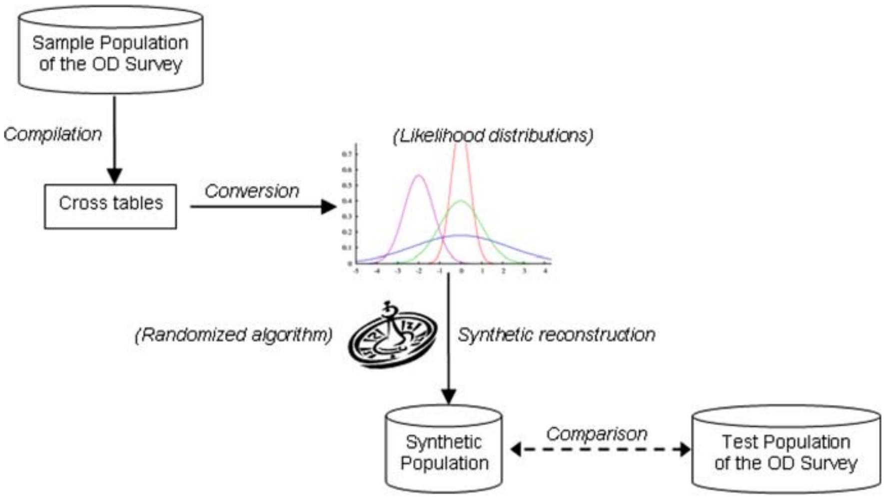



**Fig. 8.5** The entire process of estimating non-spatial attributes when generating the synthetic population

attributes of households and persons. Each cross table is then transformed into a likelihood function which is an input of the synthetic population generator. The households' creation, and the allocation of non-spatial attributes, is performed using a randomized algorithm. The attributes that we considered encompass the main characteristics of households and persons (age, gender, main occupation, possession of a driver's license, owned vehicles, etc.) that can influence travel behaviors. Finally, the generated synthetic population is compared to the OD test population, which is the entire OD population deprived of households that already have been considered in the sample population (initially used to generate cross tables).

In accordance with the *law of large numbers*, let us consider a random variable with a finite expected observations whose values are repeatedly sampled. When the number of these observations increases, their mean will tend toward the expected value of the variable. Fortunately, the size of the generated synthetic population (the number of observations in our case) was large enough to obtain an acceptable solution. Cross tables of the generated population were compiled and manually compared to the original ones. Figure 8.6 shows an example of the correlation between three variables: driver's license, age, and gender. Histograms of Fig. 8.6 prove that we obtained for the generated synthetic population a similar distribution as the one characterizing the OD sample population. Advanced goodness-of-fit measures for synthetic micro-data have been proposed in the literature (Voas 2001) and could be used in the validation step. Indeed, the usefulness of such approaches lies in their capability to automate the process of generation and validation. In our work, attribute dependencies have been detected intuitively and then statistically verified. The automatic identification of relevant correlations between variables would be an interesting way to generalize our approach and, then, to automate the process. We might use classification algorithms based on the measurement of Shannon's entropy that characterizes the homogeneity of instances in order to construct decision trees (Quinlan 1993). With such a decision tree, an automatic procedure that compiles likelihood functions can be called at runtime. Unfortunately, the construction of the decision tree and the computation of the likelihood function need to directly operate on the sample, and consequently the confidentiality of information provided by respondents would be broken.

### 8.4.2 Allocation of Spatial Attributes

Yet, vigorous debates still occur over the role played by various determinants of the relationship between urban form and commuting duration which is still poorly understood (Vandersmissen et al. 2003). Some studies are using gravitational models to represent interaction flows between home places and workplaces (Tiefelsdorf 2003), while others rely on the actual location of homes and workplaces to model relative accessibility using transport GIS in order to

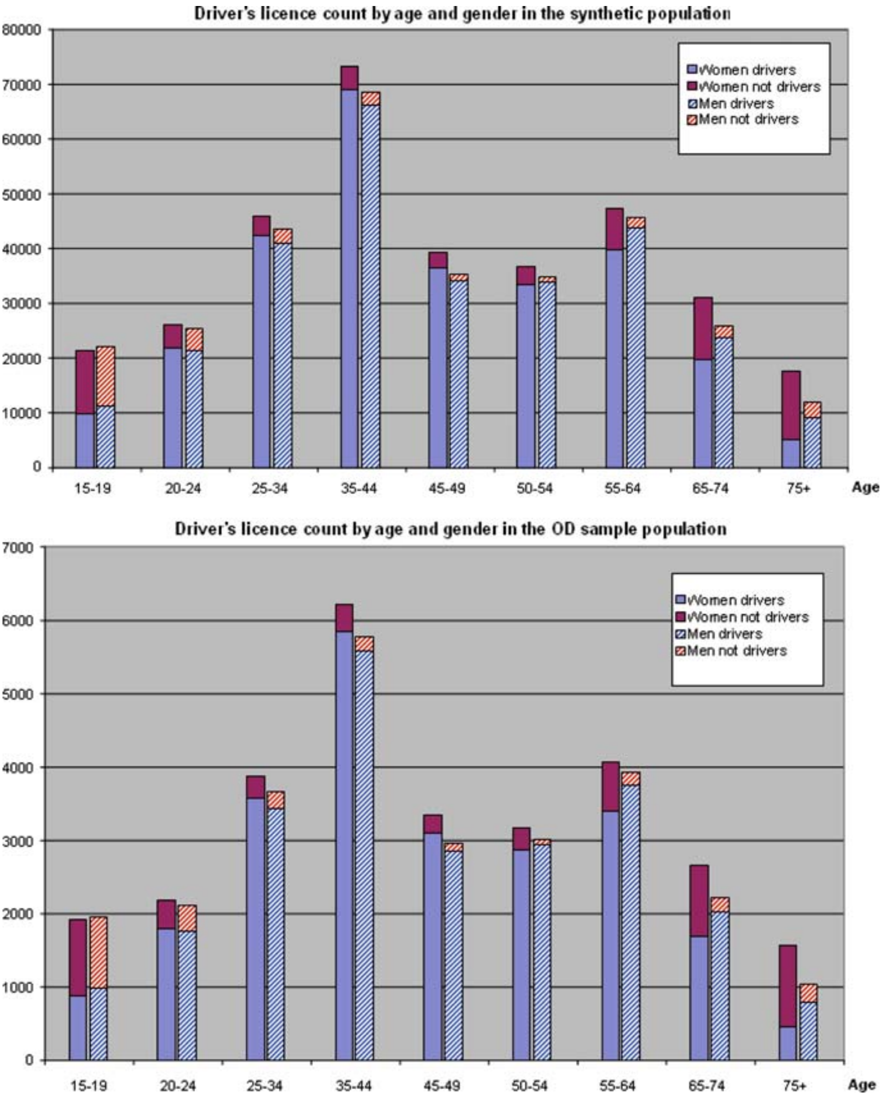



**Fig. 8.6** Example of validating non-spatial attributes estimation: Driver's license counts follow the same distribution in both histograms: the generated synthetic population (*top*) and the OD sample population (*bottom*)

implement time-geography concepts (Kim and Kwan 2003; Thériault et al. 2005). In this work, we investigate the use of a gravitational model in order to allocate a workplace to each active person of the synthetic population. We present here the assignment approach that we used to allocate regular or frequently visited places (home places, workplaces, and schools) to each active person of the synthetic population. Even if other activity places such as shopping and leisure places are not yet considered, we are able to reproduce a large part of the car traffic because 71.6% of trips in the OD survey are related to work, school, and going back home.

We first aggregated locations by using a hexagonal grid in which each cell is 250 meters wide in order to hide confidential data. The assignment of households to home places is quite simple: From the OD survey, the number of housing units is estimated for each grid cell using an expansion factor. The resulting values are then calibrated in comparison to census and assessment data. Each household is then associated with a unit according to variables such as *number of workers* and *number of children*. While having information about buildings (or blocks), we randomly allocate households to blocks within cells.

Once home places are assigned to households, workplaces are assigned to workers. To this end, we adopted an algorithm in which relationships between home places and workplaces are quantified in terms of the *predicted time to go to work* variable. This predicted value is estimated for each person of the synthetic population using the same probabilistic approach as the one used to allocate non-spatial attributes. To formally define the problem, let $P_1, P_2, \ldots, P_n$ be a set of persons who already have a home place and let $W_1, W_2, \ldots, W_n$ be $n$ candidate workplaces (Fig. 8.7). The assignment consists of determining $n$ disjoint couples $(P_i, W_j)$ in order to minimize the difference $u(i,j) = |O(i) - D(i,j)|$ where $D(i,j)$ is a travel time (or cost) from home place of $P_i$ to location $W_j$ (that we compute using the router module of TransNetSIM, which will be detailed in Section 8.5) and $O(i)$ is the "*predicted time to go to work*" related to $Pi$. Once values of $u(i,j)$ are obtained, the assignment can be efficiently solved using a *Weighted Bipartite Matching* resolution approach. Here we assign people to work places in such a way that the travel time is the nearest one to the *predicted time to work* already computed for each person according to attributes such as *profession* and *home places*. We construct the bipartite graph $G(A, B, E)$ where:

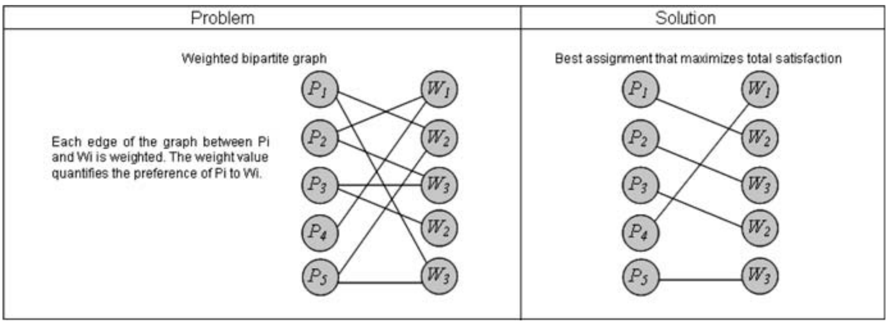



**Fig. 8.7** The weighted bipartite matching as a solution to the workplace assignment problem

- $A$ is the set of vertices representing $P_1, P_2, \ldots, P_n$ and $B$ is the set of vertices representing $W_1, W_2, \ldots, W_n$
- $E$ is the set of weighted edges between vertices of $A$ and vertices of $B$ where the edge $(i,j)$ is included in E only if $u(i,j)$ is bigger than a certain threshold $\lambda$.

In order to perform the assignment by using the bipartite matching algorithm as explained before, we classified jobs into 8 sectors (education, health, services, retail trade, wholesale trade, public services, leisure) and 5 job types (manager, professional, skilled or specialized employee, unskilled worker, unspecialized employee). Thus, we generated 40 distinct bipartite graphs. We applied the Goldberg-Tarjan algorithm (Goldberg 1997), one of the most efficient polynomial time solvers of the *minimum cost flow problem,* which is a general case of the *bipartite matching problem*. The biggest bipartite graph has $n = 31447$ vertices and was solved in 30 min by a 2.66 GHz mono-processor. Each graph was pruned in order to let its degree bound between 50 and 200 (which means that each vertex is linked to a minimum of 50 and a maximum of 200 vertices of the opposite side). The mean of gap values between predicted and allowed travel times is 4.32 min and the distribution of gap values is shown in Fig. 8.8. The log-normal distribution that we obtained with a low mean value proves the success of the algorithm since allocated and predicted travel times are equal to the minute.

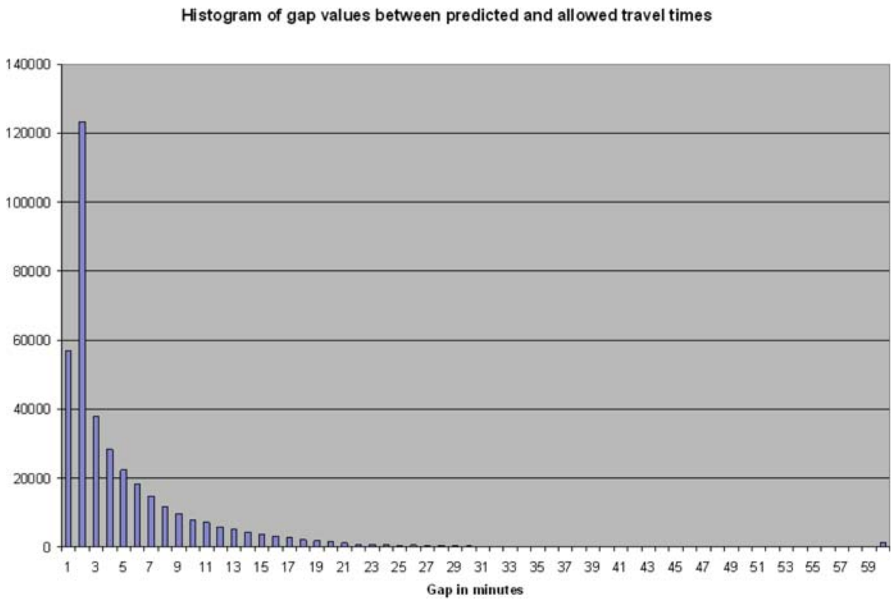



**Fig. 8.8** Low gaps between predicted and allowed travel times prove the success of workplaces' assignment by using the weighted bipartite matching solver

School locations were also assigned to the students of the synthetic population. Unlike workplaces, the assignment of school locations does not require travel time prediction because in most cases, students go to the school nearest to their home locations. Since we obtained school locations and their capacities for the Quebec-city metropolitan area, the assignment was quite simple. Nevertheless, college and university students were randomly assigned to education places because in reality, the choice of establishment depends on the area of study more than on home proximity. If we can obtain more details about schools' capacities in relation to study domains, we could then apply a similar approach as the one used for workplace assignment using the weighted bipartite matching solver.

## 8.5 TransNetSIM: Using the Meso-VUE with Associated Synthetic Population for Simulating Travel Activities

TransNetSIM is a meso-simulator that we developed using the meso-VUE model: the data structures manipulated by the simulator directly correspond to the meso-VUE presented in Section 8.3. A meso-environment certainly corresponds to the scale that provides the best level of detail to plausibly simulate traffic and related activities taking place in large urban areas, given the availability of relevant data and computational constraints. In our case, the scope of our meso-environment comprises the entire Quebec-city's metropolitan region with a network of 32,071 nodes and 80,855 links. This is the level of detail at which the entire synthetic population can plausibly be modeled. The purpose of this experiment is to plan and simulate the daily activities of the individuals of our synthetic population.

### *8.5.1 Pre-processing Data for TransNetSIM*

Figure 8.9 displays a schematic view of the input data required for the TransNetSIM simulator. Using the meso-environment's data structures, we developed a path-finding module that computes all the possible paths from each node $S$ to all other nodes $V$ of the network. To this end, we implemented the well-known Dijkstra algorithm with a Fibonacci heap data structure which has a complexity of $O(m + n \log(n))$ for time and $O(n^2)$ for space where $n$ is the number of nodes and $m$ the number of links (or edges). While the Dijkstra's algorithm is widely used, it is important to remember that it is a *single-source shortest path algorithm*. This means that it computes all paths from a source node $S$ to any other node in the network each time the algorithm is called on to find a path from $S$ to any destination $V$. Hence, obviously we save a lot of time if we call Dijkstra only $n$ times rather than $n^2$ for each couple $(S, V)$ of nodes. This is known as a naïve implementation of the *all-pairs shortest path problem*. Other

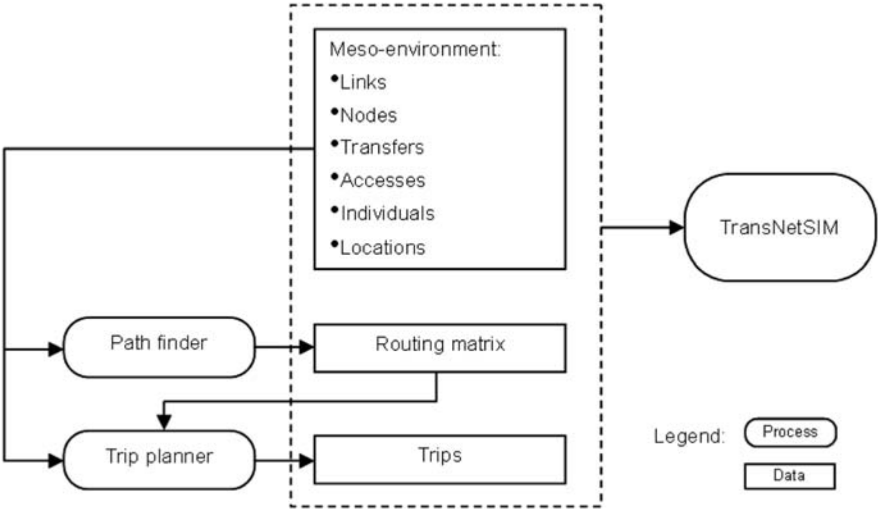



**Fig. 8.9** Pre-processing data for TransNetSIM

implementations that take into account the network specificity are available (Seidel 1992, Wagner and Willhalm 2003). However, a naïve implementation of the *all-pairs shortest path problem* using the Dijkstra's single source algorithm satisfies our constraints for the time being. Currently, the path finding module takes about 25 min to compute and to store all paths of our network in a routing matrix on a 2.66 GHz mono-processor with 2 GB RAM. The computation of the impedance of each link is based on the *travel time* constant computed according to the length of the link and its maximal speed. Our path finder module also considers turn penalties: 0.4 min for turning left, 0.2 for right and 0.1 for going through (Thériault et al. 1999). Naturally, making a U-turn and driving down a one-way street are prohibited. To this end, we revised the original form of the Dijkstra algorithm in order to consider turn penalties. Currently, the size of the routing matrix file is 986 Mb. If we consider the maximum order of the graph (the maximum number of outgoing edges for each node) we could theoretically reduce the storage space of the routing matrix to 368 Mb.

Besides accessing the structured meso-environment as our *path finder* does, the *trip planner* accesses the *routing matrix*. It schedules individual activities during an entire work day and generates trips that represent the final form of the travel demand. An activity is characterized by a *place* (a location which is a parcel in the meso-environment), a *type* (work, shopping, leisure, etc.), and some spatio-temporal constraints. Depending on the scheduled activities, and according to the required time to reach each activity place (computed directly by using the routing matrix), we generate trips.

Trips are the final representation of the travel demand while individuals modeled in our synthetic population are the initial representation. Each trip is simply defined by:

- An *origin* and a *destination* which are *Services*. Each of them represents a fixed position on a link. Let us recall that a *Service* represents either a transfer from one *Link* to another or an *Access* from one *Link* to a *Place*.
- Either a *departure time* or a *pair of elements* (references to the previous trip duration and to the last activity). If a departure time is given, the system has just to start the trip at this time of the simulation. Otherwise, the simulator has to wait for the total duration of the last activity combined with the arrival time (obtained during the simulation) of the previous trip.
- A *transportation mode*: we consider only one mode for each trip. This does not mean that multi-modality is not allowed but it is the responsibility of the *trip planner* module to chain modalities.

## 8.5.2 Simulation of Trips Using TransNetSIM

As shown in Fig. 8.9, TransNetSIM needs the routing matrix, the list of trips and the meso-environment (mainly the network) as inputs. The system starts by loading both links (including their vertices) and the routing matrix in the computer memory. Trips are also loaded in memory, but they are enhanced with the following runtime attributes: (1) *State* that takes as a value *waiting-departure*, *in-progress,* or *finished*, (2) *current link index* which is used only if the state is *in-progress*, (3) *current position* which is a time-related distance on the link, and (4) *arrival time* which is required for chained trips. Currently, the vehicles' behaviour simulated in TransNetSIM is quite simple and is based on a time-based function that updates the *current position* and the *current link index* attributes at each time step (a variable parameter specified by the user) according to links' impedances and turn penalties. Links capacities are not taken into account yet, but we will be able to compute them in the near future since the *number of lanes* and *crossing speed* attributes are available in our network.

Since origins and destinations of trips are *Services*, we need to apply a light transformation to the path-finding procedure in order to be able to exploit the routing matrix that can only be used to compute the shortest path between two nodes. In the case of a *Service* related to a directional link (a one way road), we simply take the outgoing node. In the case of a bidirectional link, the transformation comes down to comparing four paths combinations as shown in Fig. 8.10. In theory, we could also convert each *Service* into a node and split the corresponding link in order to apply shortest path queries. However, the routing matrix size would increase drastically. Let us recall that a link can be associated with a high number of services as in the case of parking lots (Fig. 8.2).

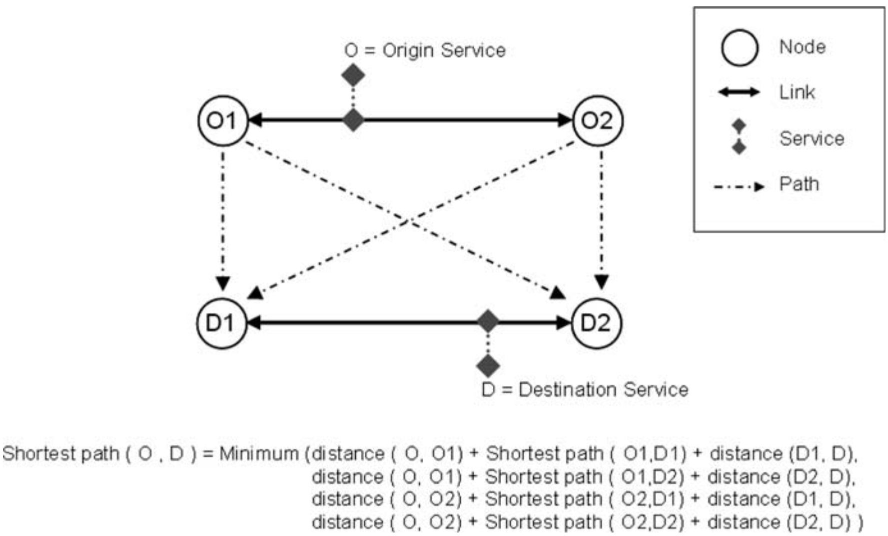



**Fig. 8.10** Shortest path computation between two Services positioned on bidirectional Links

### *8.5.3 TransNetSIM's Performance*

TransNetSIM was developed using Borland's *C+ + Builder* for *Windows* platform. Figure 8.11 shows the GUI (Graphical User Interface) of TransNetSIM. When the simulation is in progress, the GUI allows the inspection of each running trip and displays the traffic state at each *di* time step (*di* is a number that depends on a display frequency specified by the user). The simulation of 1 million random trips on the Quebec-city's network (32 K nodes and 81 K links) takes 37 s (without counting display time) on a 2.66 GHz single-processor computer with 2 GB RAM. When applying TransCAD to the same simulation data, TransNetSIM is 757 times faster (276 202 trips were simulated in 2 h and 9 min using TransCAD). This comparison test was done by developing a C+ + prototype working as a client of the TransCAD server. The vehicle updating routine is exactly the same as in TransNetSIM, except that when starting a trip we call a TransCAD path finder function. It is also worth mentioning that, thanks to the routing matrix, no route has to be stored when using TransNetSIM for the simulation, contrary to the TransCAD test.

Similarly, let us compare this aspect of our system to the well-known TRANSIMS simulator (Barrett and Beckman 2004). The TRANSIMS route planner computes the "shortest" path, subject to mode constraints, for each traveler in the system and stores routes in several files that are some of the inputs of the microsimulator. An important difference with our approach is that our path finder module, Fig. 8.9, generates a routing matrix and not travelers'

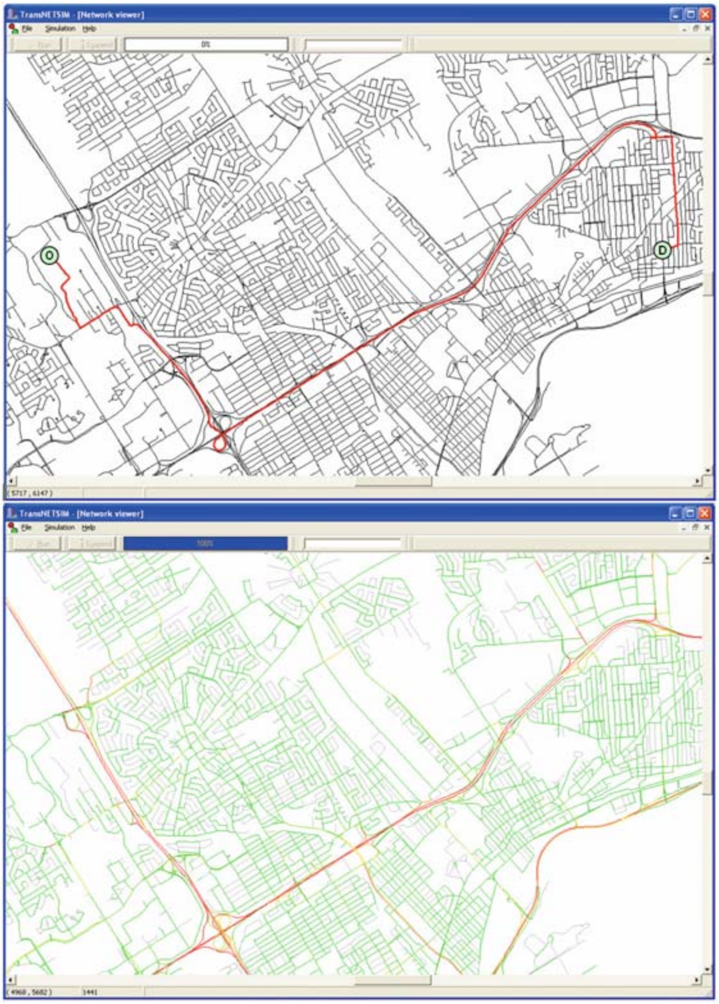



**Fig. 8.11** TransNetSIM interface with QC network of 81 K links and 32 K nodes. Screenshot on top highlights a fastest path from O to D and the screenshot on the bottom shows the state of the traffic after the simulation of one day's activities

routes. The routing matrix is a compilation of all possible shortest paths. Hence, there is no need to store huge files and there is also no need to rely on parallel computing to process all these files. Many efforts in the TRANSIMS project were spent to find the best "one to one" shortest path algorithms for the router module (Jacob et al. 1999). In our case, we proved that a "single source" shortest path algorithm is suited for setting the routing matrix in a reasonable time before starting the simulation even with a very large network.

## 8.6 Conclusion

In this chapter we introduced a novel approach to model Virtual Urban Environments (VUE) by taking into account: (1) the multiscale issue, (2) activities' locations, multi-modal transport networks and their relationships, and (3) the associated synthetic population. The new modeling approach is based on a scale-free design pattern that introduces the notion of *Service* as a way of representing locations' accesses as well as modal shift opportunities using the transportation network.

Currently, meso-scale simulations of daily travel activities have been performed with our TransNetSim VUE-based software. At the same time, our VUE modeling approach remains fully compatible with microscopic simulations, which operate on restricted and high-resolution spaces (such as those used in microscopic traffic simulators) thanks to the design pattern approach.

TransNetSim currently operates on a meso-VUE. It is functionally similar to mesoscopic traffic simulators in which vehicles are considered individually (as in a microscopic approach) while their vehicle updating behaviour handles aggregated variables such as flow densities (as in the macroscopic approach). Currently we couple TransNetSim with the commercial traffic microsimulator Vissim, which provides an environment that is compatible with our micro-VUE (Fig. 8.3) where a *micro-link* represents a lane and a *micro-node* represents what is called a *connector* in Vissim. For experimental purposes, we developed a micro-simulation that takes place on the campus of Laval University. Since the campus is considered a unique location that has several accesses in the meso-VUE, each building becomes a location in the micro-VUE in which we model all parking slots and bus stations as *transfers* and all buildings' entries as *accesses*.

Our initial description of the VUE design pattern (Section 8.2) using an object-oriented formalism made it easy to transition from the higher conceptual level to the lower logical view when we built TransNetSim. In addition, the design pattern approach avoids the drawbacks of a strictly hierarchical representation where each spatial concept is divided into concepts of the same type at a finer level (or scale). For example, the Laval University's campus, which is represented as a location in the meso-VUE, becomes an aggregation of services, links, nodes, and locations in the micro-VUE, and not only to a set of locations. Moreover, we think that we can exploit the "design pattern" concept in an even better way. Indeed, while coupling TransNetSim with Vissim is sufficient to prove the consistency and coherence of our model, a fully integrated micro- and meso-scale version of TransNetSim would benefit from the model capabilities even more.

Our work is a first step toward a long-term objective of modeling VUEs as a framework in which many multiscale urban phenomena could be integrated. In particular, we are investigating how it is possible to enhance our VUE using cellular automata, which are widely used in the simulation of urban phenomena like urban sprawl.

**Acknowledgments** This research was financed by GEOIDE, the Canadian Network of Centers of Excellence (MUSCAMAGS Project) and supported by various partners such as Centre de recherche en Aménagement et Développement (CRAD at Laval University), Ville de Québec, Ministry of Transportation Québec and Social Sciences and Humanities Research Council of Canada (CRSH).

## References

Albeverio, S., Andrey, D., Giordano, P. & Vancheri, A. (2008). *The dynamics of complex urban systems: an interdisciplinary approach*. (Berlin: Springer)

Alexander, C. (1977). *A pattern language: towns, buildings, construction*. (New York: Oxford University Press)

Barrett, C.L. & Beckman, R.J. (2004). *TRANSIMS, Volume Three*. Los Alamos Unclassified Reports 00-725 and 00-1767

Batty, M. (2005), Cells, and cities: new representational models for simulating multiscale urban dynamics. *Environment and Planning A*, 37, 1373–1394

Beckman, R.J., Baggerly K.A. & McKay, M.D. (1996). Creating synthetic baseline populations. *Transportation Research A*, 30(6), 415–429

Crucitti, P., Latora, V., Porta, S. (2006). Centrality measures in spatial networks of urban streets. *Physical Review E – Statistical, Nonlinear, and Soft Matter Physics,* 73(3), 1–5

Donikian, S. (1997). VUEMS: a virtual urban environment modeling system. *Proceedings of Computer Graphics International conference*, 1997, 84–92

Goldberg, A V. (1997). An efficient implementation of a scaling minimum-cost flow algorithm, *Journal of Algorithms*, 22(1), 1–29

Huang, Z. & Williamson, P. (2001). A comparison of synthetic reconstruction and combinatorial optimization approaches to the creation of small-area microdata, Working paper, department of Geography, University of Liverpool

Jacob, R., Marathe, M.V. & Nagel, K. (1999). A computational study of routing algorithms for realistic transportation networks. Special issue of *ACM Journal: Experimental Algorithmics,* 4, 6–25

Jiang, B. (1999). SimPed: simulating pedestrian flows in a virtual urban environment. *Journal of Geographic Information and Decision Analysis*, 3(1), 21–30

Jiang, B. & Liu, C. (2007). Street-based topological representations and analyses for predicting traffic flow in GIS. *International Journal of Geographical Information Science*, 23(9), 1119–1137.

Kim, H.M. & Kwan, M.P. (2003). Space-time accessibility measures: a geocomputational algorithm with a focus on the feasible opportunity set and possible activity duration. *Journal of Geographical Systems*, 5, 71–91

Mackaness, W.A., Ruas, A. & Sarjakoski, L.T. (2007). *Generalization of geographic information cartographic modeling and applications*. (Holland: Elsevier)

Miller, H.J. & Shaw, S.L. (2001). *Geographic information systems for transportation: principles and applications*. (USA: Oxford University Press)

Moulin, B., Chaker, W., Perron, J., Pelletier, P., Hogan, J. & Gbei, E. (2003). MAGS Project: multi-agent geosimulation and crowd simulation. In W. Kuhn, M. F. Worboys & S. Timpf (Eds.), Spatial Information Theory. Foundations of Geographic Information Science: International Conference, COSIT 2003, Ittingen, Switzerland, September 24-28 (pp. 151–168). (Berlin: Springer)

PTV, Vissim User Manual - V.4.20. Karlsruhe, Germany, http://www.ptvag.com/traffic/software-system-solutions/vissim/ -(2008)

Quinlan J. (1993). *C4.5: programs for machine learning*. (San Francisco: Morgan Kaufmann Publishers)

Rosvall, M., Trusina, A., Minnhagen, P., Sneppen, K. (2005). Networks and cities: an information perspective. *Physical Review Letters,* 94(2), 1–4

Seidel, R. (1992). On the all-pairs-shortest-path problem. *Proceedings of the twenty-fourth annual ACM symposium on theory of computing, STOC '92* (pp. 745–749). (New York: ACM Press)

Thériault M., Des Rosiers, F. & Joerin, F. (2005). Modeling accessibility to urban services using fuzzy logic: a comparative analysis of two methods. *Journal of Property Investment and Finance*, 23(1), 22–54

Thériault, M., Vandersmissen, M.H., Gosselin, M.L. & Leroux, D. (1999). Modeling commuter trip length and duration within GIS: Application to an O-D survey. *Journal for Geographic Information and Decision Analysis*, 3(1), 41–55

Tiefelsdorf, M. (2003). Misspecification in interaction model distance decay relations: a spatial structure effect. *Journal of Geographical Systems*, 5, 25–50

Vandersmissen, M.H., Villeneuve, P. & Thériault, M. (2003). Analyzing changes in urban form and commuting time. *The Professional Geographer*, 55(4), 446–463

Voas, D., & Williamson, P. (2001). Evaluating goodness-of-fit measures for synthetic microdata. *Geographical and Environmental Modelling,* 5(2), 177–200.

Wagner, D. & Willhalm, T. (2003). Geometric speed-up techniques for finding shortest paths in large sparse graphs. In G.D. Battista & U. Zwick (Eds.), *Proceedings of the 11th European symposium on algorithms* (ESA 2003) (pp. 776–787). (Berlin: Springer)

Williamson, P., Birkin, M. & Rees, P. (1998). The estimation of population microdata by using data from small area statistics and samples of anonymised records. *Environment and Planning A*, 30, 785–816

# Chapter 9
# Imageability and Topological Eccentricity of Urban Streets

**Itzhak Omer and Bin Jiang**

**Abstract** Previous studies of the influence of structural qualities of urban street network on the image of the city are based mainly on centrality and connectivity measures taken from graph and space syntax theories. The paper suggests application of the structural property of *eccentricity* for considering the structural distinctiveness or differentiation of a given street in the overall street network. Eccentricity suggested by Q-analysis and based on the perspective of multidimensional chains of connectivity. This structural property is applied to the case of the city of Tel Aviv by using a geographic database of the street network and observed data acquired from Tel Aviv residents' production of sketch maps. The study's findings provide preliminary evidence for the relevance of the structural property of eccentricity for understanding the relationships between street network and the image of the city.

**Keywords** Topological analysis · Imageability · Legibility · Street patterns · Mental maps

## 9.1 Introduction

Much evidence has been collected indicating that the physical characteristics of an urban environment have the potential to affect the acquisition of environmental spatial knowledge (Cubukcu and Nasar 2005; Garling et al. 1986; Weisman 1981). In his seminal work "The image of the city", Lynch referred to this potential effect of the urban environment with the term *legibility* of the cityscape, meaning "the ease with which its parts can be recognized and can be organized into a coherent pattern" (Lynch 1960, p. 2). The underlying order

I. Omer (✉)
The Environmental Simulation Laboratory,Department of Geography and Human Environment, Tel Aviv University Ramat-Aviv, Tel Aviv 69978, Israel
e-mail: omery@post.tau.ac.il

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_9,

derived from a legible city allows people to familiarize themselves with the environment, move about and spatially orient themselves.

According to Lynch's theory, city legibility is highly connected with the property of the urban environment called *imageability* – "that quality in a physical object which gives it a high probability of evoking a strong image in any given observer" (Lynch 1960, p. 9). Imageability can be analyzed by three components: identity, structure and meaning. "A workable image requires first the identification of an object, which implies its distinction ...... as a separable entity. Second, the image must include the spatial or pattern relation of the object to the observer and to other objects. Finally, this object must have some meaning for the observer, whether practical or emotional" (Lynch 1960, p. 8).

Streets or paths are the predominant city elements and part of them may become important elements due to certain characteristics that strengthen their image in observers' minds (Lynch 1990). Lynch mentioned several such characteristics: concentrated spatial use or activity along a street, a special façade, proximity to unique features and spatial characteristics like the width or narrowness of streets (Lynch 1960, pp. 49–51). With reference to the structural qualities of individual streets within a city, he noted that a street can acquire importance or uniqueness due to its place within the street network – such as its location in a *path intersection* or a *branching of streets* – when these features are treated as strategic points (p. 98).

Despite its great contribution to understanding the involvement of urban environment in the construction of spatial cognitive representation, Lynch's theory has been criticized mainly with respect to its intuitive character and the absence of analytic explanation of imageability. Conroy-Dalton and Bafna (2003) for example, argued that Lynch's statement "does not offer any arguments for why some elements are selected over others, claiming instead that their distinctively visual characteristics are responsible" (Conroy-Dalton and Bafna 2003, p. 59.4). In this respect, it is interesting to note that Lynch was aware of the absence of appropriate tools for estimating the quality of structure, and noted that "... there was a lack of information on element interrelations, patterns, sequences and wholes. Better methods must be evolved to approach these vital aspects" (Lynch 1960, p. 155). However, it is clear that this aim is more achievable today, with the development of advanced structural analysis methods and access to massive digital geographic databases.

Much research has shown that a topological analysis can be used as appropriate means to evaluate the effect of structural qualities of an urban element on imageability, as well as on other aspects of spatial cognition. Studies concentrating on street networks have found that topological properties, described mainly by centrality measures of network connectivity, could be reliable indicators for describing and predicting the effect of street network topology on the city's imageability (e.g. Haq and Zimring 2001; Conroy-Dalton and Bafna 2003; Shokouhi 2003; Claramunt and Winter 2007; Yun and Kim 2007;

Tomko et al. 2008). The underlying logic of this research is the assumption that streets that have greater structural importance – i.e., higher centralities – tend to be places with which people have more experience, making them more prone to imageability. It should be noted that these topological properties have also been found appropriate for evaluating wayfinding performance, mainly with reference to the overall street network or to paths within this network (Kim 2001; Penn 2003; Omer and Goldblatt 2007).

The contribution of these structural/topological studies is clear: they improve our knowledge about the involvement of urban structural qualities in the construction of the image of the city. More specifically, they enable us to investigate how the status and contribution of a given element to the connectivity in the network affect its imageability. Nevertheless, these structural/topological studies are not referring to the distinctiveness or differentiation of urban elements, which play an important role in the Lynchian theory and imageability studies (e.g: the Abu-Ghazzeh 1997; Abu-Obeid 1998; Haken and Portugali 2003; Rosvall et al. 2005): In other words, they omit the role of differentiated spatial elements and the sense of visual distinction (Conroy-Dalton and Bafna 2003).

In light of the above distinction, studies on how physical qualities of a given urban environment affect its imageability can be placed into two views. The first focuses on the identity of urban elements that stem from their distinctive visual quality, while the second stresses the structural properties of an urban element, namely, its role in the city as an integrated whole, i.e. its relative importance in the overall spatial organization.

The aim of the current study is to suggest an application of the structural property of eccentricity, which is based on the perspective of multidimensional chains of connectivity, to the structural investigation of the relationship between an urban street network and the image of the city. Eccentricity enables us to consider the structural distinctiveness or differentiation of a given urban element in the overall spatial organization. It thus has the potential to offer not only insights different from the two views described above, but also a sort of integration between them by combing the properties of distinctiveness and structure in one measure. In addition to that, the application of eccentricity is motivated by the awareness that the multidimensional chain of connectivity of the street networks has the potential to assist and complement an examination of the relation between urban street network and imageability.

In the following sections we present a conceptual discussion of the link between urban street network, imageability and topology, with a special focus on the distinction between graph theory's centrality measures and the eccentricity measure of Q-analysis. We then describe the results of the investigation on the relationship between structural properties and the urban street network's image as computed with the two kinds of measures, and a comparison between them.

## 9.2 Topology of Urban Street Networks and Imageability

A topology of urban streets takes individual streets as nodes and street intersections as edges of a connectivity graph; see Fig. 9.1, where the kite-shaped graph is taken from Krackhardt (1990). The graph forms a basis for structural analysis using the centrality measures (Jiang and Claramunt 2004) initially developed for the description of social networks (Freeman 1979). Like Freeman (1979), Krackhardt (1990) used the "kite structure" to analyse power relations among groups of actors, pointing out that "all three of these measures can be illustrated by the position of actors in this hypothetical network and it represents the smallest network I have found in which the centrality based on each of the three measures reveals different actors as the most central in the network." Three centrality measures – *degree, closeness* and *betweenness* – are used to describe the status of individual streets in terms of which streets intersect with other streets. These topological properties are called connectivity, global integration and choice, respectively in space syntax (Hillier and Hanson 1984). *Degree* indicates how many other streets are connected directly to a particular street, a characteristic that reflects the level of a street's integration with its neighboring streets. *Closeness* indicates how close a street is to other streets by computing the shortest distances between every street node to every other street node, a feature that reflects how well a street is integrated within the network. *Betweenness* centrality indicates the extent to which a street is located between pairs of streets; as such, it directly reflects the intermediate location of the specific street. The values of these measures obtained for this network are illustrated in Table 9.1; street 6 is the most connected directly, streets 4 and 5 the most connected to all the streets (directly and indirectly) and street 3 is placed in the most intermediate location.

The relationship between the structural qualities of urban streets and imageability has been investigated through the use of the values of the

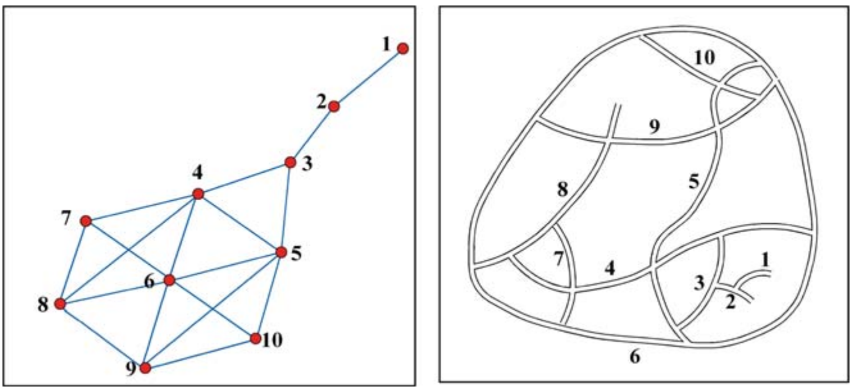



**Fig. 9.1** Topologic (*left*) and geometric (*right*) representation of a notational street network

**Table 9.1** Centrality measures for the notational street network shown in Fig. 9.1

| Street | Degree | Closeness | Betweenness | Eccentricity |
|---|---|---|---|---|
| 1 | 1 | 0.31 | 0.00 | 0 |
| 2 | 2 | 0.43 | 0.22 | 1 |
| 3 | 3 | 0.60 | 0.39 | 1 |
| 4 | 5 | 0.64 | 0.23 | 2 |
| 5 | 5 | 0.64 | 0.23 | 2 |
| 6 | 6 | 0.60 | 0.10 | 3 |
| 7 | 3 | 0.50 | 0.00 | 1 |
| 8 | 4 | 0.53 | 0.02 | 1 |
| 9 | 4 | 0.53 | 0.02 | 1 |
| 10 | 3 | 0.50 | 0.00 | 1 |

centrality measures described above when compared to people's perception tests. These studies have reported that the appearance of a street in the sketch maps of people who visited a space was found significantly correlated with one or more values of the centrality measures (e.g.: Haq and Girotto 2003; Shokouhi 2003; Yun and Kim 2007). It was found that the imageability of a given street segment can be influenced by the topological properties of closeness or integration (i.e., the short topological distance from all other street segments/axial lines; see Conroy-Dalton and Bafna 2003; Shokouhi 2003; Yun and Kim 2007), connectivity (Haq and Zimring 2001) and betweenness (Tomko et al. 2008).

Unlike the graph theory' centrality measures described above, eccentricity measure is taken from Q-analysis that enables us to deal with a multidimensional chain of connectivity. Q-analysis is a mathematical language for the description and analysis of structural complexity, initially developed from algebraic topology by Atkin (1974, 1977). It provides a computational language for the structural description of relationships among a set or between two sets of objects, using geometric and algebraic representations. For the sake of simplicity, we present the main procedure applied with Q-analysis with respect to street networks.

In Q-analysis terms, we represent the street network as a simplicial complex in which streets and street interactions are represented as simplices and vertices. We first generate an incidence matrix in terms of how one street intersects another, i.e., $\lambda = \mathrm{S} \otimes \mathrm{S}$: if the streets $\mathrm{S_i}$ and $\mathrm{S_j}$ from set $S$ (i.e., $\mathrm{S_i}, \mathrm{S_j} \in \mathrm{S}$) are intersected, then the corresponding entry of the matrix is 1, otherwise 0:

$$\Lambda_{\mathrm{ij}} = \begin{cases} 1: & if\ (s_i, s_j) \in \lambda \\ 0: & otherwise \end{cases} \tag{9.1}$$

Using Equation (9.1), our example of a street network (Fig. 9.1) can be represented as an incidence matrix as follows:

$$\Lambda = \begin{array}{c|cccccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 4 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \\ 5 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 \\ 6 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 \\ 7 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 8 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 9 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 \\ 10 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \end{array}$$

Thus, each street, $S_{i,}$ can be considered a simplex in $K_S(S;\lambda)$. For instance, street *3* is represented as a 2-dimentional simplex, $\sigma^2(3) = <2, 4, 5>$. Two key concepts of Q-analysis are q-nearness and q-connectivity. Two streets are q-near if and only if they intersect q + 1 common streets, e.g., streets *4* and *5* are 1-near since they share streets *3* and *6*. Two streets (simplices) are q-connected if they are q-near or there is a chain of streets (simplices) between them that are q-near. For instance, the streets *1* and *4* have no common street, but they are 0-connected due to the fact that both streets are 0-near with street *3*. Thus, q-nearness relates to a pair of streets that intersect while q-connectivity relates to a pair of streets that do not intersect but are connected via a chain of intermediate q-near pairs. In effect, Q-analysis is intended to detect q-connected components at every dimension of a simplicial complex ranging from 0 up to q-1. For example, the Q-analysis of the street network in Fig. 9.1 leads to the following q-connected components (equivalence classes) at different dimensions:

$$\begin{aligned} q &= 0: \{1,2,3,4,5,6,7,8,9,10\} \\ q &= 1: \{2\}\,\{3,4,5,6,7,8,9,10\} \\ q &= 2: \{3\}\,\{4,5,6,8,9\}\,\{7\}\,\{10\} \\ q &= 3: \{4\}\,\{5\}\,\{6\}\,\{8\}\,\{9\} \\ q &= 4: \{4\}\,\{5\}\,\{6\} \\ q &= 5: \{6\} \end{aligned}$$

The structural property of eccentricity is actually an assessment of the status of individual simplices within the complex. In street network terms, eccentricity indicates how a street is unique from other surrounding streets. This measure is defined by the relation between a dimension where a street is disconnected and another dimension where the street is integrated. It is formally defined by

$$Ecc(\sigma_i) = \frac{\hat{q} - \check{q}}{\check{q} + 1} \tag{9.2}$$

where $\hat{q}$ or top-q denotes the dimensionality of the simplex ($\sigma_i$); $\check{q}$ or bottom-q denotes the higher q level where the simplex is connected to any other simplex. According to this definition, if a street is q-connected at its dimensional level of $\hat{q}$, the eccentricity of the simplex is equal to 0. Alternatively, a simplex with a non-zero eccentricity is eccentric from its surrounding simplices and therefore not fully embedded within the complex, i.e., the higher the eccentricity value, the more eccentric the simplex. The denominator $\check{q} + 1$ in Equation (9.2) is meant to attach great weight to the uniqueness in the lower dimension. However, this weighted dimensional does not necessarily apply to all phenomena (Atkin 1974). In the context of assessing imageability, it is reasonable to assume that a street's dimension, i.e., its degree value, can be significant. Hence, in this paper, we use eccentricity without considering a dimension's weight but only its depth along the dimensions of the simplicial complex, i.e.

$$Ecc'(\sigma_i) = \hat{q} - \check{q} \tag{9.3}$$

According to Equation (9.3), the eccentricity values of all the streets in the kite-shaped graph presented in Fig. 9.1, other than street 1, have eccentricity values greater than 0 (see Table 9.1) i.e., there is a difference between their $\hat{q}$ and $\hat{q}$ values. These eccentricity values indicate that they are unique since they are connected uniquely or exclusively to set of streets.

Thus, unlike graph theory's centrality measurement, a street acquires a high eccentricity value if it is different from every other street in the sense of structural distinctiveness or uniqueness rather than only in the sense of importance or function in the network's flows. This means that when we refer to an individual street in a street network from this perspective, the street will receive a high value if it connects to certain streets exclusively and there are many such streets. The consideration of structural distinctiveness, uniqueness or exclusiveness can be significant with respect to the possibility of a street being imageable since a street's high uniqueness means that it can be distinguished from others streets with respect to its connectivity in the street network. Concerning this, eccentricity has then, potentially at least, a surplus value relative to other topological centrality measures. The comparison between these two measurements regarding the kite-shaped graph presented in Table 9.1 illustrates that the eccentricity values only partly correlate to the centrality measures' values. Eccentricity thus has the potential to offer a different view compared to the graph theory's centrality measures. The following empirical study of the Tel Aviv street network attempts to evaluate this potential.

## 9.3 The Tel Aviv Street Network: Structural Qualities and Imageability

The study of the Tel Aviv street network qualities and imageability was conducted in two stages. First, we constructed an aggregate urban image by means of a survey conducted among 32 Tel Aviv residents. These residents were asked to "draw a map of Tel Aviv and its dominant elements (no more than 15 elements are to be included)". An aggregate map was then constructed based on the street elements in those individual images. This aggregate urban image included the 35 streets mentioned in common by the residents. Then, the imageability level of each given street was determined by the frequency of its appearance in the aggregate urban image, that is, the number of sketch maps in which the street appeared. In the second stage, we used the geographic data of the street network of Tel Aviv (a total of 1,767 streets). From these streets, we choose the section in which all the 35 imageable streets are included. By applying this selection method, the 195 main Tel Aviv streets in the network were chosen. These streets are shown in Fig. 9.2, together with their imageability level.

Based on the collected data, we calculated the correlation between imageability and the streets' centrality values as computed by Q-analysis and graph


**Fig. 9.2** Imageability level of Tel-Aviv's main streets

theory measurement. A significant correlation was found between imageability and eccentricity as defined in Equation (9.3): r = 0.695; $R^2$ = 0.483, ($p <$ 0.0001). The correlation of imageability with degree was r = 0.665; $R^2$ = 0.443, ($p < 0.0001$), with closeness it was r = 0.398; $R^2$ = 0.158, ($p < 0.05$), while the correlation for betweenness was r = 0.417; $R^2$ = 0.174 ($p < 0.05$). This analysis indicated that the imageability level was affected mainly by the topological properties of eccentricity and degree centrality, and much less by closeness and betweennes centrality. We should note that the former properties reflect the topological centrality of the street from the local perspective, whereas the latter does so from a global perspective. The correlation of imageability with eccentricity and degree is illustrated in Fig. 9.3. We then used multiple regressions to estimate how the combination of eccentricity and degree variables influenced the level of imageability. In this analysis, the correlation becomes r = 0.736 and $R^2$ = 0.541. This finding is highly significant, $p < 0.0001$. The correlation between these topological properties and imageability is illustrated in Fig. 9.3. A comparison between the spatial patterns of these topological properties (Fig. 9.4) illustrates their similarity with the spatial pattern of imageability in the Tel Aviv street network (Fig. 9.2).

The correlation between these topological properties and imageability can be related to the concept of *cognitive skeleton* (Kuipers et al. 2003). According this concept there is small set of streets in a city that provide the structure or mental frame of reference in the overall cognitive representation. This set of streets, which often correlates with the set of imagable elements and is used with a greater frequency, basically form a topological and hierarchical organization (Kuipers 2000; Stevens and Coupe 1978). Based on this view, several studies

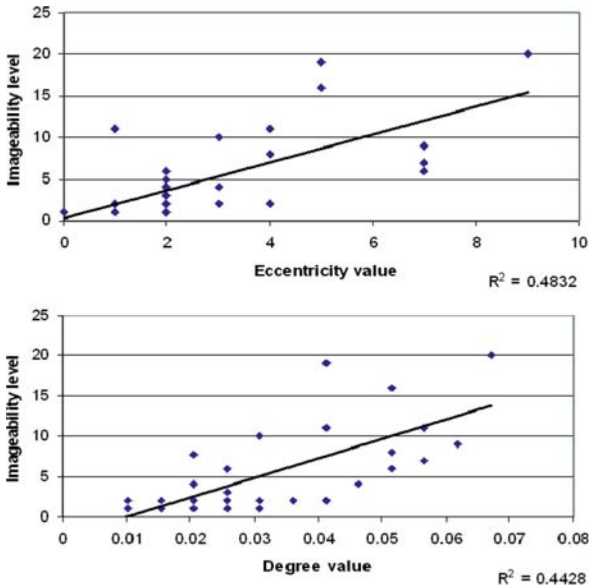



**Fig. 9.3** The correlation of imageability with the eccentricity and degree levels of Tel-Aviv's main streets

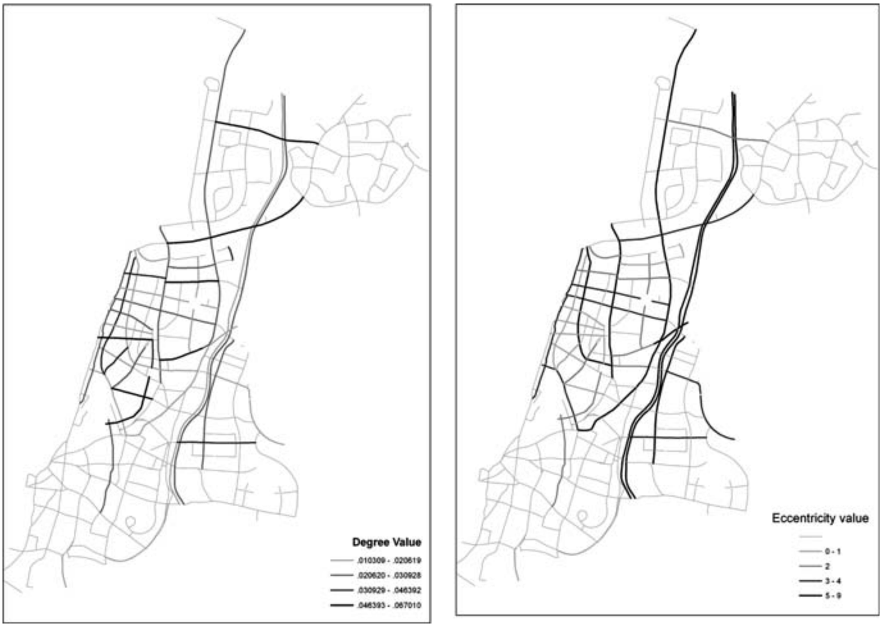



**Fig. 9.4** Degree (*left*) and Eccentricity (*right*) values of Tel-Aviv's main streets

have also suggested that a power-law-like distribution (or scale-free property) of the urban streets provide an indication for the existence of a hierarchy of streets in the network that form the cognitive skeleton (e.g. Porta et al. 2006; Jiang 2007; Tomko et al. 2008).

Examination of the distribution of the measures shows that the eccentricity and degree distribution fit significantly better with the power law distribution, $R^2 = 0.9043$ and $R^2 = 0.8443$ respectively (see Fig. 9.5), relative to the closeness

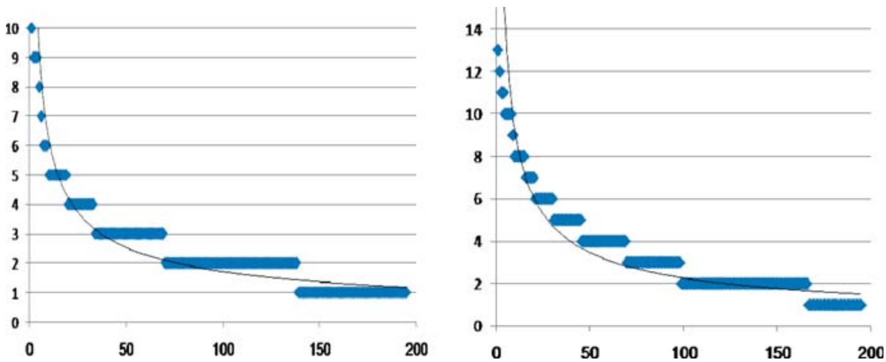


**Fig. 9.5** Power law distribution of named street eccentricity (*left*) and degree (*right*) values in the Tel Aviv street network

and betweeness distribution where $R^2 = 0.223$ and $R^2 = 0.69$, respectively. It is worth noting that the imagability level also shows power-law-like distribution $R^2 = 0.880$.

Further examination of the image of Tel Aviv's street network, based on the Q-analysis of the aggregate image of the city $K_I(I; \lambda)$ that presents in Table 9.2, provides some deep insight into the relation between imageability and structural qualities of streets. Comparison of this topological structure to the topological structure of street network indicates that there is a certain similarity between them. For example, we can take a look at the streets at dimension $q = 5$ of $K_I(I; \lambda)$ in Table 9.2, where the most nine imageable streets, which actually constitute the skeleton of the image or the cognitive skeleton of Tel Aviv city, are gathered into one connected component. Five of them belong to the same equivalence class in $K_s(s; \lambda)$ (at dimension $q = 2$).

**Table 9.2** Q-analysis of the image of Tel Aviv's street network; $K_I(I; \lambda)$ ($q \geq 4$)

| | |
|---|---|
| q = 5: | {Even Gvirol, Ben Yehuda, Dizengoff, Derech Namir, Hayarkon, Ayalon, Allenby, Arlozorov, Nordau} |
| q = 6: | {Even Gvirol, Ben Yehuda, Dizengoff, Hayarkon, Ayalon, Allenby} {Derech Namir}{Arlozorov} |
| q = 7: | {Even Ggvirol, Ben yhoda, Dizingoff, Hayrkon, Ayalon, Alenbi,} {Derech Namir }. {Arlozerov} |
| q = 8: | {Even Gvirol, Ben Yhoda, Dizingoff, Hayrkon, Ayalon} { Alenbi} {Derech Namir} |
| q = 9: | {Even Gvirol, Ben Yhoda, Dizingoff, Hayrkon, Ayalon} {Alenbi} |
| q = 10: | {Even Gvirol, Dizingoff, Ayalon} {Hayrkon} {Ben Yhoda} |
| q = 11: | {Even Gvirol , Ayalon} {Hayrkon} {Dizingoff} |
| q = 11: | |
| q = 15: | {Even Gvirol} {Ayalon} {Dizingoff} |
| q =14: | |
| q = 18: | {Even Gvirol} {Ayalon} |
| q = 19: | {Even Gvirol} |

It is interesting in this respect to note that strongly connected imageable streets also received high eccentricity values: Hayarkon, Dizengoff, Ibn Gvirol, Derech Namir and Arlozorov received eccentricity values greater than 4.0 ($Ecc' \geq 4$). This similarity implies links between the structure of tangible geographic connections of streets and the intangible conceptual connections of these same streets in the image of the city. These findings provide preliminary evidence of the relevance of eccentricity for understanding the relationship between structural qualities and imageability regarding the street network.

## 9.4 Conclusion

This paper explored the relationship between the topology of an urban street network and the image of the city by considering the structural distinctiveness of a given street in the overall street network. For this purpose, we suggested applying the structural property of eccentricity based on the perspective of multidimensional chains of connectivity.

The explicit consideration of the role of distinctiveness in a street network has the potential to improve the urban imageabilty study mainly through two ways. First, by complementing the graph theory' centrality measurement that provides a structural view on the state of individual streets, but with no consideration to the role of differentiated spatial elements. Second, by diminishing the gap between this structural view and the Lynchian view, which concentrates on the visual distinctiveness of urban elements.

The empirical study conducted using the case of Tel Aviv street network demonstrated how the structural property of eccentricity, as well as the overall perspective of multidimensional chains of connectivity, provides new insight into the study of the structural relationships between a city's street network and its image in our mind. Nonetheless, we wish to stress that these findings are based on one city only, and thus it is very methodological in nature. Hence, further study is needed to evaluate the contribution of this approach to networks and cities different in size, character and topology.

**Acknowledgments** Earlier versions of this paper were presented at the 11th AGILE International Conference on Geographic Information Science 2008, University of Girona, Spain. The authors are grateful to Hong Zhang, who helped design the sample street network in the paper and to Oz Shalom for his help in preparing the maps.

## References

Abu-Ghazzeh, T. (1997). Signs, advertising and the imageability of buildings: a perceptual selection in the view from the street in Amman, Jordan. *Habitat International,* 21(2), 255–267

Abu-Obeid, N. (1998). Abstract and sceneographic imagery: the effect of environmental form on wayfinding. *Journal of Environmental Psychology*, 18, 159–173

Atkin, R.H. (1974). *Mathematical structure in human affairs*. (London: Heinemann Educational Books)

Atkin, R.H. (1977). *Combinatorial connectivities in social systems*. (Basel: Birkhauser)

Claramunt, C. & Winter, S. (2007) Structural salience of elements of the city. *Environment and Planning B*, 34(6), 1030–1050

Conroy-Dalton, R. & Bafna, S. (2003). The syntactical image of the city: a reciprocal definition of spatial syntaxes. *4th International Space Syntax Symposium*, London.

Cubukcu, E. & Nasar, J.L. (2005). Relation of physical form to spatial knowledge in large scale virtual environments. *Environment and Behavior*, 37, 397–417

Freeman, L.C. (1979). Centrality in social networks: conceptual clarification. *Social Networks*, 1, 215–239

Garling, T., Book, A. & Lindberg, E. (1986). Spatial orientation and wayfinding in the designed environment: a conceptual analysis and some suggestions for post occupancy evaluation. *Journal of Architectural and Planning Research,* 3(1), 55–64

Haken, H. & Portugali, J. (2003). The face of the city is its information. *Journal of Environmental Psychology*, 23(4), 382–405

Haq, S. & Girotto, S. (2003). Ability and intelligibility: wayfinding and environmental cognition in the designed environment. *Proceedings of the 4th International Space Syntax Symposium*, London

Haq, S. & Zimring, C. (2001). Just down the road a piece: the development of topological knowledge in building layouts. In J. Peponis, J. Wineman & S. Bafna (Eds.), *Proceedings of the 3th International Symposium on Space Syntax* (pp. 48.41–48.17). Ann Arbor

Hillier, B. & Hanson, J. (1984). *The social logic of space*. (Cambridge: Cambridge University Press)

Jiang, B. (2007). A topological pattern of urban street networks: universality and peculiarity. *Physica A*, 384, 647–655

Jiang, B. & Claramunt, C. (2004). Topological analysis of urban street networks. *Environment and Planning B: Planning and Design*, 31, 151–162

Kim, Y. O. (2001). The role of spatial configuration in spatial cognition. In J. Peponis, J. Wineman & S. Bafna (Eds.), *International Symposium on Space Syntax* (pp. 49.41–49. 21), Atlanta

Krackhardt, D. (1990). Assessing the political landscape: structure, cognition, and power in organizations. *Administrative Science Quarterly*, 35(2), 342–369

Kuipers, B. J. (2000). The spatial semantic hierarchy. *Artificial Intelligence*, *119*, 191–233

Kuipers, B., Tecuci D., Stankiewicz, B. (2003). The skeleton in the cognitive map: a computational and empirical exploration. *Environment and Behavior,* 35, 81–106

Lynch, K. (1960). *The image of the city*. (Cambridge: MIT Press)

Lynch, K. (1990). *City sense and city design*. (Cambridge: MIT Press)

Omer, I. & Goldblatt, R. (2007). The implications of inter-visibility between landmarks on wayfinding performance: an investigation using a virtual urban environment. *Computers, Environment and Urban Systems*. 31(5), 520–534

Penn, A. (2003). Space syntax and spatial cognition or why the axial line? *Environment and Behavior*, 35(1), 30–65

Porta, P., Crucitti, P. & Latora V. (2006). The network analysis of urban streets: a primal approach. *Environment and Planning B: Planning and Design* 2006, 33, 705–725

Rosvall, M., Grönlund, A., Minnhagen, P. & Sneppen, K. (2005). Searchability of networks. *Physical Review E*. 72, 046117

Shokouhi, M. (2003). Legible cities: the role of visual clues and pathway configuration in legibility of cities. *Proceedings of 4th International Space Syntax Symposium*, London

Stevens, A. & Coupe, P. (1978). Distortions in judged spatial relation. *Cognitive Psychology*, 10, 422–437

Tomko, M., Winter, S. & Claramunt, C. (2008). Experiential hierarchies of streets. *Computers, Environment and Urban System*s, 32(1), 41–52

Weisman, J. (1981). Evaluation of architectural legibility: way-finding in the built environment. *Environment and Behavior*, 13, 189–204

Yun, Y.W. & Kim, Y.O. (2007). The effect of depth and distance in spatial cognition, *Proceedings 6th International Space Syntax Symposium*, Istanbul.

# Chapter 10
# A Spatial Analysis of Transportation Convenience in Beijing: Users' Perception Versus Objective Measurements

**Yongmei Lu, Weihong Yin, and Jing Chen**

**Abstract** This study investigates the convenience of urban transportation in Beijing municipality area. The analyses reported here consist of three parts – the spatial patterns of urban transportation Level of Service (LOS) perceived as transportation convenience by residents in Beijing, the spatial patterns of urban transportation LOS as reflected by objective transportation and traffic data, and the comparison of these two sets of patterns and measurements. The spatial unit of analyses was *Jie Dao* (街道), which is the basic civil administration unit in Chinese cities. This is one of the first studies linking subjective assessment of transportation convenience in Beijing with objective measurement; it reveals the displacement between the patterns of transprotation convenience reflected from the two aspects. Further examination are needed to explain the discrepancy between the perceived and measured transportation LOS in different *Jie Daos* in Beijing. The related investigation shall shed light on better understanding the effectiveness of different practices of urban transportation management, planning, and project investment in different types of neighborhoods in Beijing.

**Keywords** User perception · Transportation convenience · Beijing · *Jie Dao* district (街道)

## 10.1 Introduction

Urban transportation has always been an important topic for general public, city managers, and academics. Although large cities tend to have more roads to connect different parts within it (Lu and Tang 2004), the efficiency of their transportation systems is frequently a concern. This problem tends to be more

Y. Lu (✉)
Beijing Normal University, Beijing, China; Texas State University, San Marcos, TX, USA
e-mail: yl10@txstate.edu

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_10, 

severe for cities in developing countries, as revealed by literature (e.g. Sit 1996; Sheng 1997; Wang and Chai 2008). Transportation convenience refers to how easy it is to travel from one location to another. As many other convenience concepts, it is highly related to users' perception. Transportation convenience is related to not just road capacity, but also traffic situation and transportation service facilities. Therefore, the achievement of convenient transportation calls for the collaboration from engineering, planning, management, as well as public service.

This paper investigates transportation convenience within the munucipality area of Beijing. This was done from two perspectives – the quantification and analysis of citizens' perception of Beijing's transportation convenience, and the assessment of the city's transportation system based on objective data. These two aspects were investigated at the *Jie Dao* (街道) level (or translated as *Civil Administration District at Street Level*), which is the basic civil administration unit in Chinese cities. The spatial patterns of urban transportation Level of Service (LOS) is evaluated using both user perception survey data and objective transportation data. The spatial discrepancy between these two patterns are revealed and discussed.

## 10.2 Literature Review

### *10.2.1 Urban Transportation and City Livability*

The study of transportation convenience is an integral aspect for city livability (e.g. Lennard and Lennard 1995; Dang 2006; Zhang et al. 2006, Chen 2007) and quality of life (e.g. MSGUNCC 2008; NRC 2002; Shafer et al. 2000). In the western world, traffic-calming and bicylce planning movement in the 1960s and 1970s linked transportation management to livability in urban area; they promote controlling for traffic speed and cut-through traffic in neighborhoods. Starting from the Netherlands, this movement quickly spread to other countries in Europe, North America, Australia, and the developed countries in Asia such as Japan. The bi-annual conference of "International Making City Livable" was launched in 1985. This conference series has played an important role in connecting urban transportation studies directly to the improvement of city livability. In 1998, the Clinton-Gore administration of the US launched the *Clinton-Gore Livability Agenda*. This agenda puts billions of dollars toward building "livable communities" for the twenty first century, which include: *Easing traffic congestion by improving road planning, strengthening existing transportation systems, and expanding use of alternative transportation*. A similar approach was also included in the US President's 2000 Fiscal Year Livability Initiatives.

With "livable city" listed as one of the four major characteristics (or goals to pursue) for Beijing manucipality area, *The Comprehensive Urban Planning of Beijing, 2004–2020* was approved by the State Council in 2005. Among all the important factors and aspects related to the livability of Beijing, urban transportation is identified as one of the most problematic components. Beijing was ranked number 15 in 2005 for its livability among the major cities in China, according to a report by Horizon-China and *Business Week* (No Author 2005). Awful urban transportation was believed to be the most significant factor for this poor ranking for Beijing (Cheng 2006). Beijing International Institute of Urban Development Research conducted a study in 2006 on the quality of life in Chinese cities; Beijing was ranked number 14 among 287 cities (Nian 2006); Beijing's transportation was ranked as the last one among all the cities when considering people's satistifaction with the system. Therefore, the problem posed by poor urban transportation in Beijing is significant and calls for immediate attention and effective responses from all groups invovled, including government, general public, as well researchers.

### 10.2.2 UrbanTransportation Studies in China

In China, an abundance of urban transportation studies have appeared since 1970s. Most of them are about specific transportation topics in urban areas. This can be seen from the recent publications, including those on topics such as facilitating the development of functional zones through planning of urban tranportation network to (e.g. Ye and Chen 2006), optimization of the spatial separation between urban roads (Cai 2006), models and algorithms for intelligent transportation systems (e.g. Si and Gao 2006), tranportation congestion at street intersetions (Li and Hu 2006), improvement of public transit systems (Quang and Sun 2006), and urban transportation safety (Wang and Zhang 2006). In addition to topical investigation, there are comprehensive studies to evaluate urban transportation systems (e.g. Shi and Wang 2006; Wang and Wang 2006). Recent years also saw studies linking urban tranportation with urban landuse and urban functional zones (e.g. Yi 2006a, b).

From a methodological perspective, urban tranportation researchers in China have noticed the potential of GIS. Urban streets are frequently conceptualized as linear features; location and orientation of streets are closely related to the geographical condition of a place; topological information is integral to databases of street network. It is a clear trend in urban transportation literature in China that GIS and related technologies are acting as important technological supports (e.g. Zhang and Zhu 2005; Fang 2006).

However, there is a lack of studies that account for urban residents' opinion in China. As a matter of fact, this gap in urban transportation literature could be world-wide, as recognized by some scholars in 1990s (Madanat et al. 1994; Pfefer 1999). Despite some recent tranportantion studies that incorporate user' perception (e.g. Pecheux et al. 2000; Zhang and Prevedouros 2004), the need to analyze and understand travelers' perception of urban tranportation system is very clear and increasing (Lee et al. 2007).

This paper aims at adding to the literature of urban transportation in general, and that of Chinese cities in particular by analyzing transportation convenience as perceived by urban residents in Beijing. This was done through examining their opinions on the transportation asepct of urban livability in Beijing. Residents' perception of transportation convenicence is then linked to a series of objective urban transportation measurments that are based on GIS and transportation engineering data.

## 10.3 Data Used for This Study

The study area includes the eight urban districts in Beijing (namely Dong Cheng District, Xi Cheng District, Chong Wen District, Xuan Wu District, Chao Yang District, Hai Dian District, Feng Tai District, and Shi Jing Shan District) and five special urban development zones and new towns (namely Hui Long Guan, Tian Tong Yuan, Yi Zhuang New Town, Tong Zhou New Town, and Da Xing Huang Cun). The analysis units are *Jie Dao*. There are a total of 134 *Jie Dao* districts included in this study (see Fig. 10.1) with population size in each *Jie Dao* ranging roughly from 12,500 to 137,000 and areal size between 1 and

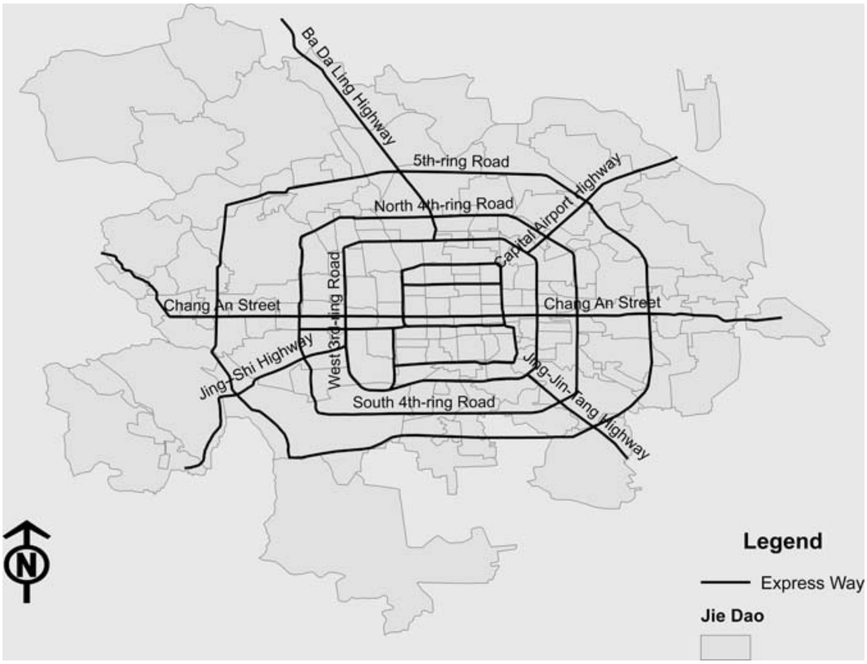



**Fig. 10.1** The study area of Beijing Municipality and the 134 *Jie Dao* Districts included for this study

124 $km^2$. Detailed information about the 134 *Jie Dao* districts is not reported in this paper but is available upon request.

### 10.3.1 Residents' Perception of Transportation Convenience

The data on urban residents' perception of transportation convenience were obtained through recoding the responses to a number of questions from a survey titled *the Questionnaire on Assessment of and Preference for Urban Residential Environment*, which was conducted in 2005 with a grant support by Beijing Municipal Natural Science Foundation. To ensure that the data reflect the perception of urban transportation by regular residents, the respondents are limited to those who have lived in Beijing for a continuous period of at least six months by the time of the survey. In addition, a series of actions were taken to control for the representativenees of samples. Such measurements include stratified sampling in residential communities and *Jie Dao*s based on population size, equal-distance sampling according to the distribution of neighborhoods and communities along streets, sampling by controling for gender and age to reflect population structure of *Jie Dao* districts, etc.. More information about this survery project can be found in Zhang et al. (2006). The data used by this study were derived from answers to five questions in Section Q9 of the questionnarie, which is titled "Assessment of Tranportation Convenience". These questions were designed to measure urban residents' perception of transportation convenience from five asepcts, namely, *use of public transit system*, *traffic congestion*, *convenience for job and school commuting*, *convenience for runing daily errands*, and *convenience for trips to the City Center*. The respondents were asked to indicate their satisfaction to these aspects on a five-point Likert scale ranging from *very satisfied* to *very unsatisfied*. The option of *not familiar with the situation* was available for those respondents who chose not to respond to the five-point scale. Among the 11,000 or so questionnaries distributed, 9,112 were returned to the field staff, and 7,647 were determined to be valid and were included for this study.

### 10.3.2 Objective Measurements for Beijing's Urban Transportation

Data for objective measurements of Beijing's urban transportation were derived from various datasets to reflect three aspects: *urban road capacity*, *urban public transit service*, and *urban traffic flow situation*. Road capacity data were obtained from "Digital Transportation Maps of Beijing" project. This project was conducted by Beijing Livable City Research Digital Platform with support from the Center for Science of Human Settlement of Beijing Union University (http://cshs.homebj.com). Information about the type and length of each street segment were extracted from this source. Urban public transit LOS was evaluated using GIS data of the spatial locations of bus stops and subway stations. These data were extracted from "Digital Transportation Maps of Beijing" project and Beijing Subway Network (http://-www.bjsubway.com/cns/index.html). The urban public transit service data were updated to the beginning of year 2006, which is close to the time of the residents' perception survey on transportation convenience. It is important to know that Beijing has made significant improvement in its public transit system and the overall transportation facilities due to the 2008 Summer Olympics Games. Traffic flow data were derived from the records of the Control Center of Beijing Bureau of Transportation Administration. The data are based on traffic congestion reports, including the records of call-for-services to "122" (a phone number designated for call-for-services to police officers for help with urban transportation problems), field reports, and traffic congestion reports from the City Transportation Administration Office. The related records for year 2005 were cleaned, analyzed, and validated to identify 59 frequently congested locations in Beijing (http://www.qianlong.com). These traffic congestion location data provide a foundation for objective evaluation of city-wide traffic congestion. The residents' transportation perception data and objective measurement data were (dis)aggregated to *Jie Dao* level for spatial pattern analyses.

## 10.4 Methodology and Procedure

### *10.4.1 Subjective Scores for Urban Transportation Convenience*

The perception of transportation convenienece for a *Jie Dao* in Beijing was quantified as a comprehensive score that was derived through a three-step approach. The first step was to recode the residents' responses to survey questions from text information to standardized ratio scores. The second step was to calculate separate scores for five individual factors for each *Jie Dao* district. The the last step was to develop a weighting schema for the five factors and to combine the factor scores into a comprehensive urban transportation convenience score for each *Jie Dao*.

*Recoding text information into scores:* The five-point scale options to reflect residents' degree of satisfaction include "very satisfied", "satisfied", "ok", "unsatisfied", and "very unsatisfied". The scores assigned to the different options are 100, 80, 60, 30, and 0 respectively. The reason for choosing the score range between 0 and 100 is to allow for more room to reveal scoring differences when multiple scores are combined later. The two scores for the two extreme situation are set to keep this score range. The score of 60 is assigned to the response of "ok" as it is a common practice in various fields in China that a 60-point grade is assigned for a "pass". The rest of the two scores are encoded considering that "satisfied" is inbetween "very satisfied" and "ok" and that "unsatisfied" is inbetween "ok" and "very unsatisfied".

*Calculating factor scores* This step derives quantitative socres in all five aspects of transportation convenience for each *Jie Dao*. Equation (10.1) explains how this was calcualted –

$$Q_{i,j} = (C_{i,j}{}^{*}100 + D_{i,j}{}^{*}80 + E_{i,j}{}^{*}60 + F_{i,j}{}^{*}30 + G_{i,j}{}^{*}0)/(T_{i,j} - U_{i,j} - M_{i,j}) \quad (10.1)$$

where $Q_{i,j}$ indicates the score *Jie Dao i* received for factor *j*, $C_{i,j}$, $D_{i,j}$, $E_{i,j}$, $F_{i,j}$, and $G_{i,j}$ represent the numbers of responses choosing "very satisfied", "satiefied", "ok", "unsatisfied" and "very unsatisfied" respetively for factor *j* by residents in *Jie Dao i*, $T_{i,j}$ represents the total responses for factor *j* from *Jie Dao i*, $U_{i,j}$ represents the number of responses choosing "not familiar with the situation" for factor *j* from *Jie Dao i*, and $M_{i,j}$ indicates the number of missing answers to factor *j* from *Jie Dao i*. To make the scores for different factors comaprable city-wide, the individual scores for each *Jie Dao* was then standardized for each of the five factors following Equation (10.2) –

$$Q'_{i,j} = \frac{Q_{i,j} - Q_{\min,j}}{Q_{\max,j} - Q_{\min,j}} \quad (10.2)$$

where $Q'_{i,j}$ is the standardized score of *Jie Dao i* for factor *j*, $Q_{\min,j}$ is the lowest score all *Jie Dao*s received for factor *j*, and $Q_{\max,j}$ is the highest score all *Jie Dao* received for factor *j*. The result of this step was a 134 by 5 matrix indicating the standardized factor scores each *Jie Dao* received, reflecting the residents' degree of satisfaction for each of the five investigated aspects of tranportation convenience.

*Deriving a comprehensive score:* The contribution to the overall urban transportation convenience by each of the five factors differs. To combine the five factor scores into a comprehensive indicator, we needed to reach an agreement regarding how important each factor is as relative to the other four. *Delphi*-technique was used to derive a weighting scheme that reflects the collective expertise from a group of 12 experts. The expert group consists of normal urban residents, researchers with specilization in urban transprotation, as well as vocational or volunteering field staff who work to help manage daily traffic control in the streets. The five factors (as listed in Table 10.1) are weighted pairwise by each of the experts. Any conflict that presented after combining the weights were reported back to the expert group and a re-assessment of the related factors were conducted. After serveral runs of adjustment by the experts, the combined weighting matrix reached an acceptable level of consistency. The final weights for the five factors are listed in Table 10.1. Note that the last three factors are about different aspects of "movability" for the residents. They are collectively referred to as *movability* in the rest of this paper. The overall perceived urban transportation convenience for each *Jie Dao* was then calculated as –

**Table 10.1** Weights of transportation factors for overall transportation convenience

| Factor | | Description | Weights |
|---|---|---|---|
| Public transit | | The easiness of using public transit | 0.22 |
| Traffic congestion | | The smoothness of traffic (no congestion) | 0.25 |
| movability | Commuting | Job or school related commuting trips | 0.34 |
| | Daily-life trip | Trips related to daily-life errands | 0.17 |
| | Trip to city center | Travel to and from the center of the City | 0.02 |

$$Q_i = \sum_{j=1}^{5} (w_j{}^* Q'_{i,j}) \tag{10.3}$$

where $Q_i$ is the overal score for the perceived urban transportation convenience for *Jie Dao i*, and $w_j$ is the weight assigned to factor *j*.

### 10.4.2 Objective Scores for Urban Transportation Convenience

For the objective measurement of urban transportation convenience in the different *Jie Dao*s in Beijing, three aspects were assessed using data of urban streets, public transit facilities, and traffic flow situations.

*Urban road capacity:* The length of each street segment was derived using GIS and was aggregated for each *Jie Dao* based on the street types. Considering the difference in service capacity of different types of urban streets, a weighting system was used to calculate the weighted street length density for each *Jie Dao*, as explained in Equation (10.4) below –

$$RC_i = (L_{i,h}{}^*4 + L_{i,1}{}^*4 + L_{i,2}{}^*3 + L_{i,3}{}^*2 + L_{i,4}{}^*1 + L_{i,s}{}^*1)/A_i \tag{10.4}$$

where $RC_i$ represents road capacity in *Jie Dao i*, $A_i$ represents the areal size of *Jie Dao i*, $L_{i,h}$ represents the length of highway in *Jie Dao i*, $L_{i,s}$ indicates the length of suburban roads in *Jie Dao i*, $L_{i,1}$ indicates the length of first class roads in *Jie Dao i*, and so on and so forth. Similar to the standardization of the subjective scores, the road capacity level measured by $RC_i$ was also standardized to derive $RC'_i$ for *Jie Dao i*.

*Urban public transit service:* The geographic location information for bus stations and subway stops were overlaid atop *Jie Dao* districts map. Each *Jie Dao* receives one point for each bus station or ordinary subway stop that fully resides in it. When a bus station or a subway stop locates at the boundary of multiple *Jie Dao*s, the one point is divided equally among the *Jie Dao* districts that share this station or stop. Since the subway transfer stations have much higher level of service capacity and do serve more people, two points were assigned to each of the subway transfer station, which were split equally among *Jie Dao*s if it is located on a boundary. As a result, each *Jie Dao* in the study area was assigned a score, $PT_i$, which was further standardized into $PT_i^{'}$ to represent the public transit service level in it.

*Urban traffic flow:* The technical approach was similar to that for public transit scoring. Each of the frequent traffic congestion locations, as explained in the data section, was mapped atop *Jie Dao* districts so that GIS operations can be conducted to determine the topological relationship between the congestion places and *Jie Dao*s. The congestion score for a *Jie Dao* is based on the average congestion time at the related frequently congested places. A recorded "congestion time" is considered to be fully within a *Jie Dao* if the congestion place is geographically within the *Jie Dao*; otherwise, the "congestion time" is equally divided among the *Jie Dao*s that are touched by a congestion place. The score each *Jie Dao* received is the total "congestion time" assigned to it from all the frequently congested places and is recorded as $TC_i$ . Similarly, $TC_i$ was standardized into $TC_i^{'}$. Note that a high traffic congestion score indicates low levels of transportation convenience; to make this score consistent with the other scores, Equation (10.5) was used to convert traffic congestion scores into traffic flow score –

$$TF_i' = 1 - TC_i' \tag{10.5}$$

where $TF_i^{'}$ is the standardized traffic flow score for *Jie Dao i.*

*Overall objective scores for urban transportation:* The overall objective score for transportation convenience for each *Jie Dao* district is a weighted sum of the three scores described above. The three objective aspects measured in this study (i.e. road capacity, public transit service, and traffic flow) are considered to be corresponding to "movability", "public transit", and "traffic congestion", which are the three aspects surveryed for residents' perception (note that three of the original five factors are collectively referred to as "movability" here). Therefore, the weighting scales for calculating the perceived transportation convenience are applied to derive the objective overall score of urban transportation service. The three objective factors discussed above receive a weight of 0.53, 0.22. 0.25 respectively.

### *10.4.3 Link the Perceived Transportation Convenience with the Measured*

The assessment based on perception of urban trasnprotation was lined with the evaluation based on objective data and measurements. This was done through two approaches. First, *Wilcoxon Signed-rank Tests* were conducted to compare the subjective scores with the objective ones. The purpose was to test for the null hypotheses that the corresponding two sets of scores are statistically the same, i.e. there is no significant difference between the residents' perceived transportation convenience in Beijing and the transportation LOS reflected by objective data. Were the hypothesis rejected, the second approach group the 134 *Jie Dao*s using *K-means Cluster* technique based on the two sets of scores measuring transportation convenience. The purpose was to reveal the patterns of statistical and spatial difference between the perceived scores and the objective scores. After the *Jie Dao*s were classified, *Kruskal-Wallis tests* were applied to examine the among-group difference for the comprehensive scores and the factor scores to validate the *K-means cluster* results. Furthermore, a series of *Mann-Whitney* tests were conducted to investigate the between-group differences. GIS maps were also used to visualize the distribution of the different types of *Jie Dao* districts and to link the spatial patterns with other urban characteristics.

## 10.5 Analysis of Results

Following the methods explained above, the scores for perceived transportation convenience in Beijing were derived. Table 10.2 reports factor scores and overall scores for a list of exemplar *Jie Dao*s. The complete list of scores is available from the authors upon request. The spatial distribution of *Jie Dao*s and their respective overall scores derived from the survey are illustrated in Fig. 10.2. Figure 10.3 and Table 10.3 report the results of transportation convenience assessment based on objective data. It can be seen from Fig. 10.2 that the overall high subjective scores for transportation convenience are concentrated in the west part of Beijing, especially the section highlighted by the oval. It consists of such *Jie Dao*s as Si Ji Qing, Xiang Shan, and Wan Liu Di Qu. They are in Hai Dian District, which is the major education and hightech industry functional zone in Beijing. In addition, there are some *Jie Dao*s spotted outside of the 5th-ring Roads that are determined by the pubic to be good in transportation convenience. On the other hand, two areas were identified by the public as being most inconvenient for transportation. One is in north Beijing – the notorious "Tian Tong Yuan – Hui Long Guan" area. The other is in Chao Yang District; it cuts through the southest corner of the city starting from inside the 2nd-ring Road and extending to outside the East 5th-ring Road. This sector includes *Jie Dao*s such as Long Tan, Pan Jia Yuan, Shi Ba Li Dian, Nan Mo Fang, and Dou Ge Zhuang etc.. These two areas are highlighted by ovals on Fig. 10.2. Moreover, examining the "ring road system" of Beijing's transportation network, two "poor transportation rings" exist. The inner ring is located between the 3rd-ring and the 4th-ring Roads; the outer ring is right outside the 5th-ring Road in the north and between the 4th-ring and the 5th-ring Roads in the southwest– south–southeast directions. It is especially bad in the zone between the extension of East Chang An Street (Jian Guo Road – Jing Tong Express Way) and Jing Jin Tang Highway to the southeast.

Figure 10.3 illustrates the objective scores for transportation LOS in Beijing. A distance-decay pattern away from the city center is evident, meaning that the *Jie Dao*s at or close to the center of Beijing tend to show relatively high objective scores for urban transportation. The *Jie Dao*s inside the 3rd-ring Roads

**Table 10.2** Scores for perceived transportation convenience in *Jie Dao* districts in Beijing

| ID | Public transit (original) | Public transit (standard-ized) | Traffic congestion (original) | Traffic congestion (standard-ized) | Urban movability (original) | Urban movability (standard-ized) | Overall transportation convenience |
|---|---|---|---|---|---|---|---|
| 1 | 75.63 | 0.71 | 59.38 | 0.57 | 66.68 | 0.76 | 0.70 |
| 2 | 65.00 | 0.51 | 55.00 | 0.47 | 59.70 | 0.62 | 0.56 |
| 3 | 73.02 | 0.66 | 63.95 | 0.68 | 65.12 | 0.73 | 0.70 |
| … | … | … | … | … | … | … | … |
| 134 | 67.46 | 0.56 | 56.35 | 0.50 | 61.14 | 0.65 | 0.59 |
| 135 | 74.04 | 0.68 | 59.81 | 0.58 | 62.48 | 0.68 | 0.66 |
| 136 | 60.83 | 0.43 | 42.89 | 0.19 | 53.42 | 0.49 | 0.40 |
| Maxi-mum | 90.94 | | 78.15 | | 78.26 | | |
| Mini-mum | 38.00 | | 34.44 | | 29.52 | | |

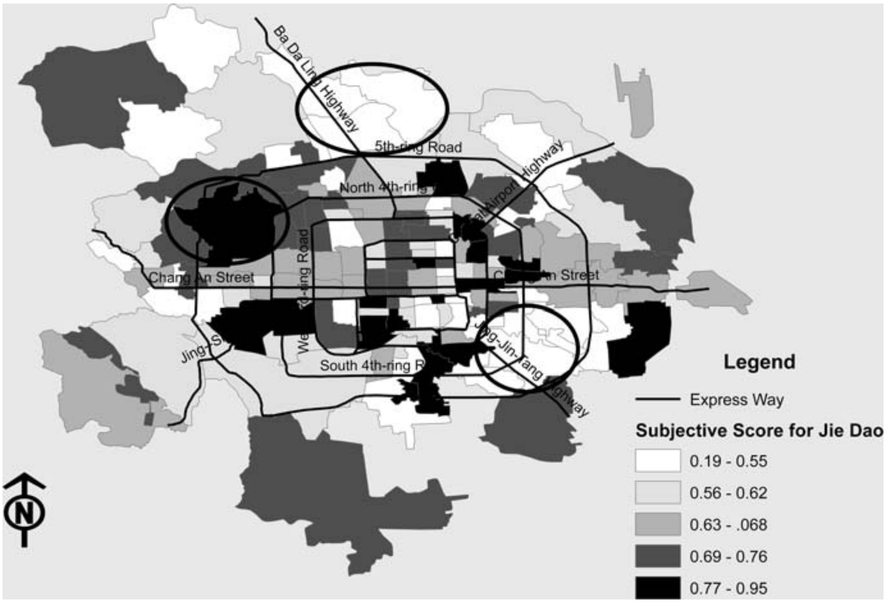



**Fig. 10.2** The perceived transportation convenience at *Jie Dao* level in Beijing

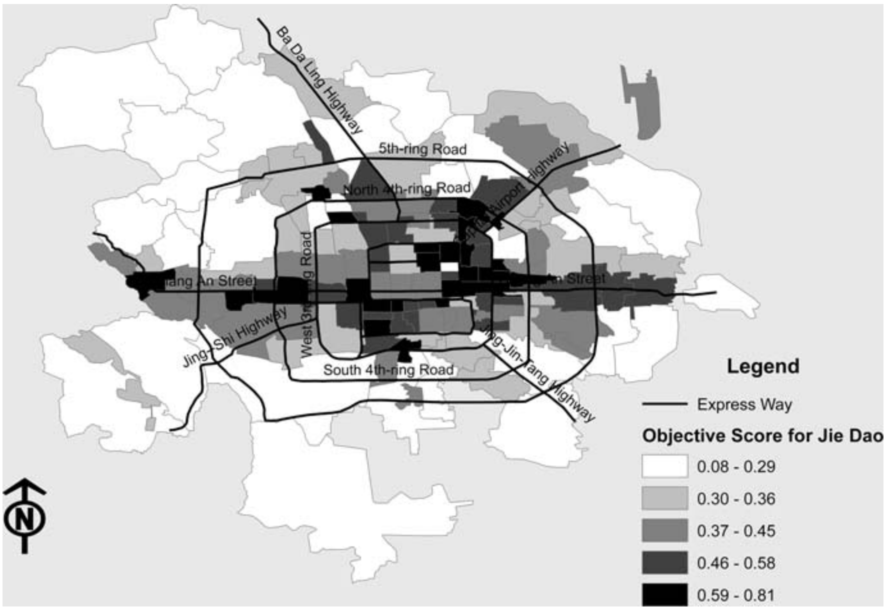



**Fig. 10.3** The objective assessment of transportation convenience in Beijing at *Jie Dao* level

**Table 10.3** Scores (standardized) for objective measurements of transportation convenience in the *Jie Dao* districts in Beijing

| ID | Public transit | Road capacity | Traffic congestion | Overall transportation convenience |
|---|---|---|---|---|
| 1 | 0.11 | 0.29 | 0.17 | 0.19 |
| 2 | 0.08 | 0.24 | 0.16 | 0.16 |
| 3 | 0.10 | 0.28 | 0.19 | 0.19 |
| ... | ... | ... | ... | ... |
| 134 | 0.09 | 0.25 | 0.14 | 0.16 |
| 135 | 0.11 | 0.35 | 0.18 | 0.21 |
| 136 | 0.07 | 0.23 | 0.10 | 0.14 |

received the highest objective scores. The scores for those between the 3rd- and the 5th-ring Roads are not as high, while most *Jie Dao*s outside of the 5th-ring Roads received low objective scores. This pattern might reflect the urban development process of Beijing – the urban built sections as well as roads and other infrastructures were constructed following a center-outwards order. However, it is interesting to notice that the *Jie Dao*s along the East 3$^{rd}$-ring Road and along the extension of the East Chang An Street were scored relatively high based on objective transportaion and traffic data but were judged poor by the residents for transportation convenience. These areas have good transportation infrastructure, but lack convenient urban transportation. This might indicate the presence of problems related to transportation design and management, which call for researchers and urban transportation management and planning departments to further examine these problems so as to achieve the full potential of the existing transportation facilities.

*Wilcoxon signed rank test* reveals that the subjective scores derived from residents' perception of urban transportation in Beijing are statistically different from the objective scores at 0.001 significance level. The difference is also significant when comparing the subjective scores with the objective ones for each of the three aspects investigated. This finding indicates that there is a significant gap between residents' perception of transportation convenience in Beijing and urban transportation LOS that is measured by urban roads, public transit, and other transportation facilities.

The *K-mean cluster analysis* classified the 134 *Jie Dao*s into four groups based on both subjective and objective scores. Table 10.4 summarizes the four groups of *Jie Dao*s. Each group is defined as "H", "L", or "M" for subjetive or objective scores by comparing the majority membering *Jie Dao*s'scores in that group with the median score city wide. For example, the first group is labeled as "LH" because 42 out of its 50 members have objective scores equal to or below the median and 43 members have above-median subjective scores. This way, the four groups are labeled as "LH", "HM", "LL", and "HL". *Kruskal–Wallis test* shows that the four groups are statistically different from each other at 0.001 significance level when considering both subjective and objective scores.

**Table 10.4** K-means clustering results of the 134 *Jie Dao* districts into four groups

| | | | | | |
|---|---|---|---|---|---|
| Group ID | | 1 | 2 | 3 | 4 |
| Group label (Obj./Subj.) | | LH | HM | LL | HL |
| Total number of *Jie Dao* districts | | 50 | 29 | 27 | 28 |
| Score for cluster center | Subjective scores | 73 | 70 | 49 | 58 |
| | Objective scores | 33 | 63 | 30 | 48 |
| Number of *Jie Dao*s with objective score | > Median score | 8 | 29 | 1 | 27 |
| | <= Median score | 42 | 0 | 26 | 1 |
| Number of *Jie Dao*s with subjective score | > Median score | 43 | 16 | 0 | 3 |
| | <= Median score | 7 | 13 | 27 | 25 |

**Table 10.5** Mann–Whitney tests for the score difference between any two of the four groups of *Jie Dao* districts (significance level)

| Significance level | Group1 (LH) | | Group 2 (HM) | | Group 3 (LL) | | Group 4 (HL) | |
|---|---|---|---|---|---|---|---|---|
| | Subj. scores | Obj. scores | Subj. scores | Obj. scores | Subj. scores | Obj. scores | Subj. scores | Obj. scores |
| Group1 (LH) | | | 0.067 | 0.000 | 0.000 | 0.158 | 0.000 | 0.000 |
| Group 2 (HM) | | | | | 0.000 | 0.000 | 0.000 | 0.000 |
| Group 3 (LL) | | | | | | | 0.000 | 0.000 |
| Group 4 (HL) | | | | | | | | |

Furthermore, *Mann–Whiteney tests* were conduted to test the score difference between any two groups and reveal that all, except the objective scores between group 1 (LH) and group 3 (LL), are significantly different at 90% or above (see Table 10.5), Thus, the classification of the 134 *Jie Dao*s into four groups are statistically sound.

Figure 10.4 shows the spatial distribution of these four types of *Jie Dao*s. It can be seen that the different types of *Jie Dao*s are clearly clustered. The districts with relatively low objective scores for urban transportation show either high or low subjective scores. Three zones of *Jie Dao*s with low objective but high subjective scores are identified as “LH” in Fig. 10.4 and Tables 10.4 and 10.5. The first of such zone is in the north part of the centrer of Beijing – between the North 2nd-ring Road and Chang An Street. The second zone forms a “ring” between the the 3rd-ring and the 5th-ring Roads in the northeast, the 3rd-ring and 5th-ring Roads in the west, and between the 2nd-ring Road and the 4th-ring Road in the south. The third “LH” zone is outside the 5th-ring Roads in the south, east, and northwest directions. The *Jie Dao*s with “LL” scores are dominant along and outside the 5th-ring Roads in the north-northeast, between the 3rd-ring and the 5th-ring Roads in the southeast and southwest, and outside of the 5th-ring Road in the southwest.

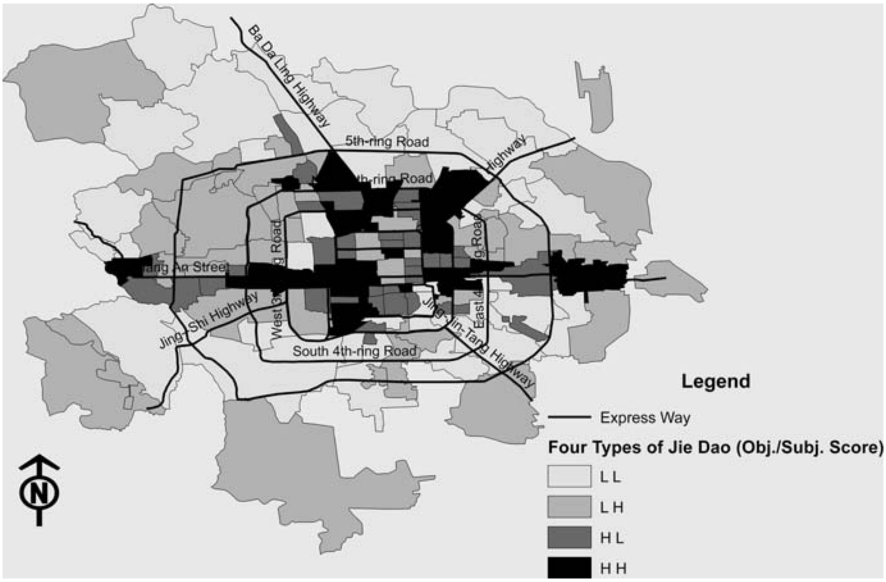



**Fig. 10.4** The results of k-mean clustering of the Jie Dao districts

The rest *Jie Dao*s have high objective scores for urban transportation. One group of them have low subjective scores (labeled as "HL"), while the other have medium subjective scores (labeled as "HM"). They together form two zones. The first is a ring inside the South 2nd-ring Road, between the North 3rd-ring Road and the North 4th-ring Road, and between the 2nd-ring and the 3rd-ring Roads in the east and west. The north–northwest section and southeast section of this ring signify "HL" *Jie Dao*s, i.e. high objective and low subjective scores. The second zone is a belt following Chang An Street and its extension. The *Jie Dao*s on the east part of this belt are mainly "HL" type. All *Jie Dao*s along the ring and belt score high for urban transportation LOS when considering road capacity, public transit infrastructure, and reported traffic flow situations; but they were all evaluated as around or below the median level by residents for their transportation convenience.

## 10.6 Discussion and Conclusions

Urban tranpostation, like many commodities consumed by urban residents, should be assessed from both supplier's perspective and user's perspective. Transportation LOS literature has been calling for efforts to incorporate user's percpetion since the 1980s (Ndoh and Ashford 1994). This study adds to the literature by accounting for urban residents' perception of transportation convenience when assessing trasnportation in a mega city in a developing country. Furthermore, it compares residents' perception of traditional objective measurements for transportation service based on capacity and availability. As an empirical study of urban transportation in Beijing, the findings of urban transportation LOS in different parts of the city as well as those about the spatial patterns of and differences between the objectively measured and subjectively perceived LOS have significant implications for urban transportation management and planning in Beijing.

Among the zones with very poor urban trasnportation, as identified by urban residents, the "Hui Long Guan – Tian Tong Yuan" area is well-known by almost everyone living in Beijing or anyone who has traveled there. These two *Jie Dao* districts mainly consists of recently built communities that were designed to function as a bed-room town for the working class of Beijing in the late 1990s. These two *Jie Dao*s grew from rural communites into urban residential town almost overnight. However, without a quality plan, these communites were built to house 100,000 residents without being able to provide them with other community and basic daily life services. Public transit service, community schools, daycare centers, community shopping centers, and entertainment faciltites are lacking. On top of the list, majority of the young working classes who purchased home there work at places inside Beijing's traditional urban districts. With the poor urban transportation infrastructure in this area, as indicated by low objective transportation scores (Fig. 10.3), the supply-demand conflict for urban transportation is very prominent. Although the Beijing government has realized this problem in recent years, the analysis of this study indicates insufficient effort or at least ineffective measurement by the government in helping people in these two *Jie Dao*s.

The objective assessment of Beijing's transportation LOS was based on road capacity, public transit service, and reported traffic congestion. It reveals that transportation condition in Beijing shows a clear distance-decay pattern starting from the city center. This pattern relfects the practice of urban transportation development of Beijing in its long history – single-centered city with most of the resources concentrating at the center of the city. This can be very problematic for a megacity like Beijing. With a 16 million population residing in the eight urban districts, it is bad planning and urban contruction practice to develop urban transportation systems from the center outwards.

It is an important contribution of this study to link the subjective scores for urban transportation LOS with objective scores. The difference between these two scores speaks to the potential problems existing in Beijing's transportation. For example, the *Jie Dao*s with high objective scores but low or medium subjective scores form a ring between the North 4th-ring Road and the South 3rd-ring Road and a west-east belt along Chang An Street and its extension. It is necessary for urban transportation planners and managers to understand what facilitates the formation of these two zones, as they are where urban transportation resources might be wasted or at least not used to full potential. They call for focused efforts aiming at (1) maximizing the usage of existing transportation infrastructure to improve supply of service, and (2) de-centralizing some urban functions and services to reduce commuting and other trip demands. Meanwhile, the "LH" *Jie Dao*s warrant in-depth investigation as well, for understanding the factors that make the residents satisfied with urban transportation in these areas, especially when the transportation infrastructure is not so good based on objective scores, should provide significant insights for the city to improve transportation LOS to fullfill residents' travel demands.

Nevertheless, the analyses presented in this paper are not without limitations. First of all, the subjective scores of urban transportation for each *Jie Dao* are derived based on the perception of residents living in the *Jie Dao*. The assessment might be biased towards the perception of local people and have ignored the perception of people traveling to or through a *Jie Dao*. This might have masked some problematic areas in Beijing that are incovenient for non-local residents to travel to or through. For example, *Zhong Guan Cun Jie Dao* is one of the most congested areas during traffic peak time but its subjective score is not low. The reason might be that the residents of *Zhong Guan Cun* do not necessarily travel in congested major streets during peak time and those who are frequently stuck in transportation are commutters who were not surveyed for their percpetion of transprotation in *Zhong Guan Cun*. Secondly, despite that stratefied sampling was conducted to ensure that the samples reflect the socio-demographic structure of residents in a *Jie Dao*, due to the high frequency of mismatch between where people actually live and where they are registered for their residence, the samples might be biased as well. However, there is no easy way to handle this "residence-registration mismatch" problem that is pretty common in China.

Future investigation of urban transportation convenience problem in Beijing should link residents' perceptions of transportation with their sociodemographic charactersitics to reveal the perceptions and preferences of different groups of residents. This is especially important when "quality of life" and "city livability" are becoming a well-accepted concept to both citizens and city government in Beijing. A convenient urban transportation system should be one that satisfies residents' travel demands. To understand who needs or prefers what regarding urban transportation services can certainly help Beijing to deliver better urban transportation services to its residents. It is also worth further effort to compare specific *Jie Dao*s from "LH" and "HL" groups to find out what the determining factors are that cause the contrast between objectively measured transportation LOS and subjective assessed one. The related findings will shed light on the identification of effective measures to improve transportation services.

**Acknowledgments** The first author wants to acknowledge the faculty development leave awarded by Texas State University and the visiting professorship by Beijing Normal University. The second author appreciates the grant support by Beijing Municipal Natural Science Foundation (No. 8073024) and by Planning Program (Major) of Beijing Philosophy and Social Science (No. 06AbCS001).

## References

Cai, J. (2006). Discussions on the theoretical reasoning of proper distances among urban roads. *Urban Transportation*, 2006(1), 55–59. (蔡军.关于城市道路合理间距理论推导的讨论[J].城市交通, 2006(1), 55–59)

Chen, J. (2007) The space analysis of livable Beijing's transportation convenience. Unpublishe master thesis. College of Arts and Sciences, Beijing Union University

Cheng, M. (2006). Beijing ranked number 15 for the raking of the most livable cities in China. *Chiese Youth Daily*, [electronic version]. Retrieved from http://zhidao.baidu.com/question/2515668.html?fr = qrl

Dang, D. (2006). Public transit is a strategic priori for developing "livable city". *Frontier Thoughts*, 2006(2), 54–55. (党德信.公交优先是建立"宜居城市"的战略选择[J].思想前沿, 2006(2), 54–55)

Fang, J. (2006). A study of urban transportation management system using GIS. *Science and Technologies for the Public*, 2006(1), 116-117. (方吉祥,基于GIS的城市交通管理系统的研究[J]. 大众科技 2006(1), 116–117)

Lee, D., Kim, T. & Pietrucha, M. (2007). Incorporation of transportation user perception into evaluation of service quality of signalized intersections. *Transportation Research Record*, 2027, 9–18

Lennard, S.H.C. & Lennard, H.L. (1995). *Livable cities observed. A source book of images and ideas for city officials, community leaders, architects, planners, and all others committed to making their cities livable*. (Carmel, CA: Gondolier Press)

Li, M. & Hu, Q. (2006). Approaches to improvement traffic congestion at urban street intersections. *Standardization of Transportation*. 2006(1), 106–108. (李敏, 胡群袖.浅谈改善现代城市交叉口交通阻塞的措施[J]. 交通标准化 2006(1), 106–108)

Lu, Y. & Tang, J. (2004). Fractal dimention of transportation network and its relationship with urban growth: a study of Dallas-Fort Worth area. *Environment and Planning B*, 31, 895–911

Madanat, S.M., Cassidy, M.J. & Ibrahim, W.H. (1994). Methodology for determining level of service categories using attitudinal data. *Transportation Research Record 1457 (TRB)* (pp. 59–62). (Washington, D.C: National Research Council)

MSGUNCC (Metropolitan Studies Group, University of North Carolina at Charlotte). (2008). *Charlotte neighborhood quality of life study 2008*, [electronic version]. Retrived from http://ww.charmeck.org/QOL/PDF/2008QualityofLife_FullReport.pdf

Ndoh, N.N. & Ashford, N.J. (1994). Evaluation of transportation level of service using fussy sets. *Airport and Airspace planning and Operations*, 1461, 31–37

Nian, J. (2006). *Report on the life of quality in Chinese cities in 2006*. (Beijing, China: Economics Press)

No Author (2005). Report on the city livability measurement in China in 2005. *Business Week*, 2005(3), 99–103

NRC (National Research Council), National Academy of Sciences. (2002). *Community and quality of life: data needs for informed decision making*. (Washington DC: National Academy Press)

Pecheux, K.K., Pietrucha, M.T. & Jovanis, P.P. (2000). User perception of level of service at signalized intersections. *The 79th Annual Meeting of the Transportation Research Board*, Washington, D.C

Pfefer, R.C. (1999). Toward reflecting public perception of quality of service in planning, designing, and operating highway facilities. *Transportation Research Record 1685* (TRB) (pp. 81–89). (Washington, D.C: National Research Council)

Quang, Y. & Sun, M. (2006). Reflections on the development of public transit as a priority. *Urban Transportation*, 2006(1), 23–27. (全永粲,孙明正等.公共交通研究,优先发展公共交通历程中的若干问题反思[J].城市交通 2006(1), 23–27)

Shi, Q. & Wang, X. (2006). An evaluation method for assessing the overall construction level of urban transportation network. *Journal of Natural Sciences of Hefei Industry University*, 2005(12), 1496–1498. (石琴, 汪秀英等.城市交通网络总体建设水平综合评价方法[J], 合肥工业大学学报(自然科学版),2005(12), 1496–1498)

Si, B. & Gao, Z. (2006). SUE traffic assignment model and algorithm – Urban transportation network with stopwarning information system. *Science and Technology for Road Transportation*, 2006(1), 120–124. (四兵锋,高自友等.停车诱导信息系统条件下的城市交通网络SUE配流模型及算法[J]. 公路交通科技, 2006(1), 120–124)

Shafer, C.S., Lee, B.K. & Turner, S. (2000). A tale of three greenway trails: user perceptions related to quality of life.*Landscape and Urban Planning*, 49, 163–178

Sheng, Q. (1997). Urban transportation in Shanghai, China: problems and planning implications. *International Journal of Urban and Regional Research* , 21(4), 589–606

Sit, V.F. (1996). Beijing – Urban transportion issues in a socialist Third World setting. *Journal of Transport Geography* , 4(4), 253–273

Wang, D. & Chai, Y. (2008). The job-house relationship and communting in Beijing, China: the legacy of Danwei. *Journal of Transport Geography* , doi:10.1016/j.jtrangeo.2008.04.005.

Wang, J. & Wang, F. (2006). Studies on the safety assessment of urban transportation accidents. *Management and Development of Mechanism*, 2006(2), 121–122. (王俊人, 王凤英. 城市道路交通事故安全评价研究[J]. 机械管理开发. 2006(2), 121–122)

Wang, W. & Zhang, Q. (2006). Some thoughts on safe and smooth urban transportation. *Transportation Standardization*, 2006(1), 79–81. ((王文涛,张启义.对城市交通安全畅通问题的几点思考[J].交通标准化, 2006(1), 79–81)

Ye, P. & Chen, X. (2006). The study of urban transportation network planning for the functional zones. *Urban Transportation*, 2006(1), 36–41. (叶彭姚,陈小鸿.功能组团格局城市道路网规划研究[J]. 城市交通, 2006(1), 36–41)

Yi, H. (2006a). Comments on interrelations between urban landuse and transportation. *Urban Transportation*, 2006(4). (易汉文; ??????[[J]. 城市交通, 2006(4))

Yi, H. (2006b). General description of land-use and transportation modeling systems. *Urban Transportation*, 2006(5). (易汉文; ??????[J]. 城市交通, 2006(5))

Zhang, L. & Prevedouros, P.D. (2004). Signalized intersection LOS that accounts for user perceptions. *The 2004 Annual Meeting of the Transportation Research Board*. Washington, D.C

Zhang, W., Yin, W., Zhang, J., Meng, B. & Gao, X. (2006). *A study of livable cities in China (Beijing)*. (Beijing, China: Social Science Academic Press) (张文中, 尹卫红, 张锦秋, 孟斌, 高晓路 等, 《中国宜居城市研究报告 (北京)》, 社会科学文献出版社, 2006)

Zhang, J. & Zhu, Y. (2005). A spatial data structure for urban transportation and roads using MapInfo. *Knowledge and Skills on Computer*, 2005(10), 33–34. (张锦宗,朱瑜粉.基于MapInfo的城市交通道路空间数据组织[J].电脑知识与技术, 2005(10), 33–34)

# Chapter 11
# Object-Oriented Data Modeling of an Indoor/Outdoor Urban Transportation Network and Route Planning Analysis

**Deelesh Mandloi and Jean-Claude Thill**

**Abstract** We present an object-oriented data model to represent the multi-modal, indoor/outdoor transportation network of an urban area, that can be used for route planning and navigation and to perform other network analyses. The data model divides the urban transportation network into two main parts, one being the network for modeling movement inside buildings and the other being the network outside buildings using multiple modes. The network inside buildings is intrinsically three-dimensional (3D), thus requiring to model vertical connectivity between floors. Such 3D indoor network is modeled using a 2.5D approach that combines the existing 2D data structures and 3D visualization techniques available in commercial off-the-shelf GIS software. The network outside buildings is modeled as a multi-modal network consisting of modes operating on streets, walkways, and public transit routes. The data model is implemented for the multi-story buildings and transportation network on the campus of the State University of New York at Buffalo, USA. The effectiveness of the data modeling scheme is evaluated by performing path finding analysis in the study area using the data model with standard and customized GIS software tools.

**Keywords** Transportation network · Object-oriented data model · Modeling movement · 3D network modeling

## 11.1 Introduction

Geographic Information Systems (GIS) are widely used to model transportation networks as they provide powerful tools for geospatial data management, visualization, presentation and analysis. A GIS usually stores a transportation

J.-C. Thill (✉)
Department of Geography and Earth Sciences, University of North Carolina at Charlotte, Charlotte, NC 28223, USA
e-mail: jean-claude.thill@uncc.edu

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_11,

network as it exists on the surface of the earth in two dimensions (2D). The node-arc data model is the foundation for storing topological relationships of connectivity in vector GIS. Although this model is suitable for many modeling tasks, it is not able to represent the complex reality of some situations such as those arising from non-planarity and the three-dimensional (3D) nature of the network, especially in densely settled environments. Hence many additions have been made to the basic network data model.

Many transportation-related applications such as emergency dispatch and route planning, evacuation planning, turn-by-turn travel advisory services, and others require accurate estimates of travel times from trip origins (e.g., emergency facilities or home) to destinations. Destinations in urban areas may well occur on an upper floors in a multi-story building with multiple entrances. So when determining the best route to a destination it is imperative to find the "best" route not only up to the entrance of the building but also inside the building. However modeling the traversable network of corridors inside the building is a complex task due to its 3D nature.

The traditional node-arc model used by a GIS to store connectivity relationships is 2D by design. The node-arc model determines the coordinates for the point and line entities in a 2D Cartesian coordinate system. Some vector GIS data models also permit the storage of a Z-coordinate for vector entities. However this Z-coordinate is just used as an attribute and not derived from the coordinate system upon which the X and Y coordinates of the vector entities are calculated. Thus such vector data models, although useful in visualizing the vector entities in 3D, are unable to determine topological relationships such as connectivity in 3D.

This chapter presents an object-oriented GIS data model capable of supporting enhanced routing and navigation in mixed, indoor and outdoor, urban environments by modeling a multi-modal and indoor/outdoor transportation network. We first propose a modeling approach for routable pathways of movement inside enclosed, multi-story, built structures (buildings). This data model is then extended to movement on infrastructure outside these structures (on sidewalks, walkways, in personal vehicles, as well as on public transportation systems) in a seamlessly integrated fashion. To model 3D connectivity inside multi-story buildings, our approach uses the object-oriented concept of connectivity groups, which has been incorporated in the network data model of commercial off the shelf (COTS) GIS software, ArcGIS. The same data modeling principles are also used to represent the network outside the buildings two-dimensionally, as a multimodal network allowing for various modes of travel.

The rest of the chapter is organized as follows. Section 11.2 provides a conceptual background on network data models used in GIS for multi-modal transportation networks. Section 11.3 develops the conceptual data model for the indoor/outdoor urban transportation network. A real-world implementation of the data model is given in Section 11.4, followed by the results of a 3D route finding application in Section 11.5. Section 11.6 presents the conclusions.

## 11.2 Network Data Models

In GIS-T, the vector data model has found widespread use to model transportation networks (Goodchild 1998). A transportation network is modeled as a network of interconnected nodes and links. This vector data model along with the ability to store topological relationship of connectivity is called a network data model. Using the network data model, a GIS can perform simple as well as sophisticated network analysis such as vehicle routing, location-allocation modeling, accessibility analysis, and many others (Thill 2000).

The connectivity between nodes and arcs in a network is stored using a node-arc representation (Miller and Shaw 2001). This representation stores the connectivity using the directionality of arcs and by enforcing rules such that every arc is bounded by a start node and an end node. Also introduction of a new node requires an arc to be split at that node. The node-arc representation thus requires that the modeled network is planar in nature. This serves some GIS-T applications well but quite often the transportation network has overpasses and underpasses that require a non-planar network representation. So the basic planar network model has been extended by using turn tables to model specific connectivity at an overpass or by using Z-level attribute values associated with street segments (Goodchild 1998, Thill 2009).

The network data model also stores the impedance to traverse an arc in each direction as a link attribute. Such network data models with arc impedances are implemented in a GIS using efficient data structures enabling network-based analysis such as determining shortest or least cost paths on extremely large datasets in finite time and using reasonable computer resources (Zhan and Noon 1998, Zhan and Noon 2000).

### *11.2.1 Modeling Multi-Modal Networks*

Transportation networks are commonly designed to be used for multiple modes of transportation. A multi-modal transportation network can be defined as one in which the movement of resources (people or freight) may occur by means of two or more different transportation modes from the point of origin to the point of destination (Southworth and Peterson 2000). Thus a multimodal network may involve the use of different transportation infrastructures such as road, rail, water, or air, or it may involve a single mode supporting different transportation services such as the public bus system and personal car, both of which typically use the same network of streets and roads.

The most important consideration in representing a multimodal network, and therefore possible movements on it, is the modeling of *transfer points* or those locations on the network where transfer from one mode to another takes place (Rietveld 1997). In urban transportation networks, these points can be the start of a walkway off a parking lot, bus stops where a person can board a bus after driving to the bus station in his/her car, or the edge of a sidewalk at a pedestrian crossing. These transfer points assume importance because connectivity between two modes is established only at these transfer points. So even if a bus route and the street network are geometrically coincident linear features, a commuter can get on a bus only at designated bus stops. Also while calculating travel impedances, substantial impedance may be incurred at these transfer points due to wait for a green light or for a scheduled service, and inconvenience.

Modeling of multimodal networks using GIS has been achieved in the past using various modeling strategies that incorporated an intelligent combination of available network modeling techniques. Southworth and Peterson (2000) presented one such methodology to model multimodal freight networks. Choi and Jang (2000) proposed another approach to model bus transit network using linear referencing methods. Miller et al. (1995) and Miller and Storm (1996) extended this approach with a virtual network design strategy based on a single physical network to which multiple modal network are logically linked.

The traditional approach to multi-modal network design is to maintain separate sub-networks for each mode in a GIS database. Using this approach, a separate logical network is later built using the physical network made up of separate modes which defines connectivity between modes. The transfer points act as the only locations where transfers between two modes are possible as shown in Fig. 11.1. Miller et al. (1995) argued that this approach can prove unsatisfactory in cases where some modes such as street network and bus routes share common infrastructure such as streets. The problem of database integrity arises when one mode is edited independently of another mode.

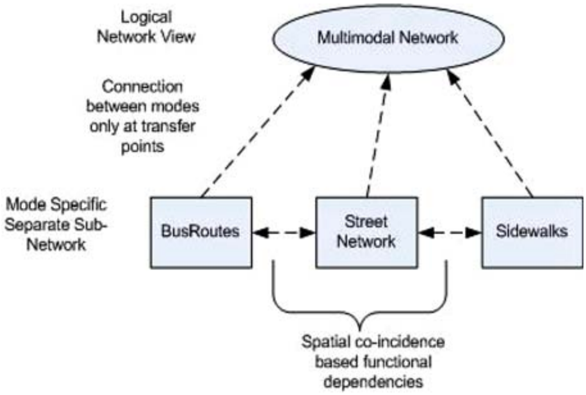



**Fig. 11.1** Multimodal network derived from separate mode specific base sub-networks

Recent advances in GIS data modeling allow for easy modeling and powerful data management of coincident features such as streets and bus routes using topology rules (e.g., Curtin et al. 2003). Furthermore, physically separating each modal sub-network provides easy data integration when the data is created and maintained by separate organizations. While building the multimodal network, data from organizations having their own protocols for data maintenance can be separately integrated and modeled for coincidence using topology rules. Also the problem of modeling appropriate travel impedances at transfer points can be avoided by using transfer arcs along with transfer points as the connecting elements between different modes. This strategy is used in the multimodal network data model developed in this chapter.

### *11.2.2 3D Network Modeling*

Most COTS GIS only have 3D visualization capabilities and lack 3D analytical capabilities such as identifying topological relationships of connectivity in 3D space (Lee 2004). A 3D GIS capable of storing 3D topological relationships should be able to answer queries such as whether an underground pipe intersects a land parcel or to determine all the corridors connected to a particular corridor on the second floor of a building. However, most COTS GIS can store topological relationships in 2D space only. The past decade has seen a growing interest for modeling street networks in ways that would respect topological relationships so as to enable the routing along network's arcs. These models extend the 2D node-arc representation through a variety of modeling techniques that range from intriguing and clever to rather awkward. The approaches proposed by Li and Lin (2006), Dueker and Butler (2000), Fohl et al. (1996), and Zhu and Li (2008) are particularly noteworthy.

In order to model topological relationships in 3D space, the GIS should support 3D topological data models. Although formulated in theory, such data models have not been yet implemented in COTS GIS (Raper and Kelk 1991). This is a significant challenge for the modeling of traversable indoor spaces, which are typically composed of multiple floors interconnected at a few locations, i.e., elevator cages, stairwells, escalators, and ramps. Some researchers like Lee (2004) have implemented 3D topological models for modeling network connectivity using dual graphs representing geometric and logical networks. Others like Groger and Plumer (2003) have proposed a 2.8D map by extending the basic 2.5D digital terrain model commonly available in GIS. Graf and Winter (2003) present an implementation of indoor routing application similar to that presented in Section 11.5, but no elaboration of the network data structures supporting the optimization algorithm is provided.

## 11.3 Modeling Transportation Networks in Urban Environments

While planning a route in an urban environment, people have to make many travel decisions such as whether to accomplish the travel using a single mode or multiple modes, whether to follow a shortest distance route to the destination or take a route that minimizes the total travel time, or to take a route that involves least effort. A GIS data model designed to support route planning (and possibly also navigation) in an urban environment must be able to model these scenarios for it to be effective. The main objective of the chapter is to propose one such routable data model that can be used to model route planning tasks in and around building that populate urban environments. Thus this section presents the conceptual and logical data models to represent and support route planning tasks in urban environments.

Movement in an urban environment not only takes place on the streets and on other transportation modes such as public transit and sidewalks, but it also occurs inside buildings where businesses or other economic activities occur. Hence the present chapter divides the data model into two functional parts, namely modeling movement inside buildings and modeling movement outdoors.

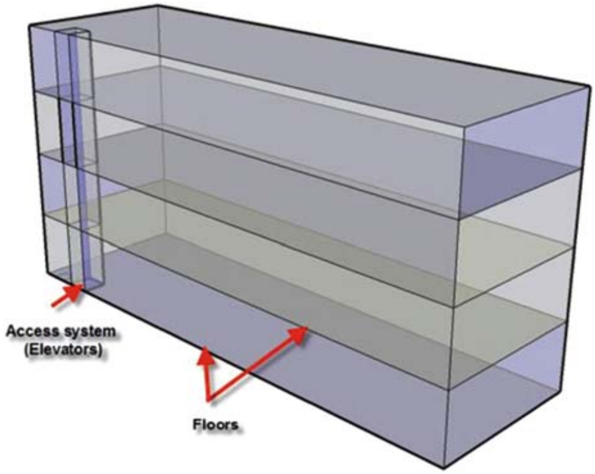
## *11.3.1 Modeling Movement Inside Buildings*

In urban areas, especially in central business districts (CBD), a significant proportion of personal travel occurs inside multi-story buildings as business and other activities take place on various floors. So when trying to reach a destination, for example on the 15th floor of a building, it is not only necessary to get the driving directions up to the door steps of the building but also within the buildings as there might be multiple routes with different travel impedances to reach the destination even inside the building. This is particularly important for individuals with limited mobility, the elderly, people with a physical handicap, pregnant women, or youngsters, for instance, for whom some indoor paths may present significant challenges. This leads to the interesting research question of modeling indoor traversable links and movements so that path planning support systems that provide turn-by-turn directions to the building's door steps can be extended to provide travel directions from the building entrance to the end destination inside the building.

### 11.3.1.1 Conceptualizing the Floor Network

A multi-story building can be conceptualized as being made of separate compartments where each compartment represents a floor of the building. These compartments are further sub-divided into rooms and corridors. The corridors act as main links connecting all the rooms and entrance/exit points on a floor. Thus the corridors form the network features along which the majority of the movements takes place. The buildings also have an access system made up of stair cases, elevators and doors that enable movement between floors and from the interior of the building to the open space outside the building. This simplified conceptual representation of a multi-story building is shown in Fig. 11.2. Thus each floor has its own network of corridors where movement takes place

Fig. 11.2 Conceptual representation of a multi-story building




and these individual floor networks are vertically connected through stair cases and elevators acting as entry and exit points for the floors.

A corridor covers an area in floor space and so it could be modeled in a GIS database as an area feature in a polygon feature class (in the same way as a roadway has an areal extent and can be represented as a polygon in a vector database or a set of grid cells in a raster database). However a network data model consists of network features made up of line and point features only. Thus, for consistency of representation, the polygon representing a corridor is converted into a line feature stored in a line feature class, along which possible movements occur. This line can be represented by the medial axis of the polygon. Such medial axis can be derived using a straight medial axis transformation (Lee 2004). The access system made up of stairs, elevators and doors can be modeled as point features through which the corridor lines connect with each other. Thus the staircase and the elevator locations on each floor can be represented as point features in a point feature class. The movement within the building can then be conceptualized as a network made up of corridor lines and elevator and stair case points coincident on these corridor lines acting as points of vertical connectivity between floors.

It is common for all floors of a building to occupy the same area in the 2D space. However the floors are separated by a certain height if viewed in 3D. So if we assume that the height between floors is constant (say, 5 m), then the coordinates representing the vertices (or shape points) that make up the geometry of corridor lines on the first floor will have a relative Z value of 0 m, the second floor corridor lines will have a Z value of 5 m, third floor as 10 m, and so on. Thus although each floor network may appear coincident when viewed in 2D (same footprint), their relative heights renders the building network they belong to three-dimensional in nature, as shown in Fig. 11.3.

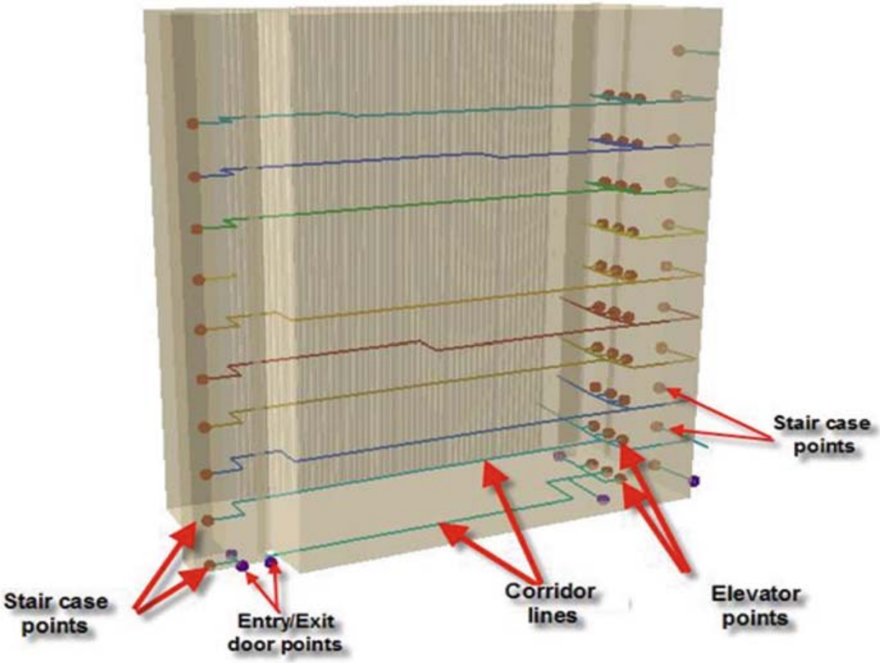



**Fig. 11.3** Floor network made up of corridor lines, elevators, stair cases and entry/exit door points in a building with three elevators, two stair cases and six entry/exit doors

The floor network described above cannot be modeled using the traditional node-arc representation as two floors may well have corridor networks that perfectly coincide in the 2D space: the vertices of the corridor lines on both floors have identical X an Y coordinates but different Z coordinates. It is important to realize that the turn tables or Z-level fields used sometimes to model relatively simple grade separations such as highway overpass and underpass, are still not sufficient to model the connectivity between floors as turn tables or Z-level fields are defined only at network nodes (or line endpoints) and not at each vertex that defines the geometry of line features (shape points). So although we can model that connectivity between floor 1 and floor 2 occurs at an elevator point only, it is difficult to specify that the connection cannot be established between coincident vertices of corridor lines. Building networks present a situation more complex than the non-planarity encountered in roadway networks, for which new modeling approaches must be devised.

The approach advocated here exploits the property that the majority of functional buildings are comprised of floors stacked on top of one another rather than a jumble of rooms haphazardly connected by hallways and piled up like interlocked Lego blocks. From this premise, the vertical connectivity between floors can be effectively represented if the network on each floor is placed in a separate group and connectivity between each group is allowed only at pre-defined points and not at any other points even if they are coincident. So by placing each floor sub-network in one connectivity group and specifying that connection between the groups occurs only at elevator and staircase points, the 3D connectivity required to represent each floor sub-network can be modeled using a 2D implementation of network connectivity relationships.

#### 11.3.1.2 Modeling Movement Between Floors

The above conceptual data model to represent floor networks provides the ability to model vertical connectivity relationships between floors. However, in order to realistically model movement inside buildings, the time or effort required to travel between floors also plays an important role in route planning. Thus the data model should have provision to model travel impedances associated with navigation inside buildings. This can be achieved by associating impedances with floor network features in the form of cost-based network attributes.

Cost-based network attributes act as impedances that are assigned proportionally along the network edge. For example a cost attribute called TravelTime can be used to store the time required to traverse a particular network edge in a given direction. As the time required to travel an edge may differ based on travel directions, two fields (e.g., FromTime, ToTime) are used to store the costs for a given edge. The measure of travel time required to traverse an edge can be derived by actual measurements or imputed from an average walking speed (which can be variable for different user groups) and the length of the corridor line. Indoor walking speed would normally be lower that outdoors.

The vertical connectivity between floors exists through staircases and elevators which are modeled as point features. Thus two floor lines connect to each other through a point feature representing an elevator or staircase point. In reality, there is travel impedance associated with such a vertical movement. In the case of an elevator, the travel impedance corresponds to the average time for the elevator to arrive at a given floor along with the time for the elevator to travel from one floor to another. In the case of a staircase, the travel impedance indicates the time it takes to walk up or down the staircase. As the data model conceptualizes the transfer points (i.e. the points where features in two connectivity groups connect with each other) as point features, the cost-based travel impedances in vertical directions are modeled using turn features. These turn features represent the cost required to transfer between two edges on separate floors. For example to store a travel time of 20 s between floor 1 and floor 2 using the elevator, a turn feature can be created from floor 1 to floor 2 with associated attribute of 20 s. It should be noted that in order to model the movement in the reverse direction, a new turn feature is required as the travel time can be different.

The turn feature is used to store the impedance associated with traveling in vertical direction. A realistic estimate of this measure for staircases and elevators can be derived using mathematical models. For travel along staircases, the travel time can be calculated by using an average walking speed and the distance traveled along the vertical direction, which can be approximated as the difference in Z levels of the floors. However this approach does not take into account the effort and likely inconvenience associated with using staircases to travel many floors. For instance, while little discomfort is incurred going one flight of stairs from floor 1 to floor 2, it is quite likely to be much more strenuous and time-consuming to use stairs from floor 1 to floor 8. So the measure of travel impedance must incorporate a sensible account of the walker's discomfort. This can be achieved by using a model in which travel time increases exponentially with the number of floors traveled, $t = m^{\beta} * (d / s)$, where t denotes the travel time, m is the number of floors traveled, $\beta$ is the inconvenience factor ($\beta \geq 1$, $\beta = 1$ indicates the absence of inconvenience), d is the distance traveled vertically, and s is the average walking speed in stairs. More complex models can also be used, as reported in Fujiyama and Tyler (2004), Gupta and Yadav (2004), Lee and Lam (2006), and others.

While using elevators to travel between floors, the travel time can be modeled using a simple linear relationship which depends on the average wait time, plus the time to reach the destination floor once the user has boarded the elevator. A simple model is as follows: $t = T + (d/s)$, where t denotes travel time, T is the average wait time, d is the distance traveled vertically, and s is the average speed of the elevator. More complex models can also be substituted for this simple relationship (e.g., Strakosch 1985).

## *11.3.2 Modeling Movement Outside Buildings*

The travel outside of buildings can be modeled on a multimodal network. While modeling a multimodal urban network it is important to consider the transfer points where transfers between modes can take place. This is important as many times two modes may use the same transportation infrastructure. Also, the delays associated with mode transfers must be incorporated in travel impedance measures modeled as network cost attributes.

### 11.3.2.1 Modeling Multiple Modes

Street Network

The streets can be modeled by line features representing the street centerlines traversed by private vehicles. The connectivity between streets can be represented using the node-arc model such that two streets connect to each other at coincident endpoints. The non-planarity of street features encountered at grade separations is modeled by Z-level attribute fields. The Z-level attribute fields define the connectivity of street edges at street intersections and connectivity between the street edges is established only if endpoints have identical Z-level attribute values. Expected delays at intersections (whether signalized or not) can be modeled using turn tables. The travel impedance is represented using a cost attribute based on the length of street features or using travel times imputed from posted speed limits and length of street features. One-way restrictions can be modeled by restriction-based network attributes. Street features can also have associated address ranges to help in determining network locations using geocoding.

### Bus Routes

The bus routes are modeled as linear features in a line feature class. The bus routes share the same geometry as the underlying street network. Thus line features representing bus routes are geometrically coincident with line features representing streets. This requires special topological relationships to be maintained between features of these two feature classes so that the geometrical edits made to one feature are also reflected in another feature to maintain referential integrity of the data. This can be achieved by using coincidence-based topological feature editing tools such as map topology in ArcGIS (ESRI 2006).

Bus service occurs only along predefined routes, while boarding, alighting, and transfer between routes are possible only at fixed bus stops. Moreover, two bus routes can share some or all the street segments; bus services serve the designated bus stops in a pre-determined order. A prudent method to model such behavior is to represent each bus route in a separate line feature class made up of street segments that define the route geometry. The stop sequence on the route can be modeled by restriction-based network attributes similar to those used to model one-way streets. For example to model a sequence of stops along a route, the route can be treated as being traversable only in the direction that gives the scheduled stop sequence and by treating all other edges as restricted, as shown in Fig. 11.4.

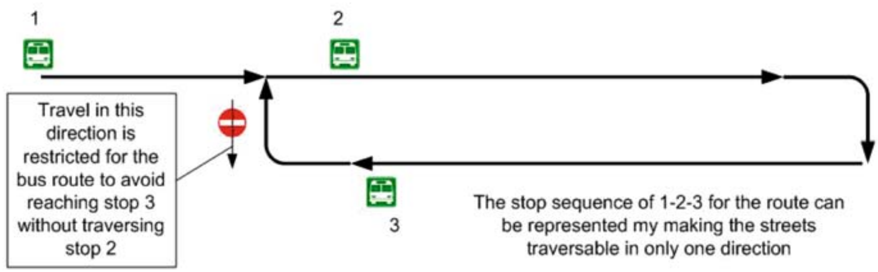



**Fig. 11.4** Modeling the stop sequence along bus routes

The travel impedance between two stops on a route can be calculated from the bus schedule or by using the speed limit and length of streets. Since each route has an underlying street feature that represents its geometry, it is easy to model travel times for two routes sharing street segments but having different travel times.

### Walkways

The walkways represent the paths along which pedestrian travel takes place. They are made up of sidewalks and crosswalks that connect two sidewalks or provide a safe passage for street crossing. Every possible pedestrian path can be represented as a line feature in a line feature class. The travel impedance along walkways can be estimated using an average walking speed and the length of walkway edges.

### Modeling Transfers Between Modes

Although the modes discussed in Section 11.3.2.1 can be used independent of one another, route planning and navigation in urban environments commonly involves the use of multiple modes. The transfer between travel modes occurs only at specific locations irrespective of the possible geometric coincidence of the modes in question. For example although the street network and bus route are coincident, people can transfer from the street network to bus routes only at bus stops. Thus bus stops act as transfer points that allow connectivity between different modes.

Bus stops act as transfer points where a passenger can embark or disembark during his/her journey. Because a sidewalk is generally not coincident with a bus route, transfer arcs are introduced, which are simple line features that connect two modes through the transfer point. This transfer arc can also be assigned travel impedance representing the delay or inconvenience associated with making a particular modal transfer as shown in Fig. 11.5.

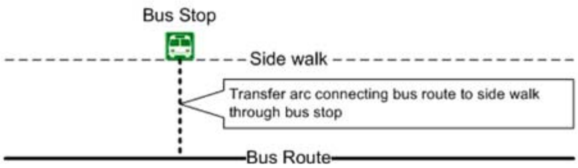



**Fig. 11.5** Modeling transfer between bus routes and walkways

The transfer between the street network and walkways occurs at parking lots or designated parking spaces along the streets. A parking lot is made up of a parking entrance through which cars enter the lot and road segments along which cars ride to reach parking spaces. These road segments within the lot are also used for walking by drivers after parking their cars. Thus a parking lot can be modeled as a complex feature comprised of a point feature representing the parking lot entrance and a set of line features representing road segments within the lot. A transfer arc connecting the parking lot entrance to the nearest point on the access street segment is used to represent travel delay associated with the use of the parking lot as shown in Fig. 11.6. Essentially, vehicles are virtually parked on this transfer arc, which renders the representation of individual parking spaces unnecessary in our modeling approach. Parking road segments

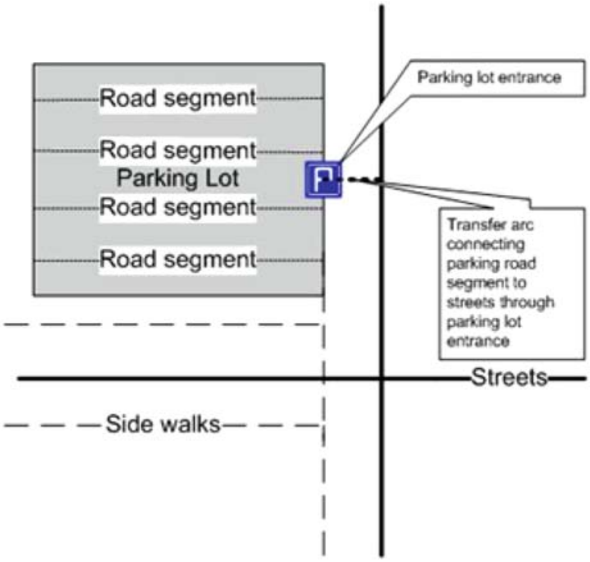



**Fig. 11.6** Modeling transfer between streets and walkways

are traversed by pedestrians only, not by cars. The transfer between the street network and the bus network occurs through walkways which act as common transfer between these two modes.

## 11.4 Data Model Implementation

The multimodal network data model for enhanced route planning in urban environments developed in Section 11.3 is implemented for a study area comprising the North Campus of the State University of New York at Buffalo (UB) located in Amherst, New York. The database consisting of the following feature classes is created in COTS GIS, ArcGIS version 9.1, from ESRI.

### *11.4.1 FloorNetwork*

The FloorNetwork feature class is a line feature class that stores the corridor lines representing the corridors on each floor in each building. These corridor lines are derived from the buildings' floor plans. After the floor plans are georeferenced, the corridor lines are digitized as lines passing through the medial axis based on the corridor area as shown on the floor plan. Only publicly accessible corridors are used as shown in Fig. 11.7. Corridor lines are assigned an appropriate floor number and a Z elevation attribute. The value for the Z attribute is chosen arbitrarily and any value based on user needs or study area

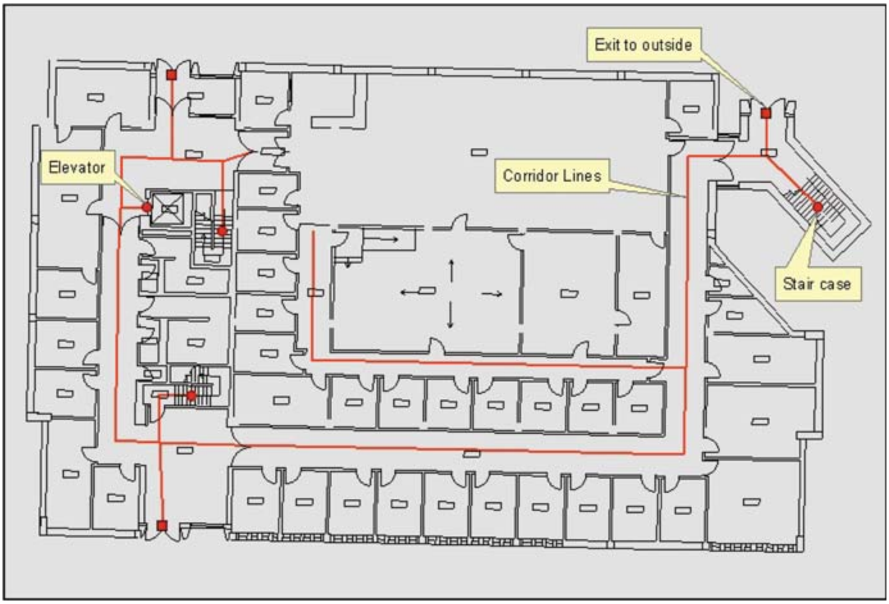



**Fig. 11.7** Corridor lines and exit points derived from a geo-referenced floor plan

preferences can be chosen. This value is used only for visualizing the lines in 3D and has no effect on analysis results.

The corridor lines for each floor of the building are stored in a separate subtype such that all corridor lines for first floor in all the buildings are in a single subtype called floor1. The total number of subtypes is 12, which is equal to the maximum number of floors in any of the buildings in the study area.

### *11.4.2 ExitPoints*

The ExitPoints feature class is a point feature class that stores the location of staircases, elevators and entry/exit doors on a particular floor within the buildings. These points act as access system for the building and establish the vertical connectivity between two floors. The door locations that facilitate entry and exit for the buildings are also digitized. The doors, staircases and elevator locations in a floor are derived from geo-referenced floor plans as shown in Fig. 11.7.

### *11.4.3 FloorTurns*

The FloorTurns feature class is a line feature class that represents the vertical connection between floors. The line feature in the FloorTurns feature class

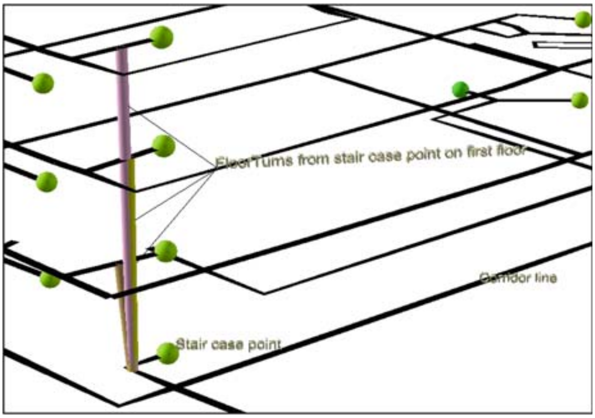



**Fig. 11.8** FloorTurns for one staircase point

starts at the corridor line of one floor and ends at the corridor line of another floor. The corridor lines identified by FloorTurn features are connected through staircase or elevator points to corridor lines on the reference floor as shown in Fig. 11.8. For every elevator or staircase point there are m-1 FloorTurn features, where m is total number of floors in a building. The primary purpose of this feature class is not to store the vertical connectivity between two floors but to store the travel time required to travel between the floors.

### *11.4.4 WalkWays*

The WalkWays feature class is a line feature class that stores sidewalks and crosswalks used by pedestrians. The walkways make up the walking mode in the network. Pedestrians are allowed to walk only on the walkways, which are represented as separate features from the street centerlines. The walkways also include the road segments found inside parking lots as these segments facilitate walking to one's destination after parking a car. The sidewalks and crosswalks are digitized as centerlines of the area they occupy, seen on one-foot digital orthophotos.

### *11.4.5 Streets*

The streets feature class is a line feature class that contains street centerlines for the streets on the UB campus. This feature class is populated from the Accident Location Information System (ALIS) streets layer available from the New York State GIS Clearinghouse (ALIS 2006).

### *11.4.6 CampusBusRoutes*

The CampusBusRoutes is a line feature class that contains the bus routes used by campus shuttles.

### *11.4.7 CampusBusStops*

The CampusBusStops is a point feature class that contains bus stop locations. These bus stops are served by the campus shuttle system. The bus stop locations are coincident with walkways to maintain proper connectivity.

### *11.4.8 ParkingLotEntrance*

The ParkingLotEntrance feature class is a point feature class that contains point features marking the entrance and exit points of parking lots. The parking lot entrance point is coincident with road segment line features from walkways to maintain proper connectivity. The location is digitized from one-foot digital orthophotos.

### *11.4.9 Sidewalk_Busroute_Transfer*

The Sidewalk_Busroute_Transfer feature class is a line feature class containing transfer arcs that connect bus stop locations to bus routes. The transfer arcs are used to establish connectivity between sidewalks and bus routes through bus stops as the bus stops are co-incident only with sidewalks. The transfer arcs also represent the average wait time before the next bus arrives at the stop. This value is derived from route schedules as half the average headway. If a stop serves multiple routes, then a transfer arc is created for each route, as the average wait time can be different for different routes.

### *11.4.10 Streets_Walkways_Transfer*

The Streets_Walkways_Transfer feature class is a line feature class containing transfer arcs that connect parking lot entrances to walkways. The transfer arcs are used to establish connectivity between road segments in parking lots (which are represented as pedestrian walkways) and street centerlines through parking lot entrances. The transfer arcs also represent the average wait time required to park a car or to exit a parking lot. This value is derived from average driving speed in the parking lot and the average distance traveled by a car before a parking space is found. This average distance is derived as one half of the cumulative length of road segments in a parking lot plus the length of the transfer arc as the car has to travel at least the transfer arc length even if the parking space nearest to the entrance is taken and at most the entire road segment length if the parking space farthest away from the entrance is available. This value is stored in FT_MINUTES and TF_MINUTES fields using an average driving speed of 10 miles/hour in the parking lot.

### *11.4.11 Creating the Network Dataset*

After creating and populating various line and point feature classes representing transportation network features, the network dataset is built to store the topological relationships of connectivity between network features. The network dataset stores the connectivity information in a set of tables along with network attributes in the form of travel impedances. This information is used for performing various network-based analyses such a finding optimal paths.

The ten feature classes presented in Section 11.4 act as network sources while creating the network dataset. The infrastructure for each transportation mode is included in a separate connectivity group. For many line feature classes (CampusBusRoutes, for instance), each subtype constitutes a separate connectivity group. The transfer points are placed in each of the connectivity groups for the modes they facilitate transfers. The connectivity scheme is given in Fig. 11.9 shows the composition of each of the 17 connectivity groups created for this case study.

| Source | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CampusBusRoutes : Green Route | | | | | | | | | | | | | | | | | ✓ |
| CampusBusRoutes : Red Route | | | | | | | | | | | | | | | ✓ | | |
| CampusBusRoutes : Yellow Route | | | | | | | | | | | | | | | | ✓ | |
| FloorNetwork : Floor -1 | | | | | | | | | | | | ✓ | | | | | |
| FloorNetwork : Floor 1 | ✓ | | | | | | | | | | | | | | | | |
| FloorNetwork : Floor 10 | | | | | | | | | | ✓ | | | | | | | |
| FloorNetwork : Floor 11 | | | | | | | | | | | ✓ | | | | | | |
| FloorNetwork : Floor 2 | | ✓ | | | | | | | | | | | | | | | |
| FloorNetwork : Floor 3 | | | ✓ | | | | | | | | | | | | | | |
| FloorNetwork : Floor 4 | | | | ✓ | | | | | | | | | | | | | |
| FloorNetwork : Floor 5 | | | | | ✓ | | | | | | | | | | | | |
| FloorNetwork : Floor 6 | | | | | | ✓ | | | | | | | | | | | |
| FloorNetwork : Floor 7 | | | | | | | ✓ | | | | | | | | | | |
| FloorNetwork : Floor 8 | | | | | | | | ✓ | | | | | | | | | |
| FloorNetwork : Floor 9 | | | | | | | | | ✓ | | | | | | | | |
| SideWalk_BusRoute_Transfer : Green Route | | | | | | | | | | | | | | | | | ✓ |
| SideWalk_BusRoute_Transfer : Red Route | | | | | | | | | | | | | | | ✓ | | |
| SideWalk_BusRoute_Transfer : Yellow Route | | | | | | | | | | | | | | | | ✓ | |
| Streets | | | | | | | | | | | | | | ✓ | | | |
| Streets_Walkways_Transfer | | | | | | | | | | | | | | ✓ | | | |
| WalkWays | | | | | | | | | | | | | ✓ | | | | |
| CampusBusStops | | | | | | | | | | | | | ✓ | | ✓ | ✓ | ✓ |
| ExitPoints : Exit from Floor 1 to Outside | ✓ | | | | | | | | | | | | ✓ | | | | |
| ExitPoints : Exit Within Floors | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | | | | |
| ParkingLotEntrance | | | | | | | | | | | | | ✓ | ✓ | | | |

**Fig. 11.9** Connectivity group settings for features classes in the network dataset

## 11.5 Developing a 3D Path-Finding Application

The effectiveness of the proposed data model to represent the 3D network is tested with a 3D path finding application on the UB campus. This application called “Network3D” is created as a toolbar in ArcMap using ArcObjects components (ESRI 2000) and the Visual Basic 6 programming language. The application uses functionality built into the ArcGIS Network Analyst extension to perform network-based analysis such as finding optimal paths, and the ArcGIS 3D Analyst extension for 3D interaction and visualization.

The Network3D application is just one of the many applications that can utilize the data model developed in the chapter. The main objective of the Network3D application is to provide the core functionality for performing 3D optimal path analysis within buildings and in their outdoor surroundings.

The following section reports some of the results derived from path finding analysis involving buildings on the UB campus. It highlights the use and flexibility of such an application in evaluating multiple routing scenarios.

### *11.5.1 Results from Path-Finding Analysis*

#### 11.5.1.1 Determining the Least-Effort Route Inside a Building

Let us consider the query for the least-effort path between two offices in the same building. Figure 11.10 shows the marked positions of an office on the

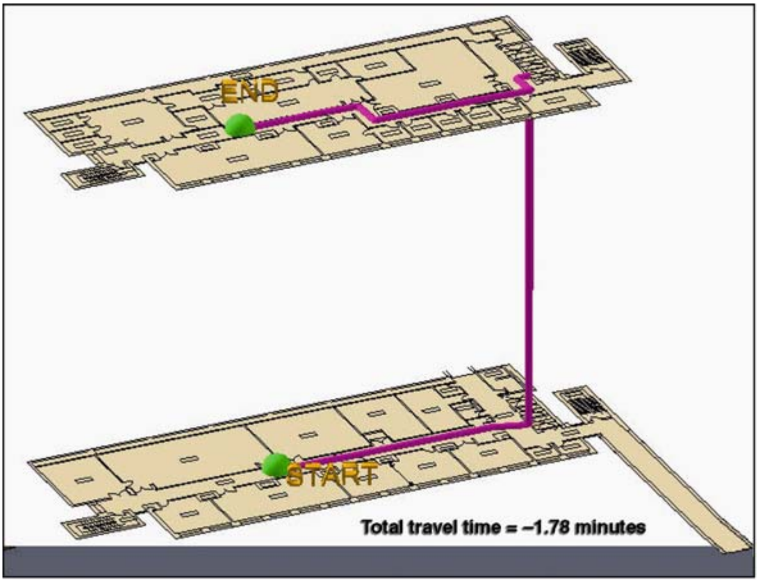



**Fig. 11.10** A least-effort route between two locations in a building

second floor of Furnas Hall and of another office on the ninth floor of this building. Also shown is the least-effort path between the said locations found by Network3D. The path is that of least effort in the sense that it minimizes the total travel time required to reach the destination. Although the destination is closer (in terms of distance) if accessed via the staircases on the left of the view, the optimal route uses the elevator at the other end of the building, which makes for a more comfortable trip over a span of seven floors. Given the assumed specification of impedances, the total travel time for this least-effort route is 1.78 min.

If should be pointed out that the optimal route does not necessarily entail taking an elevator. For instance, if we seek to determine a least-effort path between two adjacent floors (for example floors 2 and 3), results are radically different, as shown in Fig. 11.11. The indicated route uses the staircases since travel along one floor is almost identical using the stairs or the elevator, with a slight advantage for the former.

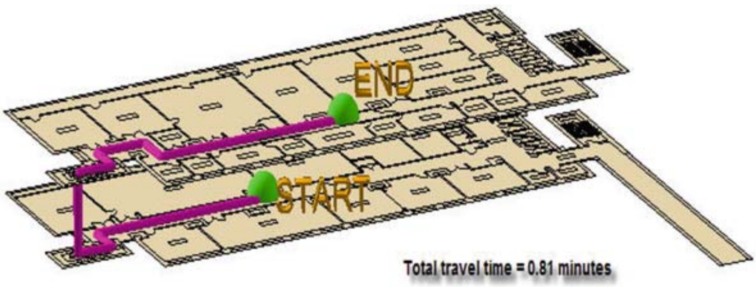



**Fig. 11.11** A least-effort route between two other locations in a building

#### 11.5.1.2 Determining a Least-Effort Route Between Two Buildings

Figure 11.12 shows a least-effort path between a location on the sixth floor of Furnas Hall and another location on the first floor of the nearby building called Bell Hall. This query requires that part of the path exists the perimeter of the buildings. The optimal route found by Network3D uses elevators as it has to traverse 5 floors to reach the nearest building exit on the first floor. The route follows the sidewalks and crosswalks after exiting the first building and then enters the second building through the nearest door and continues till the destination. The total travel time for the route is 2.46 min.

#### 11.5.1.3 Determining an "All-Inside" Least-Effort Route Between Two Buildings

As a variation on the previous query, let us assume that the query now involves finding an optimal route that avoids outdoor spaces. Figure 11.13 shows the least-effort route between the same two locations as used in Fig. 11.12, but

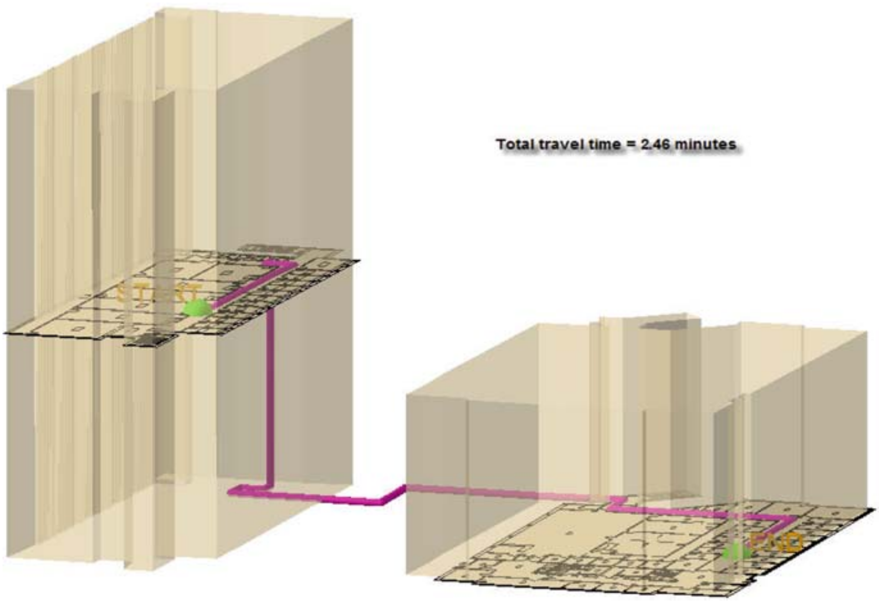



**Fig. 11.12** A least-effort route between two buildings

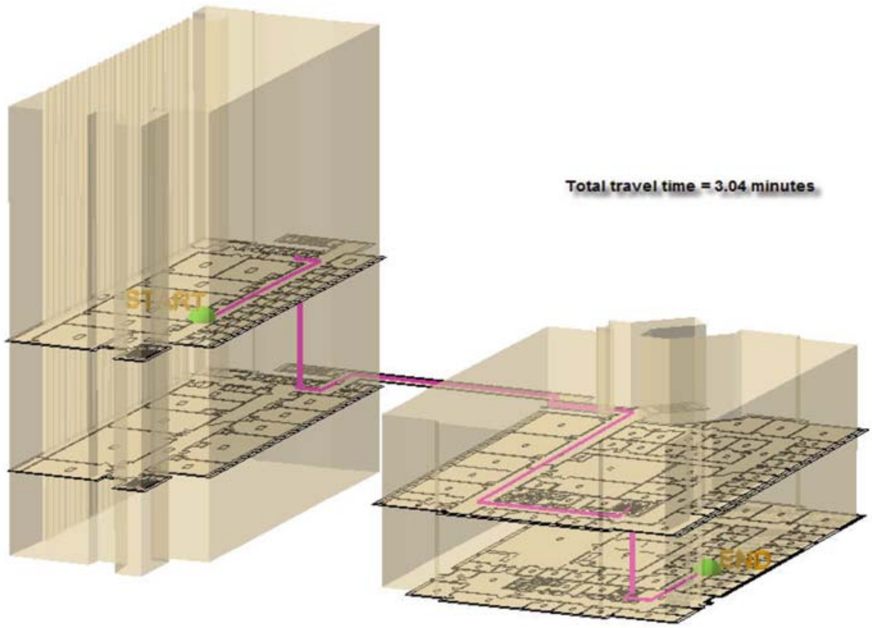



**Fig. 11.13** An all-inside least-effort route between two buildings connected by an enclosed walkway

provides a route that does not travel to the outside of each building. This is possible because there is an enclosed walkway that connects the second floor of Furnas Hall to the third floor of Bell Hall. Such enclosed walkways (or overpasses) exist throughout the academic area of the UB campus to facilitate convenient and sheltered travel during cold winter months. The identification of such interior routes is achieved by using the “InsideOnly” network restriction attribute while performing the optimal path analysis. The restriction attribute restricts the travel along walkways, streets and bus routes that connect the buildings externally and hence the solver is forced to find an “all-inside” optimal route. The total travel time for the route is 3.04 min. It should be pointed out that the constraint that the path be entirely interior could be transformed into a dual problem that seeks to identify the route that minimizes the journey outside buildings.

## 11.6 Conclusions

Using an object-oriented GIS data model, the chapter presented a methodology to model enhanced route planning in urban environments that supports personal movements inside buildings in 3D, while involving multiple transportation modes outside buildings. The deficiency of existing COTS GIS in modeling true 3D network connectivity was overcome by the innovative use of existing 2D network modeling constructs such as network datasets provided by the COTS GIS. Our solution is akin to a 2.5D approach.

The movement within the buildings was conceptualized as a network of interconnected floors made up of corridors, staircases and elevators. The vertical connectivity between floors of a building through elevators and staircases was achieved by placing each floor sub-network formed of corridor lines in a separate connectivity group and allowing the network features in different connectivity groups to connect only at access points. The inconvenience involved in using staircases for traveling many floors in the buildings was modeled using turn features between two access points with impedances indicating the effort required to travel between floors in units of time. The realistic use of different transportation modes commonly used for navigating in an urban environment was achieved by modeling a multimodal network with appropriate cost-based and restriction network attributes and by having suitable transfer points representing the modal transfers.

The data model developed in this chapter provides a platform to perform many transportation and network-based types of analysis. Spatial analysis has so far largely failed to recognize that urban environments function in dimensionality greater than two, and that the most economically and culturally vibrant urban spaces of contemporary human societies are three-dimensional. Hence how space is perceived and used by urban dwellers and visitors should be studied in all three dimensions, rather than collapsed to a 2D surface. It is our contention that accessibility analysis and location analysis in particular present great opportunity for future expansion and use of the routable network data model presented in this chapter.

Because of its scalability, our data model can be extended and modified to suit specific analyses by modeling additional transportation modes such as a subway system, and indoor accessibility facilities such as escalators and wheel chair ramps. The default average travel time based impedances used throughout the data model can be substituted with a more accurate measure of travel time or another measure of travel separation for achieving more realistic results.

There is a need for implementing topological relationships of connectivity, adjacency and containment in true 3D within a COTS GIS for many application domains. The chapter described a method to represent connectivity in 3D using readily available 2D topological data structures. But other topological relationships such as adjacency and containment remain difficult to implement using the 2D data structures of COTS GIS. Hence further research is required in formulating and more particularly, in designing and implementing efficient data structures to store true 3D topological relationships.

Future work should also be geared towards extending the indoor/outdoor network data model, which is currently designed for pre-trip route planning situations, to handle way finding considerations. In particular, indoor and outdoor environments are known to be comprehended quite differently, partly due to their different granularity (Stoffel, Lorenz and Ohlbach 2007, Rehrl et al. 2005). Hence the approach should allow for different relevance of information and different semantics in a way that respects the sensitivity of decision making and perceptions to contextual factors (Dao and Thill 2009, Yao and Thill 2005).

## References

ALIS (2006). Accident location information system data: NYS streets, New York State Office of Cyber Security and Critical Infrastructure Coordination (CSCIC), Albany, NY. http://www.nysgis.state.ny.us/gisdata/inventories/details.cfm?DSID = 932, Accessed on: March 5 2006, Last Modified: October 7 2006

Choi, K. & Jang, W. (2000). Development of a transit network from a street map database with spatial analysis and dynamic segmentation. In J.-C. Thill (Ed.), *Geographic information systems in transportation research* (pp. 129–146). (Oxford: Pergamon)

Curtin, K., Noronha, V., Goodchild, M. & Grisé S. (2003). ArcGIS transportation data model (UNETRANS), ESRI Inc. and Regents of the University of California, Redlands, CA, http://www.geog.ucsb.edu/~curtin/unetrans/TransAttributes03.zip, Accessed: October 7 2006

Dao, T.H.D. & Thill, J.-C. (2009). Contextualized space-time accessibility modeling and measurements with neurofuzzy proximity relations. *Proceedings of the 2009 International Conference on Advanced Geographic Information Systems & Web Services* (GEOWS) pp. 158–164

Dueker, K.J. & Butler, J.A. (2000). A geographic information system framework for transportation data sharing. *Transportation Research Part C: Emerging Technologies*, 8(1), 13–36

ESRI (2000). Developing applications with ArcInfo, an overview of ArcObjects An ESRI technical white paper, http://support.esri.com/index.cfm?fa = knowledgebase.whitepapers.viewPaper&PID = 43&MetaID = 296, Accessed: May 25 2006, Last Modified: July 6 2000

ESRI (2006). Creating a map topology: ArcGIS 9.1 Desktop Help Online, http://webhelp.esri.com/arcgisdesktop/9.1/index.cfm?ID = 1683&TopicName = Creating%20a%20map%20topology&rand = 241&pid = 1677, Accessed: June 17, 2008, Last Modified: January 25, 2006

Fohl, P., Curtin, K.M., Goodchild, M.F. & Church, R.L. (1996). A non-planar, lane-based navigable data model for ITS. In M.J. Kraak & M. Molenaar (Eds.), *Proceedings of the Spatial Data Handling Conference* (pp. 7B/17–29), Delft, The Netherlands, August 12–16

Fujiyama, T. & Tyler, N. (2004). An explicit study of walking speeds of pedestrians on stairs. Centre for transportation studies, University College London, London, UK

Goodchild, M.F. (1998). Geographical information systems and disaggregate transportation planning. *Geographical Systems*, 5, 19–44

Graf, M. & Winter, S. (2003). Netzwerk-Voronoi-Diagramme. Österreichische Zeitschrift für Vermessung und Geoinformation, 91(3), 166–174. (English translation "Network Voronoi Diagrams." available at http://www.geom.unimelb.edu.au/winter/pubs/graf03network.pdf)

Groger, G. & Plumer, L. (2003). Exploiting 2D concepts to achieve consistency in 3D GIS applications, *Proceedings of 11th ACM International Workshop on Geographic Information Systems* (ACM – GIS 03), New Orleans, Louisiana, 78–85

Gupta, A.K. & Yadav, P.K. (2004). SAFE-R: a new model to study the evacuation profile of a building. *Fire Safety J.,* 39, 539–556

Lee, J. (2004). A spatial access oriented implementation of a topological data model for 3D urban entities. *Geoinformatica*, 8, 237–264

Lee, J.Y.S. & Lam, W.H.K. (2006). Variation of walking speeds on a unidirectional walkway and on a bidirectional stairway. *Transportation Research Record,* 1982, 122–131

Li, X. & Lin, H. (2006). A trajectory-oriented carriageway-based road network data model, Part 1: background. *Geo-spatial Information Science*, 9(1), 65–70

Miller, H.J. & Shaw, S.L. (2001). *Geographic information systems for transportation*. (New York: Oxford University Press)

Miller, H.J. & Storm, J.D. (1996). Geographic information system design for network equilibrium-based travel demand models. *Transportation Research Part C,* 4, 371–389

Miller, H.J., Storm, J.D.& Bowen, M. (1995). GIS design for multimodal network analysis. *Proceedings of GIS/LIS'95*, 750–759

Raper, J.F. & Kelk, B. (1991). Three-dimensional GIS. In D.J. Maguire, M.F. Goodchild, & D.W. Rhind (Eds.), *Geographical information systems: principles and applications* (pp. 299–315), Vol. 1. (London: Longman)

Rehrl, X.K., Leitinger, S., Bruntsch, S. & Mentz, H.J. (2005). Assisting orientation and guidance for multimodal travelers in situations of modal change. *Proceedings of the 8th International IEEE Conference on Intelligent Transportation Systems* (pp. 406–412), Vienna Austria, September

Rietveld, P. (1997). Policy aspects of interconnectivity in networks. In C. Capriani & P. Rietveld (Eds.), *Networks in transport and communication: a policy approach* (pp. 177–192). (Aldershot: Avebury)

Southworth, F. & Peterson, B. (2000). Intermodal and international freight network modeling. *Transport. Res. C-Emer*, 8, 147–166

Stoffel, E., Lorenz, B. & Ohlbach, H. (2007). Towards a semantic spatial model for pedestrian indoor navigation. In J. Hainaut, & E. Rundensteiner (Eds.), *ER '07 workshops / advances in conceptual modeling* (pp. 328–337). (Heidelberg: Springer)

Strakosch, G.R. (1985). *Vertical transportation: elevators and escalators*. (New York: Wiley)

Thill, J.-C. (2000). Geographic information systems for transportation in perspective. *Transport. Res. C-Emer*, 8, 3–12

Thill, J.-C. (2009). Transportation applications of geographic information systems. In M. Madden (Ed.), *Manual of geographic information systems* (pp. 1035–1049). (Bethesda, MD: ASPRS)

Yao, X. & Thill, J.-C. (2005). How far is too far? A statistical approach to context-contingent proximity modeling. *Transactions in GIS*, 9(2), 157–178

Zhan, F.B. & Noon, C.E. (1998). Shortest path algorithms: an evaluation using real road networks. *Transportation Science*, 32(1), 65–73

Zhan, F.B. & Noon, C.E. (2000). A comparison between label-setting and label-correcting algorithms for computing one-to-one shortest paths. *Journal of Geographic Information and Decision Analysis*, 4(2), 1–11

Zhu, Q. & Li, Y. (2008). Hierarchical, lane-oriented 3D road network model. *International Journal of Geographical Information Science*, 22(5), 479–505

# Part V
# Analyzing and Modeling Urban Grown, Urban Changes and Impacts

# Chapter 12
# Integration of Remote Sensing with GIS for Urban Growth Characterization

**Xiaojun Yang**

**Abstract** This chapter explores the utilities of remote sensing and GIS integration for urban growth characterization by using a loose-coupling approach. Central to this approach is a time series of satellite imagery for detecting urban land changes. The change detection output is combined with other biophysical and socio-economic data to examine some causal factors leading to the observed changes. The historical urban extent data derived by remote sensing are used to calibrate a dynamic model, which is further applied to assess the spatial consequences of future urban growth under different scenarios. This approach has been applied to a case study site, Atlanta, a typical postmodern American metropolis having undergone rapid demographic and economic growth during the past several decades. Remote sensing-based change detection reveals far-reaching suburbanization in Atlanta during 1973–1999, as indicated by large urban land area growth that substantially outpaced the rate of population growth, as well as more scattered urban spatial patterns. The growth and change revealed by remote sensing are found to be highly correlated with population and economic growth and accessibility conditions. Future urban growth simulations suggest that numerous edge cities would emerge and smaller ones would coalesce together to form larger urban clusters. If current growth patterns do not alter, the process of suburbanization would deplete vegetation and open space by around 2030. The research approach reported here contrasts with many other urban growth studies that focus on the remote sensor data analysis without enabling GIS-based spatial analysis and modeling to go beyond data description and exploration into areas of simulation and prediction.

**Keywords** Remote sensing · GIS · Dynamic modeling · Suburbanization · Landscape changes · Driving force analysis · Scenario development and simulation · Atlanta

---

X. Yang (✉)
Department of Geography, Florida State University, Tallahassee, FL 32306, USA
e-mail: xyang@fsu.edu

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_12, 

## 12.1 Introduction

The study of urban growth and landscape change has received more attention than ever, mainly due to several major reasons. Firstly, information on urban growth and landscape change is critical among decision makers, planners, and local land users who rely on such information to assess the physical extent, characteristics, and consequences of past and future urban development (e.g. Yang 2007). Secondly, urban landscape composition and structure have been an important subject for urban scholars who study the geometry of urban spatial form and function in order to develop theories and models of urban morphology (e.g. Batty and Longley 1994; Longley 2002; Lo 2004). Thirdly, urban growth and landscape change have been considered to be primary boundary conditions for many environmental models. Specific to urban climate models, landscape composition and structure data are needed in order to calculate the areally-averaged surface temperatures as well as the areal values of surface emittance and moisture availability (e.g. Lo et al. 1997; Lo and Quattrochi 2003). Lastly, urbanization has been considered as a major form of global change, given the fact that more than half of the global population is now residing in cities and urban areas are the home of major global production and manufacture centers (Kaplan et al. 2009). Therefore, the study of urban growth and landscape change has been incorporated into various programs within the global environmental change research framework (e.g. Auch et al. 2004).

Characterization of urban growth and landscape change involves the procedures of monitoring and modeling, which further require reliable an information base and robust analytical techniques. Conventional field-based survey and investigation methods are still valid but often logistically constrained. Remote sensing and GIS, given their cost-effectiveness and technological soundness, have increasingly been used to develop useful sources of information supporting decision making in various urban applications (e.g. Yang 2003, 2005; Weng and Quattrochi 2007). Nevertheless, the urban environment, because of its complex and dynamic landscapes, challenges the applicability and robustness of remote sensing and geospatial technologies (Jensen and Cowen 1999). Encouragingly, recent innovations in data, technologies, and theories in the wider arena of remote sensing and geospatial technologies have permitted us with invaluable opportunities to advance the studies on the urban environment.

This chapter explores the utilities of remote sensing and GIS integration for urban growth and landscape change characterization. It will discuss technological needs to support urban growth and landscape change characterization, and examine different approaches that can be used for geospatial technological integration. Then, it will present a case study focusing on a large metropolis.

## 12.2 Integrating Remote Sensing and GIS for Urban Growth Research

Characterizing urban growth and landscape change involves the integration of various geospatial technologies to form a technical framework for data acquisition, processing, analysis, and synthesis (Yang 2007). These technologies include, but are not limited to, remote sensing and image analysis, GIS, and spatial analysis and modeling. Remote sensing, through sensors mounted on various airborne or space-borne platforms, provides the most important source of data for mapping physical and cultural attributes of the urban landscape that can be used to monitor progressive urban development (Paulsson 1992; Jensen and Cowen 1999; Yang 2002). Deriving useful information from remote sensor imagery can be realized through a collection of digital analysis techniques, such as image interpretation, image classification, image transformation, and change detection (Jensen 2005). GIS offers the power to integrate biophysical and socio-economic data, which can help to understand the forces driving urban growth and development (LaGro and DeGloria 1992; Foresman et al. 1997; Nedović-Budić 2002; Burgi et al. 2004). Spatial modeling technologies depict complex structures of the objects or events using mathematical equations that can be used to explore the inherent dynamics of complex urban systems and assess different scenarios for future urban growth (e.g. Alberti 1999; Clarke et al. 2002; Cheng and Masser 2003; Yang and Lo 2003; Pijanowski et al. 2005). The capabilities of these geospatial technologies can be substantially reinforced when linking them together.

There are two major reasons for this integration. The current development of GIS has been more successful in data inventory and information management than in spatial analysis and modeling. GIS is highly dependent upon other existing sources of data, particularly the census datasets. It is difficult to provide updated information without input data from remote sensor imagery. In order to analyze urban dynamics, GIS needs to be linked with remote sensing and image analysis. On the other hand, with the development of GIS, some advanced utilities for spatial analysis are available. Nevertheless, we still lack robust analytical and modeling tools essential for scientific research. In order to understand complex urban dynamic changes, GIS needs to be linked with other spatial analysis and modeling packages.

Basically, there are three possible approaches to integrate relevant geospatial technologies for observation and measurement, data integration and analysis, and modeling: full, tight, or loose integration. Each approach has its advantages and disadvantages (Goodchild et al. 1992). The full integration approach is to integrate techniques of image analysis, spatial analysis, and dynamic modeling fully within a GIS software package. The advantage of this approach is that the software vendors would take care of programming and technical supports. The major disadvantage is that the current GIS market demands heavy database construction and information management, thus constraining the development of sophisticated data analysis and modeling functions within the software. Therefore, it is not realistic to expect GIS software vendors to adopt this approach without substantial pressure from the market for such facilities. The tight integration approach is to embed certain data analysis, dynamic modeling, and interactive visualization utilities with a GIS software package using GIS macro, script, or conventional programming (e.g. Batty and Xie 1994a,b; Yeh and Li 2003; Pijanowski et al. 2005). As contemporary GIS software has become quite flexible and user-friendly, user-customized applications can be tightly coupled through various programming facilities. However, the tight integration through GIS programming languages is challenged by the difficulty in dealing with various data structures. Because of this limitation, some sophisticated utilities may not be easily coupled. In addition, users are overly burdened with the technical task of programming. The loose integration involves packages or modules of standard GIS, image processing, statistics, or dynamic modeling. These geospatial technologies can be integrated by the interchange of ASCII files between packages or using a common binary file format (e.g. Landis 1995; Clarke and Gaydos 1998). Since no extra software needs to be written, users can avoid heavy programming tasks. In addition, users can make use of the full the ranges of the sophisticated utilities available from different software packages. The drawback of this approach is that data conversions from different packages can be tedious and error prone.

Although the need of integrating remote sensing, GIS and dynamic modeling has been widely recognized, no consensus has been agreed on how such integration should be achieved. All three approaches suffer from different drawbacks. Nevertheless, they provide certain solutions for various applications. This study will adopt a loose coupling strategy to link remote sensing, GIS, and spatial modeling in order to take full advantage of different methods and technologies.

## 12.3 The Case Study Site

A case study focusing on the city of Atlanta has been conducted to demonstrate the utilities of remote sensing and GIS integration for urban growth and landscape change characterization. The study site includes the 10 counties under the Atlanta Regional Commission (ARC) as well as three additional counties, i.e. Coweta, Forsyth, and Paulding, which have shown similar growth pattern as the ARC counties (Fig. 12.1).

For the past three decades, Atlanta has been one of the fastest growing metropolises in the United States as it emerged to become the premier commercial, industrial, and transportation urban center of the southeast. Population increased 27% between 1970 and 1980, 33% between 1980 and 1990, and 40% between 1990 and 2000. The city has expanded greatly as suburbanization consumes large areas of agricultural and forest land adjacent to the city,

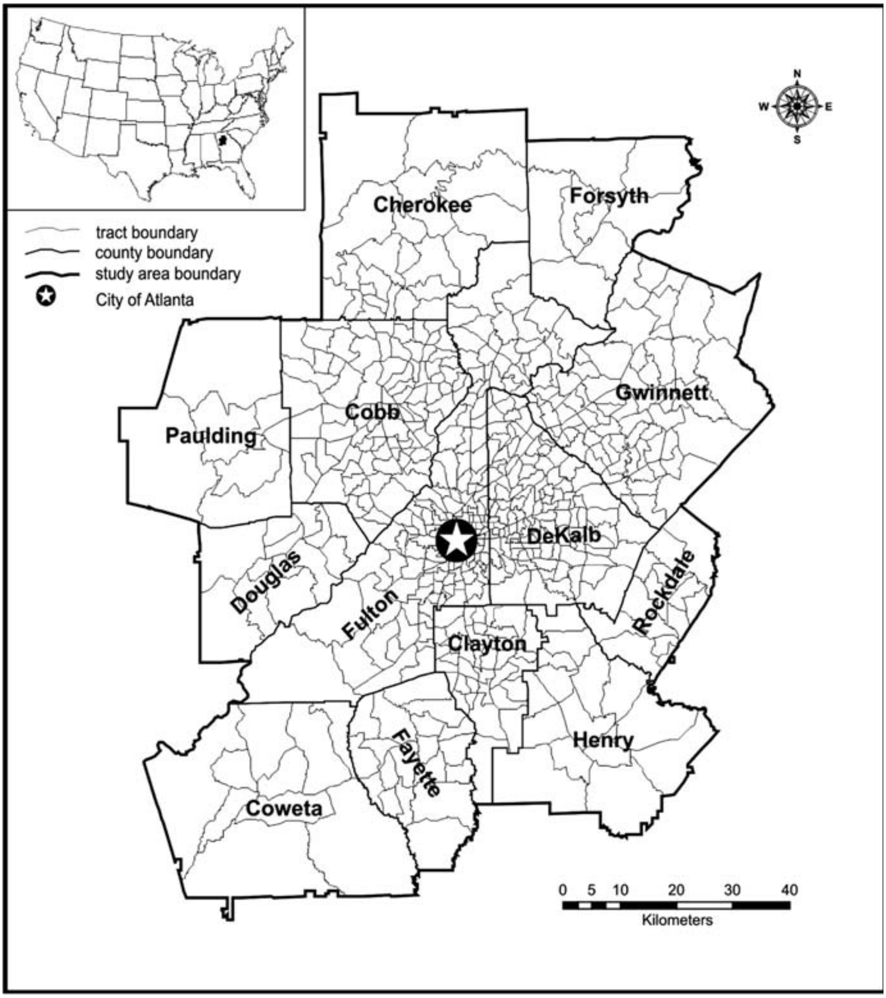



**Fig. 12.1** Location of the study site. It consists of the ten counties under the Atlanta Regional Commission (ARC) and three additional counties (Forsyth, Paulding, and Coweta) that show a similar growth pattern. The city of Atlanta is shown

pushing the periurban fringe farther and farther away from the original urban boundary. This rampant suburban sprawl has provoked concerns over losses of large areas of primary forests, inadvertent climate repercussions, and the degradation of the quality of life in this region (Bullard et al. 2000; Lo and Quattrochi 2003). By using population trends, land use, traffic congestion, and open space loss, Sierra Club's 1998 Annual Report ranked Atlanta as America's most sprawl-threatened large city (SIERRA CLUB 1998).

On the other hand, because of the significant physical growth, Atlanta's urban spatial structure and pattern have changed dramatically (Yang 2002), making the city a 'hot spot' in urban studies. Urban geographers have recognized Atlanta as one of the few typical postmodern metropolises in North America (Hall 2001). Soja (1989) first used the term "postmodern" to describe cities that have undergone restructuring in the United States after the rise of post-Fordist industrial organization, which is characterized by a flexible subcontracted production system based on small-size and small-batch units. The postmodern city exhibits unique urban forms, architectural styles, and socioeconomic characteristics. Architecturally, it has multi-centered business districts with office glass towers and stylish buildings. Socially, it is increasingly minoritized and polarized along class, income, racial and ethnic lines (Dear and Flusty 1998). Los Angeles is in fact the typical postmodern city, and many cities in the United States Sunbelt are growing in the fashion of Los Angeles. Atlanta is a good example of such a city because it exhibits all the characteristics of the postmodern city described above.

Research on Atlanta's internal structure has led to the formulation of the urban realms model to depict the multi-nuclei nature of the city in contrast to the conventional single-core urban form (Hartshorn and Muller 1989). Thus, Atlanta is an ideal city to study postmodern urban dynamics and environmental consequences of accelerating urban growth. Over the past decade, the author has been involved in various research projects aiming to understand the dynamics of change in Atlanta by using geospatial technologies. The case study will examine urban growth and landscape change dynamics. Specifically, this study will have three major components: (1) To examine urban spatial growth by using a satellite time series; (2) To analyze the major forces driving the observed growth and change; and (3) To imagine, test, and choose between two possible future urban growth scenarios.

## 12.4 Data Acquisition, Processing and Analysis

The research methodology adopted here was actually based on a loose integration strategy, which consists of several major components: data acquisition and assemblage, image processing and change detection, spatial analysis, and predictive modeling and simulation (Fig. 12.2).

### *12.4.1 Data Acquisition and Assemblage*

A time series of Landsat images were used as the primary data to measure spatio-temporal urban growth pattern at 6–8 year intervals, beginning in 1973. The Landsat Multispectral Scanner (MSS) data were used for the period before 1982, when Thematic Mapper (TM) data are not available. After that period, TM and Enhanced Thematic Mapper Plus (ETM+) data were used. Eight predominantly cloud-free images were acquired by the three sensors for Atlanta

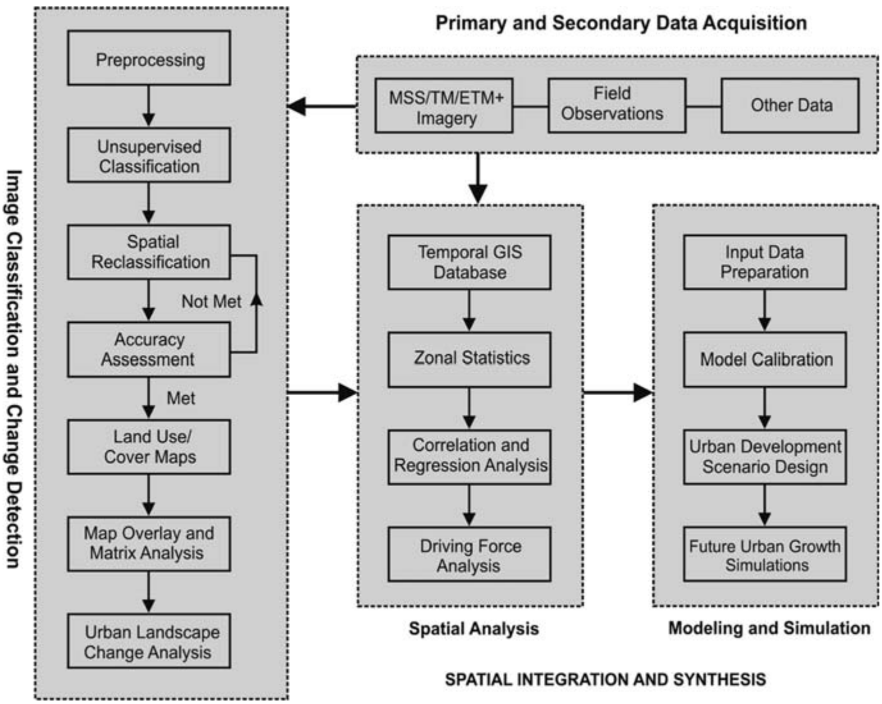



**Fig. 12.2** Flowchart of the technical framework adopted. It comprises several major components: primary and secondary data acquisition, image classification and change detection, spatial analysis, and predictive modeling and simulation

during 1973–1999 (Table 12.1). Most of the scenes were acquired during the spring or earlier summer seasons, when vegetation is in the stage of vigorous growth. The 1998 and 1999 scenes are the two exceptions. The 1998 TM scene, acquired in the winter season, was used to improve vegetation mapping. The ETM + scenes were acquired in late summer because they were the only scenes free from clouds available between April and September 1999. Because of good image quality, the 1988 MSS scene was used as the reference image for relative radiometric normalization of other MSS images and no classification was attempted for this scene because a higher resolution scene for 1987 was available.

To support change detection, driving force analysis, and predictive modeling, a variety of existing data were assembled, which include (Table 12.1): digital images of Advanced Thermal and Land Applications Sensor (ATLAS) on May 11, 1997; contact prints of aerial photographs for 1986–1988; 1993 USGS digital orthophotos; the 1988–1990 land-cover classification from Georgia Department of Natural Resources; USGS 7.5-min Digital Elevation Model (DEM); US census survey data for 1980, 1990 and 2000; and major road networks from the AND Global Database. In addition, GPS-guided field

**Table 12.1** List of the major themes in the spatial database

| Group | No. | Theme | Date or year | Source(s) | Format | Purposes[a] |
|---|---|---|---|---|---|---|
| | 1 | MSS images | 4-13-1973, 6-11-1979, 5-14-1988 | | | |
| 1 | 2 | TM images | 6-29-1987,7-31-1993 | USGS EROS data center | | A |
| | | | 7-29-1997, 1-02-1998 | | | |
| | 3 | ETM+ image | 9-09-1999 | | | |
| | 4 | High-density urban | | | | B,C,D |
| | 5 | Low-density urban | | | | |
| 2 | 6 | Cultivated/exposed land | 1973, 1979, 1987, 1993, 1999 | Classified from Group 1 | | B,D |
| | 7 | Cropland/grassland | | | | |
| | 8 | Forest | | | Rasters | B,D |
| | 9 | Water | | | | |
| 3 | 10 | Elevation | | USGS 7.5′ DEM | | C |
| | 11 | Slope | 1970s | Computed from 11 | | C, D |
| | 12 | Hillshaded relief | | | | D |
| | 13 | Excluded areas | 1973 and 1999 | Miscellaneous | | D |
| 4 | 14 | Total population | 1973, 1987, 1999 | Interpolation from 1980 and 1990 census data | | C |
| | 15 | Per capita income | | | | |
| | 16 | Cities | 1990 | US Bureau of Census | Points | C |
| | 17 | Roads | 1973, 1987, 1999, 2025 | AND global database updated with satellite imagery | Lines | |
| 5 | 18 | Major highway exits | | (Group 1); 2025 ARC Transportation Plan | Points | C,D |
| | | | 1973, 1987, 1999 | ESRI Database updated with satellite imagery | | |
| | 19 | Large shopping malls | | (Group 1) | Polygons | |
| | 20 | Census-tract boundary | | | | |
| 6 | | | 1980, 1990, 2000 | US Bureau of Census | | B,C,D |
| | 21 | County boundary | | | | |

[a]For purposes, A = land use/cover classification; B = landscape pattern and change analysis; C = urban landscape change driving force analysis; and D = urban growth simulation and predictive modeling.

surveys were conducted to help establish the relationship between image signals and ground conditions.

### *12.4.2 Image Processing of Remotely Sensed Data*

The image processing procedures adopted include preprocessing, classification, spatial reclassification, and accuracy assessment. Both geometric rectification and radiometric normalization were attempted for image preprocessing. The georeferencing strategy adopted was actually an image-to-image registration. The 1997 TM image had been accurately georeferenced by Space Imaging EOSAT, which was further used as the reference to rectify other scenes. The 1973 and 1979 MSS images acquired by two earlier satellites are very different in contrast, despite the identical processing conducted. To help restore a common radiometric response among them, the relative radiometric normalization method developed by Hall et al. (1991) was applied to the two MSS images by using the 1988 MSS scene as the reference. Radiometric normalization was not attempted to the Landsat TM images because they have been processed to good radiometric quality.

The 1973, 1979, 1987, 1993 and 1999 images were classified by using a two-step unsupervised method. Firstly, the Iterative Self-Organizing Data Analysis (ISODATA) algorithm was used to identify spectral clusters from image data, excluding the thermal band for the TM and ETM+ scenes. Then, the resultant clusters were assigned into one of the six land use/cover classes: high-density urban use (commercial and industrial buildings and large open transportation facilities); low-density urban use (predominately residential); cultivated/exposed land (areas with sparse vegetation cover); cropland/grassland (grasses, other herbaceous vegetation and crops); forest land (coniferous, deciduous, and mixed forests); and water.

Reclassification procedures were used to reduce the two types of misclassification errors on the initial maps produced through the unsupervised classification. The first type consisted of boundary errors due to the occurrence of spectral mixing within a pixel, which has been suppressed by using a modal filter. The second type consisted of spectral confusion errors due to the spectral similarity among several land classes, which is inevitable for an image acquired with a broad-band sensor and tends to be more perceptible in an urban scene than in a rural one. Defining the spectral confusions involves the use of spatial and contextual properties. To this end, an image interpretation method was employed, because it allows the combined use of spectral and spatial contents, as well as human wisdom and experience, thus providing the most powerful means for spectral confusion identification. Image interpretation can be incorporated effectively into a digital classification procedure with the use of on-screen digitizing, multiple zooming, Area Of Interest (AOI) functionality, and other relevant GIS tools, such as overlaying and recoding. In addition, several image processing programs permit advanced tools for geoprocessing through which some "manual" operations can be implemented automatically. Spectrally confused clusters were first identified, and AOI layers were created by on-screen digitizing. The AOI layers served as masks for splitting confused clusters. Finally, GIS reclassification functionality was employed to recode the split clusters into correct land classes. This was an interactive process until acceptable accuracy was obtained.

Due to the limited availability of reference data, it is impossible to perform accuracy assessment for each map exhaustively. The accuracy assessment strategy adopted here was to assess each type of imagery covering the study area through a standard method described by Congalton (1991). Results revealed that the land use/cover maps classified from the TM or ETM+ images yielded slightly better accuracies than those from the MSS images. Overall, all these maps met the minimum 85% accuracy stipulated by Anderson et al. (1976), indicating that the image processing procedures adopted here have been effective in producing compatible land use/cover data over time, despite the differences in spatial, spectral, and radiometric resolution of the three generations of Landsat data used in this project.

### *12.4.3 Change Detection*

Change detection focuses on urban growth and the nature of change. To analyze urban growth, the spatial distribution of each of the two major urban land classes was extracted from each map in the time series. The change in urban land was summarized by using the GIS minimum dominate overlay method, which allows the smallest amount of high-density or low-density urban use in the earliest year to show up fully, while only the net addition in the following years in the time series is shown. In this way, the urban extent of five dates was summarized on one map. By assigning a unique color to the net addition of each year on the combined map, the progressive growth of high-density or low-density urban land can be perceived. The statistical summary of urban addition for each period was also generated.

The nature of change was to analyze land conversions using cross-tabulation or matrix analysis that assigns a unique class for each coincidence in two input layers, thus capturing different combinations of change. This method was used to characterize the land conversions for the periods 1973–1987 and 1987–1999. Given the total number of land classes, there are 16 possible combinations for each period. Because this study focused on the conversion of forest, cropland/grassland, or cultivated/exposed land into urban uses, only nine combinations were selected for further analysis, while the other were merged into a single unit. These combinations of conversion are: cultivated/exposed land into high-density urban use (C1); cultivated/exposed land into low-density urban use (C2); cropland/grassland into high-density urban use (C3); cropland/grassland into low-density urban use (C4); forest into high-density urban use (C5); forest into low-density urban use (C6); and cultivated/exposed land or cropland/grassland or forest into water (three conversions here combined into one) (C7). C0 is for all other combinations which are not considered here.

### *12.4.4 Spatial Statistical Analysis*

The purpose of spatial statistical analysis was to analyze the forces driving the observed changes. This necessitated the incorporation of additional data into the analysis. A total of 12 data layers were prepared, which are grouped into five major categories (Table 12.1):

1. *Statistical boundaries*. The 1980 and 1990 county and census-tract boundaries for the study area were extracted from the 1992 and 1995 TIGER street centerline files.
2. *Land use/cover data*. Four classes, namely, high-density urban, low-density urban, cropland/grassland, and forest, for 1973, 1987 and 1999, were used here.
3. *Topographic measures*. Terrain elevation and slope were derived from USGS 7.5-min DEM.
4. *Total population and per capita income*. These two measures were estimated for 1973, 1987 and 1999 by using data from several census years. This estimation was based on the assumptions that the human population and per capita income grew exponentially and their rates of increase were constant within two immediate census years. At the county level, the total population and per capita income for the three specific years were computed through exponential interpolation with data from the four census years, namely, 1970, 1980, 1990 and 2000. At the census-tract level, a dasymetric mapping method described by Langford and Unwin (1994) was adopted to harmonize the original data sets before applying exponential interpolation due to the change of tract boundaries through time. The 1980 and 1990 census-tract boundaries are quite different as the latter includes more than 100 new tracts. Through dasymetric mapping the original 1980 data set was remodeled by using the 1979 low-density urban extent. The 1980 remodeled data were summarized with the 1990 census-tract boundary. This compensated for the difference in tract boundaries between the 1980 and 1990 data. For exponential interpolation, the growth rate for 1980–1990 was used for the periods 1970–1979 and 1990–1999 but adjusted with the overall rates of increase for the entire study area.
5. *Location measures*. Urban center proximity, highway proximity, node point proximity, and shopping mall proximity were generated for the analysis at the tract level. In doing so, a weighting buffer grid was created from urban centers, major highways, nodes, and large shopping malls respectively. Then, each grid was converted into a binary image with 1 as the area within the buffers and 0 as the background.

Spatial statistical analysis was conducted at three different levels of aggregation, the entire study area, county, and census tract. For the entire area, visual analysis is adequate to reveal trends with 3 years of data for one single observation. At the county level, there are 13 observations for 38 variables. At the tract level, there are 444 observations for 38 variables. These variables include: population densities, population density changes, per capita income, and per capita income changes; mean elevation, mean slope, percentages of county or tract in urban center, road, node, and shopping mall buffers; proportions in high-density urban use and low-density urban use as well as their changes over time, all for 1973, 1987, and 1999. Correlation analysis was used to determine if an association exists as well as the magnitude and direction of the significant association. At the tract level, a multivariate regression was also conducted to examine the relationship between two urban class proportions (and their changes) as dependent variables with a group of independent variables. The stepwise variable selection was employed to determine which variable to include in the final model. A total of 12 models were computed, which relate urban uses, terrain conditions, demographic, economic variables, and location measures to one another.

### *12.4.5 Predictive Modeling and Simulation*

The SLEUTH urban growth model (Clarke and Gaydos 1998) was used for predictive modeling and simulation here. It is a cellular automaton model whose behavior is controlled by the coefficients of diffusion, breed, spread, slope resistance, and road gravity. This model was chosen because it is dynamic and future-oriented, conforming to the essential requirement of urban growth simulation in this project. The behavior rules guiding urban growth in the model consider not only the spatial properties of neighboring cells but also existing urban spatial extent, transportation, and terrain slope. The driving force analysis (Sections 12.4.4 and 12.6) found that transportation and terrain conditions were significant factors driving landscape changes in Atlanta. These behavior rules therefore have realistically accounted for the driving forces in the formation of edge cities in Atlanta. Therefore, the SLEUTH model was used here as a tool to imagine, test, and choose between two possible future urban growth scenarios under different environmental and development conditions for Atlanta.

To implement the model, a set of input data were prepared:

1. *Urban extent*. A time series of urban extent layers was derived by combining the two urban classes of the land use/cover maps for 1973, 1979, 1987, 1993, and 1999.
2. *Roads*. The 'road' layers contain not only major road networks but also node points and large shopping malls. The major highways and node points were extracted from the AND global highway database and then updated with the

1973, 1987 and 1999 images. Three layers of large mall polygons were extracted from the above images. A weighting system was established for highways, nodes, and malls, respectively. The layers of highways, nodes, and shopping malls in the same year were combined to form a single ‘road’ layer. In this way, a ‘road’ layer was produced for 1973, 1987, and 1999, respectively. The ‘road’ layer for 2025 was produced by overlaying the 1999 roads with the improved roadways and new roadways proposed for 20205 by Georgia Department of Transportation.

3. *Excluded area for development*. Two layers of excluded areas were assembled. The first layer is a binary image, consisting of water extracted from the 1973 MSS image overlaid with various types of public land. These areas were not allowed for urban development. This layer was primarily used for the model calibration. For future growth prediction, another layer was built, with probabilities of exclusion included. All excluded areas in the first layer were preserved and assigned a value of 100. Additionally, this layer contains three levels of buffer zones around major streams in the study area.
4. *Slope and shaded relief*. They were derived from USGS 7.5 min DEM. Each layer was resampled into 240 m in grid size.

The next step was to calibrate the model for determining the best value of each control coefficient previously mentioned. This involved the use of statistical measures to quantify the spatial fit between the modeled results and historical urban extent data. Due to the time and computational resources constraints, the calibration was broken down into three phases. The coarse calibration was to block out the widest range for each control coefficient. The fine calibration was to narrow down the ranges to approximately 10 or less. The final calibration was to determine the best combination of the control coefficients. One more step was conducted to determine the starting values used for future growth simulation. The final values were: diffusion (71), breed (10), spread (32), slope resistance (73), and road gravity (100).

Two possible planning scenarios for future urban development in Atlanta under different policies and environmental conditions were designed and simulated. The first scenario assumes the factors for growth remain unchanged; thus, it may be termed as ‘continuation’. It provides a benchmark for comparison with the alternative growth strategy. To implement this scenario, the values of growth control coefficients obtained from the model calibration were used as the starting values. The 1999 urban extent data was actually used. The second scenario considers a hybrid growth strategy in which both conventional suburban development and alternative growth efforts are addressed. This scenario also considers environmental conservation by limiting development around several predefined stream buffer zones. To implement these ideas, the starting values for five growth control coefficients used in the first scenario was changed in order to slow down the growth rate and to alter the growth pattern. The actual control coefficients used for the second scenario were: diffusion (100), breed (100), spread (15), slope resistance (40), and road gravity (200). The proposed transportation improvements and new additions were considered here. The simulation of future growth was from 2000 to 2050 for both scenarios.

## 12.5 Urban Spatial Growth

Based on Fig. 12.3 (Upper), the spatial expansion of high-density urban use is clearly visible. In 1973, the high-density urban use was small, occupying only 2.85% of the total land area for Atlanta (Table 12.2). The outward spread is quite clear in the 1979 and 1987 patterns, following the major transportation routes. Between 1973 and 1979, the net addition was 8292 hectares, or 27.90% increment. The net addition was 16,265 hectares, or 42.79% increment, between 1979 and 1987. These additions were primarily concentrated in several inner counties, such as Fulton, Cobb, DeKalb, and Clayton. Significant growth took place by 1993 and 1999. These new development areas were highly concentrated in northern and northeastern areas. The linearly concentrated pattern became more multinucleated. The 1999 distribution shows further enhancement of this transition as the spread took place largely along transportation routes and around urban centers, particularly in some peripheral counties, such as Coweta, Cherokee, Forsyth, Paulding, and Fayette. In 1999, the high-density urban use occupied 87477 hectares, or 8.38% of the total land area, which is about 194.31% increase in land when comparing with 1973. The daily increment was approximately 6 hectares or 15 acres between 1973 and 1999.

The evolution of spatial pattern of low-density urban use, mainly residential, is clearly perceived in Fig. 12.3 (Lower). In 1973, the low-density urban use occupied 76910 hectares, or 7.36% of the total area (Table 12.2). Although the low-density urban land shows signs of spreading outward, its large share was clearly concentrated in the inner city core and several inner counties. A somewhat linear pattern can also be seen along several major transportation routes. Thus, the spatial pattern of low-density urban use in 1973 can be described as a form of concentration mixed with some degree of dispersal. Significant growth occurred in 1979 and 1987, with net addition of 52,264 and 48,651 hectares, respectively (Table 12.2). Most of the new additions occurred outside the central city core, concentrating in four inner counties of DeKalb, Clayton, Fulton, and Cobb, and in three exterior counties of Gwinnett, Rockdale, and Fayette. Growth in the northern, northwestern, and northeastern directions was quite clear. The spatial distribution pattern became more dispersed. Low-density urban use continued to grow after 1987. Most of the new development, however, took place in the exterior metropolis. This widely spread-out pattern is a major indicator of suburbanization. Quantitatively, the low-density urban use occupied 282,959 hectares or 27.10% of the total area in 1999, indicating a

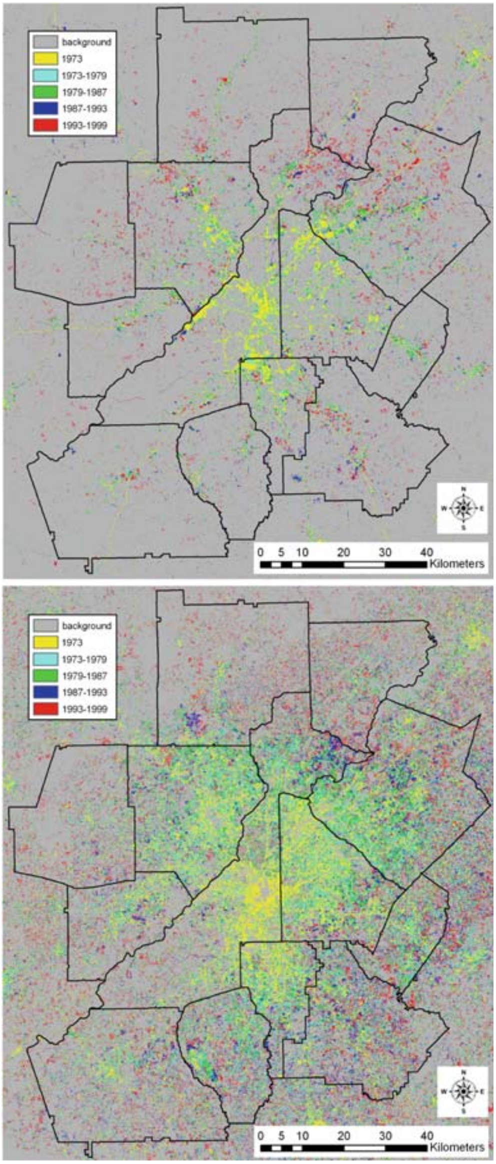



**Fig. 12.3** Urban spatial growth during different periods: the *upper figure* is high-density urban use and the *lower figure* for low-density urban use. Note that the county boundaries are shown. The original figures were in color. Readers can contact Dr. Xiaojun Yang (xjyangusa@yahoo.com) to obtain an original copy

267.91% increase between 1973 and 1999. The daily increment was about 22 hectares or 54 acres for the same period.

Table 12.2 also indicates that the continuing decline in cropland/ grassland and forest. The shrinking pattern of these two classes was proportional to the growth of the two urban classes. In general, the decline of cropland/grassland and forest land predominately took place in the interior metropolis before 1987 but in the exterior after 1987. Quantitatively, cropland/grassland occupied

**Table 12.2** Land use/cover proportions for the 13 counties in Atlanta

| | | 1973 | | 1979 | | 1987 | | 1993 | | 1999 | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| No | Land Use/Cover | Area (ha) | (%) | Area (ha) | (%) | Area (ha) | (%) | Area (ha) | (%) | Area (ha) | (%) |
| 1 | High-density urban | 29722 | 2.85 | 38015 | 3.64 | 54280 | 5.2 | 67633 | 6.48 | 87477 | 8.3 |
| 2 | Low-density urban | 76910 | 7.36 | 129174 | 12.37 | 177825 | 17.03 | 214484 | 20.54 | 282959 | 27.1 |
| 3 | Cultivated/exposed land | 14534 | 1.39 | 20595 | 1.97 | 15511 | 1.49 | 21132 | 2.02 | 5358 | 0.51 |
| 4 | Cropland/grassland | 159345 | 15.26 | 117365 | 11.24 | 117686 | 11.27 | 96700 | 9.26 | 101122 | 9.68 |
| 5 | Forest land | 750366 | 71.85 | 724967 | 69.42 | 663673 | 63.55 | 625984 | 59.94 | 545148 | 52.2 |
| 6 | Water | 13404 | 1.28 | 14166 | 1.36 | 15306 | 1.47 | 18348 | 1.76 | 22217 | 2.13 |
| In total | | 1044281 | 100 | 1044281 | 100 | 1044281 | 100 | 1044281 | 100 | 1044281 | 100 |

159,345 hectares or 15.26% of the total study area in 1973. It declined to 101,122 hectares (or 9.68%) by 1999. This represents a decrease of 36.54%, or a daily rate of 6 hectares (15 acres). Similarly, forest declined from 750,366 hectares (or 71.85%) in 1973 to 545,148 hectares (or 52.20%) in 1999, thus representing a decrease of 27.35%, or a daily rate of 22 hectares (53 acres) in land area.

The nature of change is quite clear from Table 12.3. From 1973 to 1987, the loss of forest land has contributed to the overwhelming share of the growth of the two major urban classes. The high-density urban use had a net addition of 36,384 hectares, among which 62.17% came from the loss in forest land (C5) and 33.24% resulted from the loss in cropland/grassland (C3). The loss in cultivated/exposed land only contributed to 4.59% of the increase in high-density urban use (C1). For the low-density urban use, 70.37% (C6) and 28.18% (C3) of the increase came from the loss in forest land and cropland/grassland, respectively. The loss of cultivated/ exposed land only accounted for 1.45% (C1) of the net addition in low-density urban use. From 1987 to 1999, the magnitude of these conversions generally increased and forest and cropland/grassland conversions still overwhelmed the growth in two major urban uses.

## 12.6 Urban Growth and Landscape Change Driving Forces

For the entire study area, the two urban classes have increased while the cropland/grassland and forest classes have declined since 1973. At the same time, population density and per capita income have rapidly increased, thus suggesting their impact on the urban growth and landscape changes (Table 12.4). The mean elevations for high-density urban use and forest have increased, while those for low-density urban use and cropland/grassland have decreased during the same period. This indicates that the land available for high-density urban development tends to be with higher elevations, while residential-dominated urban development favors relatively lower terrain. The mean slope gradient for each class tends to increase, suggesting that terrain slope gradient has been an influential resistance for urban development.

At finer spatial levels, the results of analysis are much more complicated as the landscape changes and demographic, economic, and terrain characteristics have become more differentiated. The results of analysis for the two major urban classes are discussed below at the county and tract levels.

### *12.6.1 High-Density Urban Use*

At the county level, the proportions of high-density urban use for 1973, 1987, and 1999 were positively correlated with population density, urban center

**Table 12.3** Land conversion statistics for different periods

| Maps[a] | | | Land conversion statistics | | | |
|---|---|---|---|---|---|---|
| | | | 1973–1987 | | 1987–1999 | |
| Year A | Year B | Nature of change code | Area(ha) | (%) | Area(ha) | (%) |
| 3 | 1 | C1 | 1669 | 0.16 | 3029 | 0.29 |
| 3 | 2 | C2 | 1627 | 0.16 | 4709 | 0.45 |
| 4 | 1 | C3 | 12094 | 1.16 | 11085 | 1.06 |
| 4 | 2 | C4 | 31765 | 3.04 | 40127 | 3.84 |
| 5 | 1 | C5 | 22621 | 2.17 | 28586 | 2.74 |
| 5 | 2 | C6 | 79319 | 7.60 | 94559 | 9.05 |
| 3 or 4 or 5 | 6 | C7 | 3653 | 0.35 | 6226 | 0.6 |
| All other combinations | | C0 | 891534 | 85.37 | 855960 | 81.86 |

[a] The land use/cover coding system: 1 = high-density urban, 2 = low-density urban, 3 = cultivated/exposed land, 4 = cropland/grassland, 5 = forest, and 6 = water. Note that approximately 75% of total pixels remain unchanged.

**Table 12.4** Land use/cover, demographic, economic, and topographic measures for the entire study area

| | | | | | 1973–1987 | | 1987–1999 | |
|---|---|---|---|---|---|---|---|---|
| Items | | 1973 | 1987 | 1999 | Change | (%) | Change | (%) |
| Land use/cover Proportion (%) | High-density urban | 2.846 | 5.198 | 8.377 | 2.352 | 82.626 | 3.179 | 61.159 |
| | Low-density urban | 7.365 | 17.028 | 27.096 | 9.664 | 131.212 | 10.068 | 59.122 |
| | Cultivated/exposed land | 1.392 | 1.485 | 0.513 | 0.094 | 6.722 | −0.972 | −65.457 |
| | Cropland/grassland | 15.259 | 11.270 | 9.683 | −3.989 | −26.144 | −1.586 | −14.075 |
| | Forest | 71.855 | 63.553 | 52.203 | −8.302 | −11.553 | −11.350 | −17.859 |
| | Water | 1.284 | 1.466 | 2.127 | 0.182 | 14.190 | 0.662 | 45.152 |
| Per capita income (×1000) | | 5.509 | 18.368 | 29.244 | 12.859 | 233.413 | 10.877 | 59.216 |
| Population density (per hectare) | | 1.614 | 2.331 | 3.159 | 0.717 | 44.404 | 0.828 | 35.543 |
| Mean elevation (100 feet) | High-density urban | 9.578 | 9.626 | 9.663 | 0.048 | 0.504 | 0.037 | 0.381 |
| | Low-density urban | 9.516 | 9.484 | 9.432 | −0.032 | −0.340 | −0.052 | −0.549 |
| | Cropland/grassland | 9.653 | 9.563 | 9.463 | −0.090 | −0.931 | −0.010 | −1.041 |
| | Forest | 9.469 | 9.494 | 9.525 | 0.025 | 0.263 | 0.031 | 0.331 |
| Mean slope (%) | High-density urban | 4.862 | 4.986 | 5.557 | 0.124 | 2.550 | 0.571 | 11.452 |
| | Low-density urban | 5.509 | 5.812 | 6.220 | 0.303 | 5.500 | 0.408 | 7.020 |
| | Cropland/grassland | 5.081 | 5.327 | 5.277 | 0.246 | 4.842 | −0.050 | −0.939 |
| | Forest | 8.095 | 8.303 | 8.542 | 0.208 | 2.569 | 0.239 | 2.878 |

proximity, node proximity, mall proximity, and highway proximity. The change in the high-density urban use proportion from 1973 and 1987 was positively correlated with population density change. Income and terrain factors did not exhibit significant statistical relationships with either the high-density urban use proportion or its change through time.

At the census tract level, more delicate relationships were revealed. The proportions of high-density urban use for 1973, 1987, and 1999 correlated positively with road proximity, node proximity, urban center proximity, population density, and mean elevation, but negatively with mean slope gradient. The variables accounting for 45–49% of the variance in the 3 years' land class proportions often included road proximity, mean elevation, mean slope, per capita income, and population density. The change in high-density urban use proportion from 1973 to 1999 correlated positively with road proximity, node proximity, urban center proximity, mean elevation, and mall proximity, but negatively with mean slope. The independent variables that explained more than 53% of the variance in the class proportion change were population density change, mean slope gradient, mean elevation, mall proximity, urban center proximity, and road proximity.

Statistical analysis indicates that population was more significant than location factors in explaining high-density urban use change at the county level. At the tract level, however, location and terrain conditions were clearly more significant than population. This is because commercial and industrial developments (the two major components of high-density urban use) were attracted to highways, node points, and urban centers.

### *12.6.2 Low-Density Urban Use*

At the county level, the proportion of low-density urban use (mainly residential) for 1973, 1987, and 1999 correlated positively with population density. Other variables that correlated positively with low-density urban use proportion often included urban center proximity, node proximity, mall proximity, highway proximity, and per capita income. The mean slope gradient was found to correlate negatively with the proportion of low-density urban use in 1987 and 1999. The change in low-density urban use proportion from 1973 to 1999 correlated negatively with urban center proximity, mall proximity, node proximity, and highway proximity.

At the tract level, the proportion of low-density urban use for 1973, 1987, and 1999 correlated positively with population density, node proximity, and highway proximity, but negatively with mean slope gradient. Other variables that correlated positively with the low-density urban use proportion for at least one year include per capita income and mall proximity. The variables accounting for 42–71% of the variance in the three years' class proportions include population density (3 years), mean slope gradient (3 years), road proximity (1973 and 1999), mall proximity (1987 and 1999), node proximity (1987), and mean elevation (1973). The change in the proportion of low-density urban use from 1973 to 1999 correlated positively with population density change, but negatively with road proximity and node proximity. The independent variables that explained about 41% of the variability in the class proportion change during the same period were population density change, mean elevation, road proximity, and mean slope gradient.

Population has been the most important factor for explaining changes in the proportion of low-density urban use at both the county and tract levels. At the tract level, it is very clear that highway proximity and node proximity were conducive to the rapid growth of low-density urban use, manifested by suburban housing. However, as more suburban housing has been built, its correlation with the location measures became less important, as indicated by the decreasing magnitude of correlation coefficient with time as well as the negative correlation between the proportion of 1973–1999 low-density urban use change and these location measures. It is also interesting to note that the proportion of low-density urban use correlated positively with per capita income for the years 1973 (0.62) and 1987 (0.16) only. The weakening correlation in more recent years indicates that at the beginning suburban housing was affordable only to people with higher income, but in recent years, as the general affluence of the population increased, more people have been able to afford suburban housing.

## 12.7 Future Urban Growth Scenario Simulation

The progressive urban development as projected into the period of 2000–2050 under two different scenarios can be well perceived from Fig. 12.4. It is clear that a Los Angeles-like metropolis characterized by huge urban agglomerations would emerge by around 2030 if current development conditions were still valid (Scenario One). The vegetation area and open space in the 13 metro counties (excluding the north-western mountainous area) would be very limited. In contrast, the simulated urbanization under the second scenario appears to be relatively restrictive, indicating that the effort of slowing down urbanization through model parameterization has been quite efficient.

Statistical measures reveal much more information (Table 12.5). Under the first scenario, the total urban area for 2050 would be 1286,692 hectares. The total net increment in urban area with at least 50% probability would be 793,561 hectares, or 43.6 hectares per day on the average, representing an increase of 160% between 1999 and 2050. The net urban increment as projected

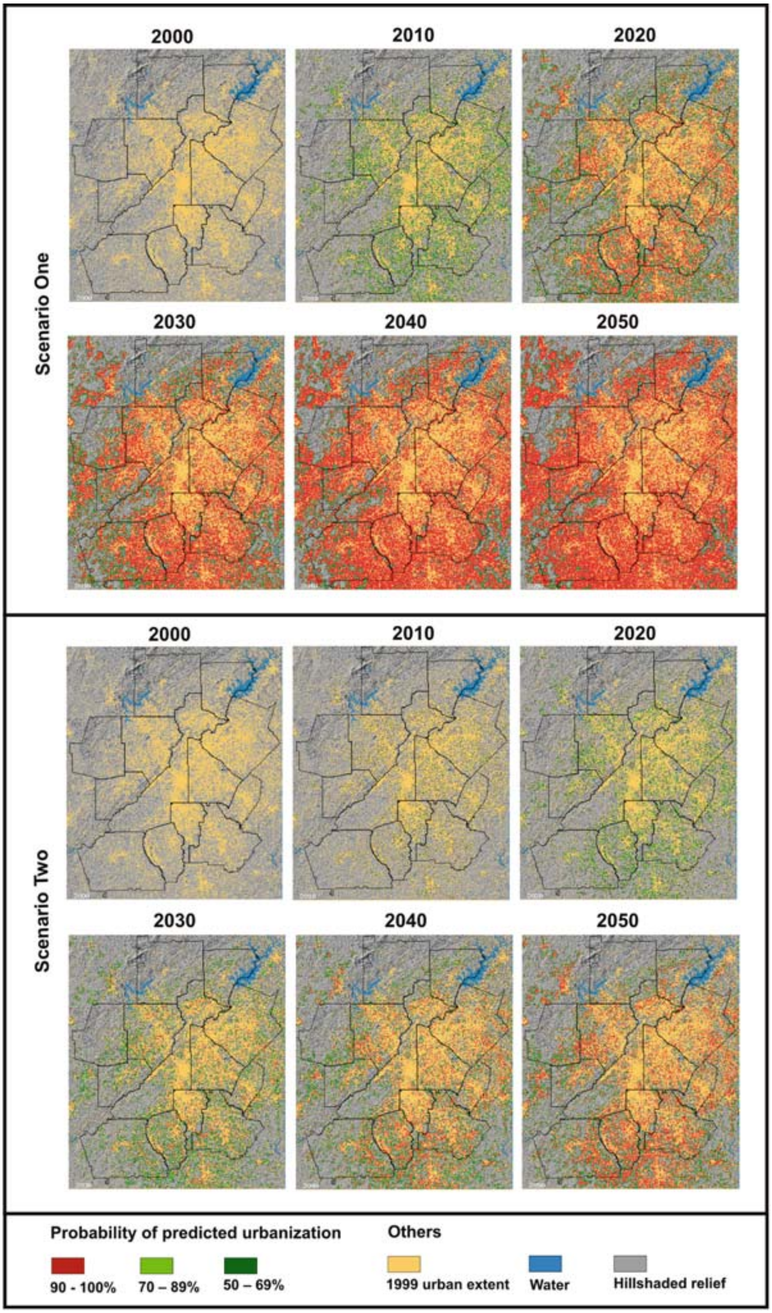



**Fig. 12.4** Simulation of future urban growth under two different scenarios. Note that the county boundary is overlaid. The dimension of each image is approximately 118 × 138 km. The original figures were in color. Readers can contact Dr. Xiaojun Yang (xjyangusa@yahoo.com) to obtain an original copy

has been overwhelmingly concentrated within 1999–2030, further confirming the finding based on graphical outputs that a huge metropolis would take shape by 2030. As a result of such a dramatic growth, urban land would occupy approximately 78.67% of the total modeled area by 2050. The average slope of urban land would increase from 4.87% in 1999 to 8.32% in 2050, implying that many steep, woody areas would be converted into urban use. Forest distribution in Atlanta was found to be positively correlated with terrain slope (Section 12.6). Under Scenario Two, by 2050 the total urban area would be 906134 hectares, or approximately 55.40% of the entire area. The total net urban increment would be 413,003 hectares, or 22.2 hectares per day, indicating an increase of 84% between 1999 and 2050. Apparently, the magnitude of urban growth as projected under this scenario has been substantially suppressed. The mean slope steepness of urban land would decrease from 4.87% in 1999 to 4.46% in 2050, implying that many low-lying, flat areas would be converted into urban uses.

The spatial distribution of simulated urbanization under the two scenarios can be discerned from Fig. 12.4. For Scenario One, the projected urban additions for 1999–2010 are largely adjacent to the 1999 urban pixels, which can be viewed as a ‘continuation’ of urbanization. This corresponds to the fact that the overwhelming share of net urban additions under the first scenario was accounted for by organic growth. This type of urban growth actually represents the expansion of existing urban pixels into their surroundings. The projected urban additions during 2010–2030 are largely distributed over places far away from the 1999 urban pixels. Many projected additions are also found in western, north-western, and south-eastern parts. Some large urban clusters can be clearly recognized. The projected urban additions after 2030 are predominately scattered over the western and south-eastern parts. Under Scenario Two, the projected urbanization for 1999–2010 has been very limited. Most of the new additions are for 2010–2030, represented by blue and green pixels in Fig. 12.4. Numerous large urban clusters can be clearly recognized, particularly in southern and western parts.

Based on the above comparisons, it is found that the results of these two scenarios are quite different. Scenario One illustrates that the unchecked urban growth would consume almost all of the entire vegetation and open space in the metropolitan area, with an exception in the north-western mountainous area. The dramatic urban growth, as projected under this scenario, would change the city’s spatial form substantially, with numerous edge cities scattered over a huge area. This would greatly deteriorate the quality of life in Atlanta. Scenario Two would cut down the rate of urban growth by approximately 50% when compared to Scenario One. It would preserve much more greenness and open space, including predefined buffer zones along large rivers, streams, and lakes. These preserved zones, although relatively small in area (about 27,358 hectares), contain the most important fresh water supply, wetlands, and floodplains for

**Table 12.5** Results of scenario simulations for future urban growth

| | | | 2010 | | 2020 | | 2030 | | 2040 | | 2050 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Area (hectare) | (%)[a] | Area (hectare) | (%) | Area (hectare) | (%) | Area (hectare) | (%) | Area (hectare) | (%) |
| Scenario One | Urbanization Probability(%) | 50–59[b] | 41985 | 2.57 | 29889 | 1.83 | 21767 | 1.33 | 15978 | 0.98 | 12390 | 0.76 |
| | | 60–69 | 51817 | 3.17 | 40602 | 2.48 | 29716 | 1.82 | 22067 | 1.35 | 17447 | 1.07 |
| | | 70–79 | 60653 | 3.71 | 58343 | 3.57 | 43419 | 2.65 | 33615 | 2.06 | 26173 | 1.60 |
| | | 80–89 | 61269 | 3.75 | 94654 | 5.79 | 76326 | 4.67 | 58291 | 3.56 | 46276 | 2.83 |
| | | 90–100 | 27740 | 1.70 | 263820 | 16.13 | 469428 | 28.70 | 603821 | 36.92 | 691281 | 42.26 |
| Total urban area[c] | | | 736595 | 45.03 | 980433 | 59.94 | 1133787 | 69.32 | 1226903 | 75.01 | 1286692 | 78.67 |
| Scenario Two | Urbanization Probability(%) | 50–59 | 29843 | 1.82 | 37123 | 2.27 | 31444 | 1.92 | 26346 | 1.61 | 22239 | 1.36 |
| | | 60–69 | 22038 | 1.35 | 44444 | 2.72 | 41507 | 2.54 | 35983 | 2.20 | 31110 | 1.90 |
| | | 70–79 | 7188 | 0.44 | 49144 | 3.00 | 54950 | 3.36 | 50970 | 3.12 | 45942 | 2.81 |
| | | 80–89 | 559 | 0.03 | 43338 | 2.65 | 75635 | 4.62 | 80185 | 4.90 | 74920 | 4.58 |
| | | 90–100 | 0 | 0.00 | 11048 | 0.68 | 75848 | 4.64 | 160865 | 9.83 | 238798 | 14.60 |
| | Total urban area | | 552758 | 33.79 | 678223 | 41.46 | 772514 | 47.23 | 847480 | 51.81 | 906134 | 55.40 |

[a] It is computed by using urban area divided by the total modeled area (1,635,656 hectares).
[b] This is the probability of predicted urbanization.
[c] It contains 1999 urban area (493,131 hectares)

the metropolis. Therefore, Scenario Two should be most desirable for future growth planning in Atlanta.

## 12.8 Conclusions

This study has demonstrated the utility of a loose-coupling approach to integrate remote sensing, GIS and dynamic modeling for urban growth and landscape change characterization. The entire research has gone through several major stages beginning with data acquisition and collection. Central to this research was a time series of satellite images acquired by three Landsat sensors, namely, MSS, TM and ETM + , which has been used to produce land use/cover maps through image classification. The combined use of unsupervised classification and GIS-based spatial reclassification procedures has been quite effective to resolve the spectral confusion among different land classes, a typical problem with broad-band sensors and within an urban image scene. The time series of land classification maps was further used to analyze urban spatial growth and the nature of change. This was built upon the combined use of post-classification comparison and GIS overlay techniques that made possible the production of single-theme change maps, which emphasize spatial dynamics. The driving force analysis was conducted through a zonal-based approach to integrate biophysical and socio-economic data. The major challenge here has been to harmonize various data that differ in formats, projections, parameter measuring, and sampling methods. The adoption of multi-level observation strategy was found to be useful for managing the MAUP problem. The last phase has been to incorporate several major driving forces into a dynamic model, which has been calibrated with historical urban extent data derived from the satellite series. The model has been used as a tool to imagine, test, and choose between two different scenarios under various environmental and development conditions. Based on this research, it is found that the integration of remote sensing, GIS, and dynamic modeling has mutually reinforced the utility of these techniques. The integration has provided insights that would otherwise not be available if spatial data were not organized in a GIS environment and GIS were not integrated with remote sensing and dynamic modeling. On the other hand, only through this integration can geospatial technologies be effective for urban growth and landscape change characterization.

This study has established a well-documented regional case study focusing on Atlanta, a typical postmodern American metropolis having undergone rapid demographic and economic growth during the past several decades. This project reveals a rampant growth in urban land during 1973–1999, which substantially outpaced the rate of population growth in Atlanta. Urban spatial form experienced a transition from linear concentration to multinucleated pattern for high-density urban use (mainly commercial, industrial and transportation) and from concentration mixed with some degree of dispersal to more dispersed pattern for low-density urban use (mainly residential) in Atlanta. Driving force analysis indicates that proximity to highways and shopping centers had promoted urban growth and increased affluence of the population had induced rapid suburbanization in Atlanta. Future urban growth simulations suggest that numerous edge cities would emerge and smaller ones would coalesce together to form larger urban clusters. If current growth patterns do not alter, the process of suburbanization would deplete vegetation and open space by around 2030. These findings should be useful to those who need to manage resources and provide services to people living in this rapidly changing environment. Given that many major metropolises across the world face the growing problems caused by rapid urban development, the technical frameworks developed in the current study focusing on Atlanta should be applicable to other metropolises. This can improve our understanding of complex urban growth dynamics, thus facilitating a sophisticated approach to sustainable urban development.

## References


Alberti, M. (1999). Modeling the urban ecosystem: a conceptual framework. *Environment and Planning B: Planning & Design*, 26(4), 605–630

Anderson, J.R., Hardy, E.E., Roach, J.T. & Witmer, R.E. (1976). *A land use and land cover classification system for use with remote sensor data*. USGS Professional Paper 964, Sioux Falls, S. D., USA

Auch, R., Taylor, J. & Acededo, W. (2004). *Urban growth in American cities: glimpses of U.S. urbanization*. (USGS Circular 1252)

Batty, M. & Longley, P. (1994). *Fractal cities: a geometry of form and function*. (London: Academic Press)

Batty, M. & Xie, Y. (1994a). Modelling inside GIS: Part 1. model structures, exploratory spatial data analysis and aggregation. *International Journal of Geographical Information Systems*, 8(3), 291–307

Batty, M. & Xie, Y. (1994b). Modelling inside GIS: Part 2. selecting and calibrating urban models using ARC-INFO. *International Journal of Geographical Information Systems*, 8 (5), 451–470

Bullard, R.D., Johnson, G.S. & Torres, A.O. (Eds.) (2000). *Sprawl city: race, politics, and planning in Atlanta*. (Washington, D.C.: Island Press)

Burgi, M., Hersperger, A.M. & Schneeberger, N. (2004). Driving forces of landscape change – current and new directions. *Landscape Ecology*, 19(8), 857–868

Cheng, J.Q. & Masser, I. (2003). Urban growth pattern modeling: a case study of Wuhan city, PR China. *Landscape and Urban Planning*, 62(4), 199–217

Clarke, K.C. & Gaydos, J. (1998). Loose-coupling a cellular automaton model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographic Information Science*, 12(7), 699–714

Clarke, K.C., Parks, B.O. & Crane, M.P. (Eds) (2002). *Geographic information systems and environmental modeling*. (Upper Saddle River, NJ: Prentice-Hall)

Congalton, R. (1991). A review of assessing the accuracy of classification of remotely sensed data. *Remote Sensing of Environment*, 37, 35–46

Dear, M. & Flusty, S. (1998). Post modern urbanism. *Annals of the Association of American Geographers*, 88(1), 50–72

Foresman, T.W., Pickett, S.T.A. & Zippere, W. C. (1997). Methods for spatial and temporal land use and land cover assessment for urban ecosystems and application in the greater Baltimore-Chesapeake region. *Urban Ecosystems*, 1(4), 201–216

Goodchild, M.F., Haining, R. & Wise, S. (1992). Integrating GIS and spatial data analysis: problems and possibilities. *International Journal of Geographic Information Systems*, 6(5), 407–423

Hall, T. (2001). *Urban geography* (2nd). (London: Routledge)

Hall, F.G., Strebel, D.E., Nickeson, J.E. & Goetz, S.J. (1991). Radiometric rectification: toward a common radiometric response among multidate, multisensor images. *Remote Sensing of Environment*, 35:11–27

Hartshorn, T. & Muller, P. (1989). Suburban downtown and the transformation of metropolitan Atlanta's business landscape. *Urban Geography*, 10(4), 375–395

Jensen, J.R. (2005). *Introductory digital image processing: a remote sensing perspective* (3rd). (Upper Saddle River, NJ: Prentice Hall)

Jensen, J.R. & Cowen, D.C. (1999). Remote sensing of urban/suburban infrastructure and socio-economic attributes. *Photogrammetric Engineering and Remote Sensing*, 65(5), 611–622

Kaplan, D.H., Wheeler, J.O., and Holloway, S.R. (2009). *Urban geography* (2nd). (New York: Wiley & Sons)

LaGro, J.A. & DeGloria, S.D. (1992). Land-use dynamics within an urbanizing nonmetropolitan county in New-York-State (USA). *Landscape Ecology*, 7(4), 275–289

Landis, J. (1995). Imaging land use futures: applying the California urban futures model. *Journal of the American Planning Association*, 61(1), 438–457

Langford, M. & Unwin, D.J. (1994). Generating and mapping population density surfaces within a geographical information system. *The Cartographic Journal*, 31, 21–26

Lo, C.P. (2004). Testing urban theories using remote sensing. *GIScience & Remote Sensing*, 41 (2), 95–115

Lo, C.P. & Quattrochi, D.A. (2003). Land-use and land-cover change, urban heat island phenomenon, and health implications: a remote sensing approach. *Photogrammetric Engineering and Remote Sensing*, 69(9), 1053–1063

Lo, C.P., Quattrochi, D.A. & Luvall, J.C. (1997). Application of high-resolution thermal infrared remote sensing and GIS to assess the urban heat island effect. *International Journal of Remote Sensing*, 18(2), 287–304

Longley, P.A. (2002). Geographical information systems: will developments in urban remote sensing and GIS lead to 'better' urban geography? *Progress in Human Geography*, 26(2), 231–239

Nedović-Budić, Z. (2002). Geographic information science implications for urban and regional planning. *URISA Journal*, 12(2), 82–93

Paulsson, B. (1992). *Urban applications of satellite sensing and GIS analysis*. (Washington DC: The World Bank)

Pijanowski, B.C., Pithadia, S., Shellito, B.A. & Alexandridis, K. (2005). Calibrating a neural network-based urban change model for two metropolitan areas of the Upper Midwest of the United States. *International Journal of Geographical Information Science*, 19(2), 197–215

SIERRA CLUB (1998). *Sprawl: the dark side of the American dream*. Retrieved 25 December 2008 from. Http://www.sierraclub.org/sprawl/report98/report.asp

Soja, E. (1989). *Postmodern geographies: the reassertion of space in critical social theory*. (London: Verso)

Weng, Q. & Quattrochi, D. (Ed.) (2007). *Urban remote sensing*. (Boca Raton: CRC Press)

Yang, X. (2002). Satellite monitoring of urban spatial growth in the Atlanta metropolitan region. *Photogrammetrical Engineering and Remote Sensing*, 68 (7), 725–734

Yang, X. (Ed.) (2003). Special issue: remote sensing and GIS for urban analysis. *Photogrammetric Engineering and Remote Sensing*, 69(9), 937–1036

Yang, X. (Ed.) (2005). Special issue: urban remote sensing. *Computers, Environment, and Urban Systems*, 29(5), 497–616

Yang, X. (2007). Integrating remote sensing, GIS and dynamic modeling for sustainable urban growth management. In V. Mesev (Ed.), *Integration of GIS and remote sensing* (pp. 173–198). (Chichester, UK: John Wiley)

Yang, X. & Lo, C.P. (2003). Modelling urban growth and landscape changes in the Atlanta metropolitan area. *International Journal of Geographical Information Science*, 17(5), 463–488

Yeh, A.G.O. & Li, X. (2003). Simulation of development alternatives using neural networks, cellular automata, and GIS for urban planning. *Photogrammetric Engineering and Remote Sensing*, 69(9), 1043–1052

# Chapter 13
# Evaluating the Ecological and Environmental Impact of Urbanization in the Greater Toronto Area through Multi-Temporal Remotely Sensed Data and Landscape Ecological Measures

**Dongmei Chen, Wenbao Liu, Jie Tian, and Peter Luciani**

**Abstract** Urbanization is a critical factor affecting the ecological and environmental balance of the Greater Toronto Area (GTA), the most populous metropolitan area in Canada, in the past three decades. The purpose of this chapter is to examine the relationship between landscape change and population increase patterns as well as to evaluate ecological impacts of urbanization in the GTA. Multi-temporal remotely sensed data have been used to derive vegetation changes from 1992 to 2003. Land use change is derived from historical land use maps. Five landscape fragmentation indices are calculated for different periods using FRAGSTATS software. Population change is compared with land use and vegetation changes. The landscape fragmentation rate is then compared with the population growth rate. Our results show that the mean normalized difference vegetation index (NDVI) is negatively correlated with the percentage of the urban settlement land and population density at the census tract (CT) level. Changes in the percentage of urban land use show relatively weak correlations with the five fragmentation indices. It was found that shape and fractal dimension indices are better at characterizing the urbanization process than the indices of diversity, contagion, and percentage of like adjacencies.

**Keywords** Urbanization · Landscape analysis · Remote sensing · Fragmentation · Environmental change

## 13.1 Introduction

Urbanization is a critical factor affecting the ecological balance of the earth system. One of the greatest challenges is to understand the impact of urban systems that evolve over time and space, as the outcome of the interaction between human behaviors and bio-physical processes. A wide variety of

D. Chen (✉)
Department of Geography, Queen's University, Kingston, ON K7L 3N6, Canada
e-mail: chendm@queensu.ca

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_13,

processes may contribute to urban development. Over the last several decades, a range of models from social-economic, transportation, land use/cover change, and other perspectives have been developed to meet land management and planning needs (Brown et al. 2005; Muller and Middleton 1994; Pijanowski et al. 2005; Veldkamp and Lambin 2001; Verburg and Veldkamp 2005). Relatively, fewer studies have focused on the relationship among environmental, ecological, and social-economical changes caused by urbanization.

Landscape disturbance is a common problem caused by changing land covers and landscape patterns of urbanization (Wang and Moskovits 2001). Such disturbance could lead to low productivity of the ecosystem due to extra travel time, wasted space along borders, inability to use certain types of facilities, etc. Landscape complexity and heterogeneity are key aspects in assessing ecological processes and biodiversity within regions (Southworth et al. 2004). Past research has demonstrated that aspects of landscape pattern or structure influence the occurrence and distribution of species (Forman 1995). To measure landscape structure, indices describing landscape patterns, such as contagion, diversity and shape complexity, have been widely calculated from discrete landscape representations using software programs like FRAGSTATS (McGarigal et al. 2002). For continuous landscape representations, spatial statistics like local spatial autocorrelation measures and surface fractal dimension have been used (Pearson 2002; Read and Lam 2002; Seixas 2000; Southworth et al. 2004).

The GTA is one of the fastest growing and most populous metropolitan areas in Canada exhibiting rapid economic development, urbanization, and population growth in the past three decades (Bell et al. 1999; Bell 2001). Consequently this urbanization has led to a series of environmental problems including reduced farmland and biological diversity, high rates of soil erosion and water quality degradation, and air pollution (Yap et al. 2005). This change will continue to impact the structure and function of the ecosystem in this region (Blais 2000). Based on a forecast by the Urban Development Services of the Toronto City Planning, the population of the Greater Toronto Area will increase to 7.45 million by 2031. The increased population will add more anthropogenic pressure upon local lands, which, if not managed effectively, can eventually lead to further environmental degradation. In turn, such degradation can affect and potentially slow down if not threaten the pace of economic development in this region. Despite the importance of understanding the impact of past urbanization on sustainable development in this region, there have been few comprehensive analyses of land-use/cover, vegetation/other natural habitats, and socio-economic change across time and space in this heartland region of Canada.

Remotely sensed data is very useful for mapping vegetation and land use/cover changes in urban areas by overcoming many limitations of traditional surveying techniques to obtain a continuous and extensive inventory of ecosystems (Ridd and Liu 1998; Ward et al. 2003; Rogan and Chen 2004; Gillanders et al. 2008; Torres-Vera et al. 2009). With the use of remote sensing, it is possible to map and monitor the spatial extent of various factors influencing and contributing to environmental degradation, such as changes in vegetation cover, impervious surface, land use type, and human activities. Remotely sensed data can also provide the necessary area-based land use/cover parameters to run different urban change models as well as plausible scenarios used to simulate future response to different planning/management scenarios (Clarke and Gaydos 1998; Veldkamp and Lambin 2001; Lo and Yang 2002; Sohl et al. 2007; Tang et al. 2007; Yuan 2009). However, prior research has predominantly focused upon land use/cover patterns caused by urban growth. There is a paucity of research analyzing the impact of land use/cover, population, and ecological change jointly as key urbanization factors.

This paper aims to evaluate the relationship between landscape change and population increase patterns as well as ecological impacts of urbanization in the GTA. Multi-temporal remotely sensed data are employed to derive vegetation changes from 1992 to 2003. Land use change is derived from historical land use maps. Population change at the census level is compared with land use and vegetation changes from 1992 to 2003. The landscape ecological indices are calculated from reclassified land use maps of the early 1990s and 2000s and compared with percentages of urban land use.

## 13.2 Method

### *13.2.1 Study Area*

The GTA is located in the southern portion of Ontario, Canada covering 7,125 km$^2$ (Fig. 13.1), including regional municipalities of Halton, Toronto,


**Fig. 13.1** Study area in Southern Ontario (the thick *dark line* is the boundary of the GTA)

York, Peel and Durham regions. The total population of GTA was 3.89 million in 1991 and increased to 4.68 million in 2001. A large part of the GTA are covered by farmlands and forests protected by the Greenbelt. The land use in other GTA area is predominantly urban including major transportation routes and pockets of expanding urban development. Between 1990s and early this century, urban areas and associated land uses/covers have significantly expanded. Areas that are not urbanized comprise various agriculture lands, forests, wetlands and water bodies.

### *13.2.2 Data Sets*

In order to analyze the relationship between population increase, land use, and vegetation changes from the early 1990s to the early 2000s, census data in 1991 and 2001 census years have been collected from Statistics Canada for each census tract (CT) in the GTA. This data is then linked to the census tract boundary file. Our analysis focused on the CT level of seven digits to ensure the CTs are geographically comparable between the two time instances. Population change from 1991 to 2001 is then calculated for each CT.

Multi-temporal Landsat TM images are used for vegetation change detection across different years. The ideal situation for change detection is that images are acquired at the same or very similar dates and times of different years. Two Landsat TM images acquired on June 3, 1992 and May 22, 2003, respectively, were ordered from the Canada Center for Remote Sensing (CCRS) containing no or minimal cloud cover. The TM images are compared for locational association and were found to need only a minor systematic translation to align to almost exact registration. All images, once aligned, were projected in the NAD 1983 UTM Zone 17 North projected coordinate system. Figure 13.2 shows the two TM images acquired in 1992 and 2003 in grey scale.

In order to measure the vegetation changes in GTA, a simple and well-known vegetation index, the normalized difference vegetation index (NDVI), is generated for each image to represent the continuum of landscape changes. NDVI is calculated from the reflectance measured in the red visible and near infrared bands by using the following equation (Jensen 2005):

$$\mathrm{NDVI} = (V_{NIR} - V_{RED})/(V_{NIR} + V_{RED}) \tag{13.1}$$

where $V_{NIR}$ and $V_{RED}$ represent the spectral values at the near infrared band (Landsat Band 4) and red visible band (Landsat band 3), respectively.

NDVI is often correlated to a variety of vegetation characteristics that include quantity, productivity, biomass, etc. By design, NDVI varies between –1 and 1. An area containing a dense vegetation canopy usually exhibits high positive values while free standing water and soils generally exhibit very low positive or even slightly negative NDVI values (Myneni et al. 1995). An NDVI

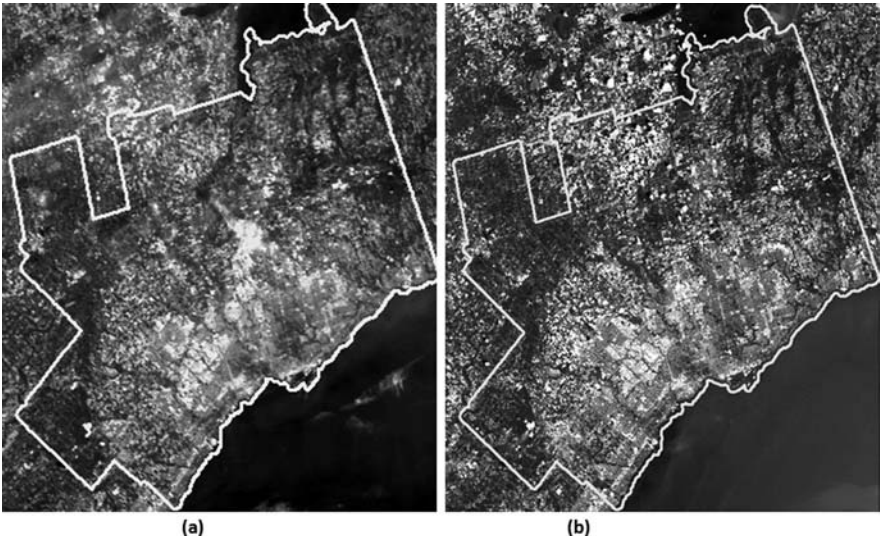


Fig. 13.2 Landsat TM images acquired at (a) June 3, 1992 and (b) May 22, 2003

change map has been generated by subtracting the 1992 NDVI map from the 2003 NDVI map. The summary statistics of NDVI changes for each census tract have been generated by zonal analysis in ArcGIS.

Initially land use maps were to have been generated from the aforementioned TM images through image classification. However, after several tests, the classification results were found to be highly subjective and not conductive to this analysis given the producer's/ analyst's choice and interpretation of classes, training sites, and accuracy measures. As such, two historical land use maps generated by the Ontario Ministry of Natural Resources (OMNR) using the Ontario Land Cover Data Base Classification Scheme have been collected. These maps depict the land use/covers within two time periods (early 1990s and early 2000s) approximately ten years apart. The former map is created solely based upon satellite imagery, while the latter is a compilation of data from various sources including topographic maps, aerial photographs, satellite imagery, and ground surveying. Based upon a review of the metadata for the maps, both are considered accurate. The later map (2000s) was provided in vector, it has been converted to raster from vector format to facilitate the subsequent analysis. As the classification schemes for the two maps differed, the original land use/cover classes were aggregated into five broad classes, which are: water, naturally vegetated land, agricultural land, settlement/developed land, and open land. The consistent classification schemes are necessary for temporal comparison. These classes are selected as being significant to this study and as per convention within studies of ecological and environmental impacts of urbanization.

### 13.2.3 Landscape Ecological Measures

Landscape ecological measures capturing fundamental aspects of landscape pattern that influence ecological processes are calculated in FRAGSTATS for the two reclassified land use maps. Spatial heterogeneity or complexity in the landscape can be viewed as a continuum of variability and complexity ranging from low to high heterogeneity. This continuum can be measured either directly by measuring complexity and variability, or indirectly by measuring the departure from homogeneity (Li and Reynolds 1995). To date many landscape indices have been developed in the past (McGarigal et al. 2002). This study employs shape, diversity, and aggregation indices.

Patch shape complexity indices relate to the geometry of patches. Shape indices capture overall shape complexity rather than a value for each unique shape (Gustafson 1998; McGarigal et al. 2002). Shape index can be calculated with several methods, including perimeter-area ratios, complexity measurements of patch shape as compared to a standard shape (square), and fractal dimension (Milne 1988; Forman 1995). Most commonly used shape metrics are based on perimeter-area relationships. The area-weighted mean patch fractal dimension is a spatial configuration index that measures the degree of complexity, based on fractal analysis, whereas the area-weighted mean shape index compares the shape complexity to a standard shape.

The diversity indices include metrics such as richness, evenness, dominance and diversity, referring to the number of patch types, the relative abundance of different patch types, and a composite of both (McGarigal et al. 2002). As these composition metrics require integration over all patch types, the diversity metrics are only applicable at the landscape level (McGarigal et al. 2002). Multiple indices such as Shannon's, Simpson's and Modified Simpson's evenness and diversity indices can be calculated to measure both landscape evenness and diversity in FRAGSTATS. All these diversity indices are composition metrics reflecting a composite measure of richness and evenness at the landscape level.

Contagion and interspersion indices assess the spatial aggregation in the landscape. Contagion can be defined as the tendency of cells of similar patch types to occur in large aggregated or contagious distributions. The contagion index (CONTAG), a spatial configuration metric, measures dispersion at the landscape level (McGarigal et al. 2002). The metric units are given as a percentage and a value approaching zero indicates that the patch types are maximally disaggregated and interspersed.

The study area has been divided into 112 grids (cells) each with a size of $6\times6$ km$^2$, to calculate landscape metrics for both reclassified landuse maps. Five landscape level metrics, including the shape dimension (SHAPE AM), the fractal dimension (FRAG AM), Shannon's diversity index (SHDI), the contagion index (CONTAG), and the percentage of like adjacencies (PLADJ), were calculated within each grid.

## 13.3 Results and Analysis

Figure 13.3 maps the total population and population density changes at the CT level from 1991 to 2001. The total population increased from 3.89 million to 4.68 million from 1991 to 2001. It is evident in Fig. 13.3 that the population increase occurred mainly at the urban-rural fringe of the City of Toronto and other urban centers within the GTA.

From the early 1990s to early 2000s, the total area of settlement and developed land in the GTA has increased by 513 km$^2$. Meanwhile, the area of agricultural land and naturally vegetated land has decreased by 114 and 423 km$^2$, respectively. As evident in Fig. 13.4, most of the land use/cover changes, similar to population change, have occurred at the urban-rural fringe within the northern portion of the GTA. In this area agricultural lands and naturally vegetated lands had been converted to new settlement or development areas.

The NDVI change map between 1992 and 2003 is shown in Fig. 13.5. The NDVI change values range from –0.560 to 0.512. The mean NDVI for the entire GTA is 0.246 in 1992 and had a slight increase to 0.263 in 2003. This increase may be the result of vegetative growth from 1992 to 2003. As expected, the NDVI change occurs most in changed urban, suburban, and agriculture areas, while little change appears in the developed urban areas.

The Pearson's product moment correlation coefficients between population density, percentage of settlement, and the mean NDVI changes at the CT level are listed in Table 13.1. As expected, mean NDVI is negatively correlated with the percentage of urban settlement land and population density. Urbanized areas tend to contain more impervious surfaces resulting in relatively low NDVI

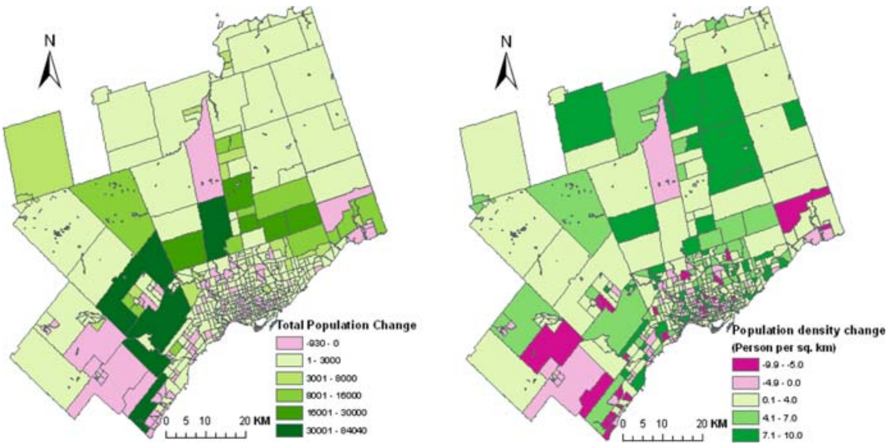



**Fig. 13.3** Total population and population density changes from 1991 to 2001 at the census tract level in GTA

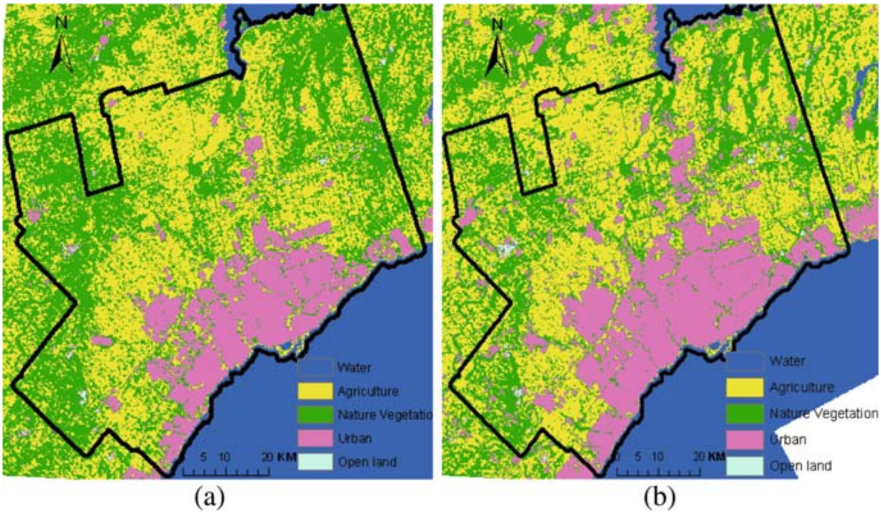



**Fig. 13.4** Land use maps of GTA for early 1990s (**a**) and early 2000s (**b**)

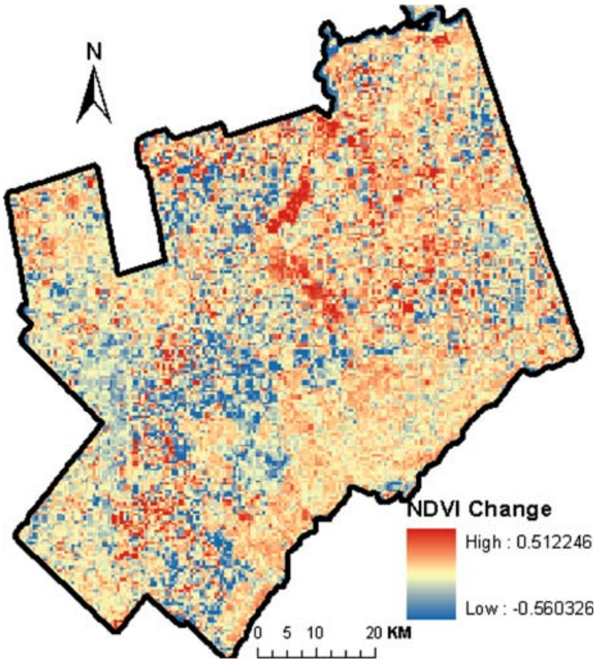



**Fig. 13.5** The NDVI change from 1992 and 2003 TM images

**Table 13.1** Correlation between NDVI, percentage of settlement and developed land, and population at the CT level

| | Early 1990s | | | Early 2000s | | | Change | | |
|---|---|---|---|---|---|---|---|---|---|
| | r | p | n | r | p | n | r | p | n |
| NDVI vs. PSD | −0.58 | <0.001 | 487 | −0.47 | <0.001 | 487 | −0.15 | <0.001 | 487 |
| NDVI vs. POP | −0.42 | <0.001 | 487 | −0.36 | <0.001 | 487 | −0.07 | <0.141 | 487 |
| PSD vs. POP | 0.44 | <0.001 | 487 | 0.39 | 0.02 | 487 | 0.03 | 0.504 | 487 |

PSD represents percentage of settlement and developed land; POP represents population density

values. In comparison, vegetated areas are normally associated with higher NDVI values. However, NDVI change is not significantly correlated to population density change. This observation likely indicates that population density change is not a major factor towards vegetation change at the CT level. Other factors such as land use planning strategies may be more significant factors.

At the grid level, the negative correlation between mean NDVI and the percentage of settlement and developed land and the population density became stronger than those at the CT level (Table 13.2). This increase is reasonable considering the impact of analysis scale and is consistent with the results from other research involving the Modifiable Areal Unit Problem (MAUP) studies (Cressie 1996). More particularly, population density has a very strong correlation (0.91 for 1990s and 0.87 for 2000s) with the percentage of settlement and developed land. It is reasonable that the percentage of settlement and developed land should be negatively correlated with the average NDVI, as urban development is often at a cost of vegetation reduction. The correlation strength (0.66 for 1990s and 0.62 for 2000s) between NDVI and population density is relatively weaker as shown in Table 13.2. The relationship between these two variables is more indirect with an NDVI value often reflecting the level of urbanization.

The five landscape fragmentation indices have been calculated at the grid level for each of two land use maps. These indices have been correlated with the percentage of settlement and developed land. The urban land use percentage is positively correlated with PLADJ and CONTAG, but negatively correlated with SHAPE, FRAG, and SHDI (Table 13.3). In particular, the correlations with FRAG, SHAPE, and PLADJ are stronger than those with SHDI and

**Table 13.2** Correlation between NDVI, percentage of settlement and developed land, and population at the grid level

| | Early 1990s | | | Early 2000s | | | Change | | |
|---|---|---|---|---|---|---|---|---|---|
| | r | p | n | r | p | n | r | p | n |
| NDVI vs. PSD | −0.81 | <0.001 | 112 | −0.8 | <0.001 | 112 | −0.28 | 0.002 | 112 |
| NDVI vs. POP | −0.66 | <0.001 | 112 | −0.62 | <0.001 | 112 | −0.09 | 0.351 | 112 |
| PSD vs. POP | 0.91 | <0.001 | 112 | 0.87 | <0.001 | 112 | 0.58 | <0.001 | 112 |

PSD – percentage of settlement and developed land; POP – population density

**Table 13.3** Correlation between percentage of settlement and developed land, and the fragmentation metrics at the grid level

| | Early 1990s | | | Early 2000s | | | Change | | |
|---|---|---|---|---|---|---|---|---|---|
| | r | p | n | r | p | n | r | p | n |
| SHAPE vs. PSD | −0.62 | <0.001 | 112 | −0.54 | <0.001 | 112 | −0.08 | 0.381 | 112 |
| FRAG vs. PSD | −0.65 | <0.001 | 112 | −0.61 | <0.001 | 112 | −0.07 | 0.445 | 112 |
| SHDI vs. PSD | −0.18 | 0.053 | 112 | −0.41 | <0.001 | 112 | 0.01 | 0.912 | 112 |
| CONTAG vs. PSD | 0.21 | 0.027 | 112 | 0.46 | <0.001 | 112 | 0.03 | <0.724 | 112 |
| PLADJ vs. PSD | 0.58 | <0.001 | 112 | 0.56 | <0.001 | 112 | −0.03 | <0.759 | 112 |

PSD – percentage of settlement and developed land; FRAG – the fractal dimension; Shape – the shape dimension; SHDI – Shannon's diversity index; CONTAG – the contagion index; PLADJ – the percentage of like adjacencies.

CONTAG. However, the changes in percentage of urban land use show relatively weak correlations with the five fragmentation indices.

Each fragmentation index is compared spatially across grids within the individual image to examine the spatial patterns of landscape fragmentation. To foster fragmentation analysis at the grid level, the values of each index have been classified into three classes: low, middle, and the high-value classes. Figure 13.6 presents the spatial distributions of SHAPE index as an example for the five indices. The SHAPE index has shown a clear spatial gradient across the study site. This gradient generally corresponds to the three dominant land use areas: the urban, the urban-rural fringe, and the rural areas.

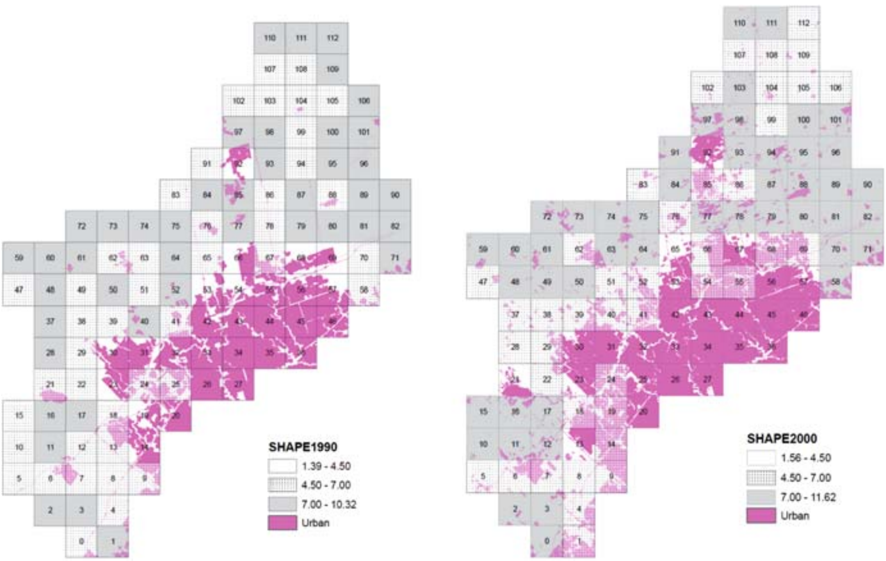



**Fig. 13.6** Spatial patterns of landscape fragmentation measured by SHAPE metrics in the early 1990s and 2000s

Similar maps with Fig. 13.6 have been generated and analyzed for the other four indices. These maps also indicate that the fragmentation indices reasonably characterize the urbanization processes. These five indices, however, characterize landscape fragmentation zones differently. From low to high values, SHAPE and FRAC metrics characterize the degree of urbanization from high to low. In other words, the urban area is the less fragmented zone, the urban-rural fringe is the more fragmented zone and the rural area is the most fragmented zone. In contrast, from low to high values, CONTAG and PLDJ indices characterize the degree of urbanization from low to high. Nevertheless, the spatial pattern of the SHDI index differs from those of other metrics in that the low value class characterizes the urban-rural fringe area. This finding implies that fragmentation indices do reveal the transition process of land use changes from the dominant land use of vegetation to agriculture and eventually to urban developed land.

Fragmentation indices have also been compared temporally for the two images on a grid basis to examine the impact of urbanization on landscape fragmentation during the decadal period. Figure 13.7 presents the spatial patterns of changes in SHAPE and FRAC indices during the decade. The large changes are generally located in the grid areas with significant land use changes in and outside of the urban region between 1990 and 2000. Similar maps for the other three indices are produced for visual interpretation and analysis. The maps of the spatial changes of these indices do not exhibit clear

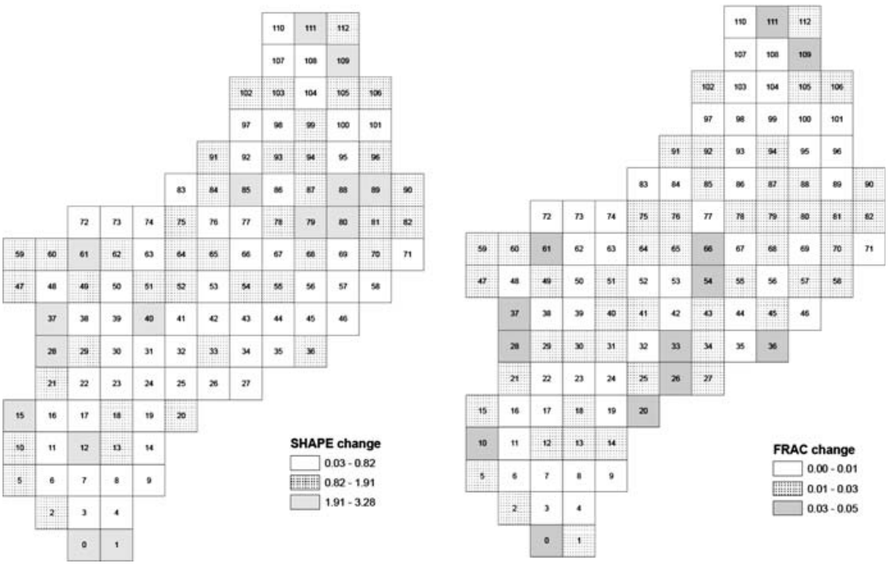



**Fig. 13.7** Spatial patterns of absolute change in SHAPE and FRAC indices between 1990 and 2000

patterns that correspond to the land use changes from 1990s to 2000s. This finding suggests that the SHAPE and FRAC indices are better to characterize the urbanization process compared to the other three indices.

## 13.4 Conclusions

This paper examines the relationship between land use, vegetation change, and population change patterns, and evaluates ecological impacts of urbanization in the GTA. Multi-temporal remotely sensed data have been employed to derive NDVI changes from 1992 to 2003. Land use change has been derived from historical land use maps. Population change is compared with land use and vegetation changes from 1992 to 2003 at both the census track (CT) level and across a grid of 6 km x 6 km cells. The five landscape fragmentation indices including the fractal dimension (FRAG), the shape dimension (SHAPE), Shannon's diversity index (SHDI), the contagion index (CONTAG), and the percentage of like adjacencies (PLADJ) have been calculated at the grid level for the two land use maps. These indices are compared both spatially across grids and temporally over the decade to examine the spatial patterns of landscape fragmentation and the impact of urbanization on landscape fragmentation over the decadal period.

Results indicate that mean NDVI is negatively correlated with the percentage of urban settlement land and population density. This correlation is stronger at the grid level than at the CT level. However, NDVI change is not correlated with population density change. The changes in the percentage of urban land use show relatively weak correlations with the five fragmentation indices. It is found that fragmentation indices do capture the land use transition process from a dominant land use of vegetation to agriculture and eventually to urban developed land. The urban land use percentage is positively correlated with PLADJ and CONTAG, but negatively correlated with SHAPE, FRAG, and SHDI. However, spatial changes of PLADJ, CONTAG, and SHDI do not capture as clear patterns of land use change as SHAPE and FRAC do.

As many studies have demonstrated, landscape ecological measures and correlation analysis are strongly dependent on the analysis scale (Turner 1989; Cressie 1996; Lausch and Herzog 2002; Li and Reynolds 1995; Gustafson 1998). The impact of analysis scale is evident from the correlation analysis between NDVI, population density, and land use at the census tract and grid levels. Our landscape fragmentation analysis is based on the aggregated land use data at a regional scale. It should be noted that the correlation between fragmentation measures and land use change may change when the analysis is based on finer land use classification scheme and analysis scale. This dependence represents an opportunity for future study.

## References

Bell, D.V.J. (2001). Sustainable urban communities in Canada: From Rio to Johannesburg. York Centre for Applied Sustainability York University. Prepared under contract for Stratos as background documentation for the report of the Government of Canada to the World Summit

Bell, D.V.J., Tilman, A. & Cheung, A. (1999). Advancing sustainability in the great lakes basin., York Centre for Applied Sustainability York University. Prepared for Environment Canada

Blais, P.M. (2000). Inching toward sustainability: The evolving urban structure of the GTA., Reports of Neptis Foundation

Brown, D.G., Page, S., Riolo, R., Zellner, M. & Rand, W. (2005). Path dependence and the validation of agent-based spatial models of land use. *International Journal of Geographic Information Science*, 19(2), 153–174

Clarke, K. & Gaydos, J. (1998). Loose-coupling a cellular automation model and GIS: long-term urban growth prediction for San Francisco and Washington/Baltimore. *International Journal of Geographic Information Science*, 12(7), 699–714

Cressie, N. (1996). Change of support and the modifiable areal unit problem. *Geographical Systems*, 3, 159–180

Forman, R.T.T. 1995. *Land mosaics: the ecology of landscapes and regions*. (Cambridge: Cambridge University Press)

Gillanders, S.N., Coops, N. C., Wulder, M.A., Gergel, S.E. & Nelson, T. (2008). Multi-temporal remote sensing of landscape dynamics and pattern change: describing natural and anthropogenic trends. *Progress in Physical Geography*, 32 (5), 503–528

Gustafson, E.J. (1998). Quantifying landscape spatial pattern: what is the state of the art? *Ecosystems*, 1, 143–156

Jensen, J.R. 2005. *Introductory digital image processing: a remote sensing perspective*. (Upper Saddle River, NJ: PEARSON Prentice Hall)

Lausch, A. & Herzog, F. (2002). Applicability of landscape metrics for the monitoring of landscape change: issues of scale, resolution and interpretability. *Ecological Indicators*, 2, 3–15

Li, H. & Reynolds, J.F. (1995). On definition and quantification of heterogeneity. *Oikos*, 73: 280–284

Lo, C.P. & Yang, X. 2002. Drivers of land-use/land-cover changes and dynamic modeling for the Atlanta, Georgia Metropolitan area. *Photogrammetric Engineering & Remote Sensing*, 68(10), 1073–1082

McGarigal, K., Cushman, S.A., Neel, M.C. & Ene, E. (2002). *Spatial Pattern Analysis Program for Categorical Maps*. Computer software program produced by the authors at the University of Massachusetts, Amherst

Milne, B.T. (1988). Measuring the fractal geometry of landscapes. *Applied Mathematics and Computation*, 27, 67–79

Muller, M.R. & Middleton, J. 1994. A Markov model of land-use change dynamics in the Niagara Region, Ontario, Canada. *Landscape Ecology*, 9(2), 151–157

Myneni, R.B., Hall, F.G., Sellers, P.J. & Marshak, A.L. (1995). The interpretation of spectral vegetation indexes. *IEEE Transactions on Geoscience and Remote Sensing*, 33, 481–486

Pearson, D.M. (2002). The application of local measures of spatial autocorrelation for describing patterns in north Australian landscapes. *Journal of Environment Management*, 64, 85–95

Pijanowski, B.C., Pithadia, S., Shellito, B.A. & Alexandridis, K. (2005). Calibrating a neural network-based urban change model for two metropolitan areas of the Upper Midwest of the United States. *International Journal of Geographic Information Science*, 19(2), 197–215

Read, J.M. & Lam, N.S.-N. (2002). Spatial methods for characterizing land cover and detecting land cover changes for the tropics. *International Journal of Remote Sensing*, 23(12), 2457–2474

Ridd, M. K. & Liu, J. (1998). A comparison of four algorithms for change detection in an urban environment. *Remote Sensing of Environment*, 63(2), 95–100

Rogan, J. & Chen, D. (2004). Remote sensing technology for land cover and land use mapping and monitoring. *Progress in Planning*, 61(4), 301–325

Seixas, J. (2000). Heterogeneity from remote sensing images: the case of desertification in southern Portugal. *International Journal of Remote Sensing*, 21(13–14), 2645–2663

Sohl, T.L., Sayler, K.L., Drummond, M.A. & Loveland, T.R. (2007). The FORE-SCE model: a practical approach for projecting land cover change using scenario-based modeling. *Journal of Land Use Science*, 2(2), 103–126

Southworth, J., Munroe, D. & Nagendra, H. 2004. Land cover change and landscape fragmentation – comparing the utility of continuous and discrete analysis for a western Honduras region. *Agriculture, Ecosystems and Environment*, 101, 185–205

Tang, J., Wang, L. & Yao, Z. (2007). Spatial-temporal urban landscape change analysis using the Markov chain model and a modified genetic algorithm. *International Journal of Remote Sensing*, 28(15), 3255–3271

Torres-Vera, M.A., Prol-Ledesma, R.M., & Garcia-Lopez, D. (2009). Three decades of land use variations in Mexico City. *International Journal of Remote Sensing*, 30(1), 117–138

Turner, M.G. 1989. Landscape ecology: the effect of pattern on process. *Annual Review of Ecology and Systematics*, 20, 171–197

Veldkamp, A. & Lambin, E.F. (2001). Predicting land-use change. *Agriculture, Ecosystems and Environment*, 85, 1–6

Verburg, P.H. & Veldkamp, A. (2005). Introduction to the special issue on spatial modeling to explore land use dynamics. *International Journal of Geographic Information Science*, 19(2), 99–102

Wang, Y. & Moskovits, D.K. (2001). Tracking fragmentation of natural communities and changes in land cover: applications of landsat data for conservation in an urban landscape (Chicago Wilderness). *Conservation Biology*, 15(4), 835–843

Ward, D., Phinn, S. & Murray, A. (2003). Monitoring growth in rapidly urbanizing areas using remotely sensed data. *The Professional Geographer*, 52(3), 371–386

Yap, D., Reid, N., De Brou, G. & Bloxam, R. (2005). Transboundary air pollution in Ontario, Ontario Ministry of the Environment, Toronto

Yuan, F. (2009). Land-cover change and environmental impact analysis in the Greater Mankato area of Minnesota using remote sensing and GIS modeling. *International Journal of Remote Sensing*, 29(4), 1169–1184

# Chapter 14
# Modeling Urban Effects on the Precipitation Component of the Water Cycle

**Marshall Shepherd, Willis Shem, Lauren Hand, Michael Manyin, and Dmitry Messen**

**Abstract** Precipitation is an important component of the global water cycle and a proxy for changing climate. Proper understanding and quantification of spatio-temporal precipitation variability is critical for a range of meteorological, hydrological, and climate processes. Past and current literature has presented theories and observational studies on how urbanization affects precipitation. Assessment of the urban environment's (land use, aerosols, thermal properties) impact on precipitation will be increasingly important in ongoing climate diagnostics and prediction, global water and energy cycle (GWEC) analysis and modeling, weather forecasting, freshwater resource management, urban planning-design, and land-atmosphere-ocean interface processes. This chapter presents a review of findings and methods related to "urban rainfall effect" studies with an emphasis on numerical modeling strategies. Numerical modeling of atmosphere-land interactions enables controlled experimentation to address fundamental research questions.

**Keywords** Urbanization · Precipitation · Water cycle · Modeling · Climate change

## 14.1 Introduction and Motivation

Both natural and human-induced climate variations appear as changes in the global water cycle (Chahine 1992). The Intergovernmental Panel on Climate Change (Trenberth et al. 2007) noted that Earth's mean temperature has increased in recent decades in response to primarily anthropogenic activities. In this context, higher evaporation and precipitation rates might occur, which could lead to an overall acceleration of the global water cycle (Webster et al. 2005; Trenberth 2005).

M. Shepherd (✉)
University of Georgia, Athens, GA, USA
e-mail: marshgeo@uga.edu

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_14,

Precipitation is a key link in the global water cycle and a proxy for changing climate; therefore proper assessment of the urban environment's (land use, aerosols, thermal properties) impact on precipitation will be increasingly important in ongoing climate diagnostics and prediction, global water and energy cycle (GWEC) analysis and modeling, weather forecasting, freshwater resource management, urban planning-design, and land-atmosphere-ocean interface processes. These facts are particularly critical if current projections for global urban growth are accurate. Hand and Shepherd (2009) noted that in 2008, more than half the world's population lived in urban areas. The United Nations projects that this could reach 81% by 2030 (UNFPA 2007). Human activity in urban environments has impacts at local to global scales by changing atmospheric composition; impacting components of the water cycle; and modifying the carbon cycle and ecosystems.

## 14.2 Historical and Current Perspectives on the Urban Rainfall Effects

While the climate change discussion has focused on greenhouse gas forcing and its effects, the IPCC (Trenberth et al. 2007) noted a growing interest in understanding what role urban land cover/land use (LCLU) and pollution has on climate change. The urban heat island (UHI), for example, is well studied.

The intensity of a UHI is related to the size of an urban region; it is defined as the difference between rural and urban temperatures (Oke 1981). The magnitude of the UHI attains a maximum value sometime after sunset (Oke 1987) although UHI circulations are also observed during the day due to urban–rural pressure gradients and strong vertical mixing (Shreffler 1978; Fujibe and Asai 1980).

UHI circulations are caused by differential heat capacity and thermal inertia between rural and urban regions (Fig. 14.1). UHI-related temperature gradients depend strongly on both land use and urban parameters (e.g. built-up ratio, green surface ratio, sky view factor, etc.) (Oke 1987). A surplus of surface energy over urban regions can be traced to enhanced surface sensible heat flux, ground heat storage, and anthropogenic heating as well as reduced evapotranspirational cooling. Urban regions typically have lower albedo values than rural areas and therefore, more absorption of shortwave radiation energy at the surface. A reduction in sky view factor (SVF) below roof level reduces radiative loss and turbulent heat transfer and contributes positively to the UHI anomaly (Unger 2004). Gradients in the surface energy budget components between urban and rural regions ultimately cause thermally direct mesoscale circulations similar to sea breeze or mountain-valley circulations.

UHI-related dynamics and urban surface characteristics have been shown to affect the spatio-temporal variability of rainfall (Shepherd 2005). Urban effects on rainfall can be traced back to the early 1900s (Horton 1921) but it was

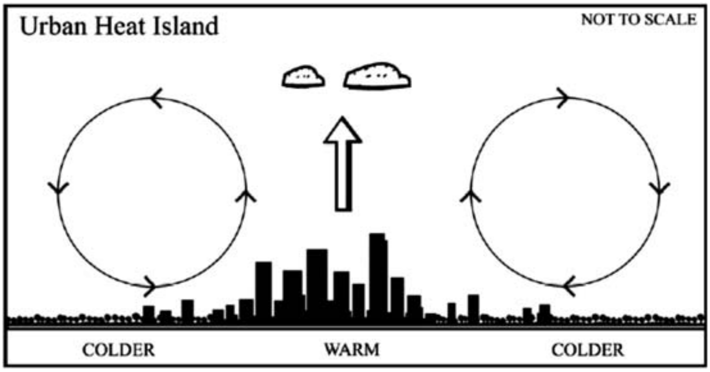



**Fig. 14.1** Urban heat island circulation during the night with light winds caused by the surface gradient between urban and rural regions (from Simpson 2006)

Helmut Landsberg's classic text "The Climate of Towns" (1956) that popularized the notion that urban areas could affect rainfall patterns. A series of studies emerged in the 1960s and 1970s that suggested increased rainfall downwind of urban areas (Landsberg 1970; Huff and Changnon 1972a, b). Enhancement of convective rainfall by large urban areas in the United States was studied during the Metropolitan Meteorological Experiment (METROMEX) (Changnon et al. 1977; Huff 1986). Observations from METROMEX showed enhanced precipitation by urban effects typically 25–75 km downwind of a city during summer months (Huff and Vogel 1978, Changnon 1979, Braham 1981). Precipitation amounts were enhanced by 5–25% over background values (Changnon et al. 1981). The size of an urban area was shown to influence the horizontal extent and magnitude of urban enhanced precipitation (Changnon et al. 1991). Increased convection downwind of large urban areas such as Tokyo was related to enhanced surface convergence over the urban region (Fujibe 2003). Maximum radar reflectivities were observed downwind of New York City due to enhanced convective development caused by the urban environment (Bornstein and LeRoy 1990). The diagram in Fig. 14.2 is a conceptualization of the "urban rainfall effect" (URE) based on Shepherd et al. (2002).

There is renewed debate on how the urban environment affects precipitation variability. Possible mechanisms for urbanization to enhance precipitation or convection include one or a combination of the following: (1) enhanced convergence due to increased surface roughness in the urban environment (e.g., Changnon et al. 1981, Bornstein and Lin 2000); (2) enhanced sensible heat fluxes (e.g. Huff and Vogel 1978, Thielen et al. 2000); (3) destabilization due to urban heat island (UHI)-thermal perturbation of the boundary layer and resulting downstream translation of the UHI circulation or UHI-generated convective clouds (e.g., Shepherd et al. 2002; Shepherd and Burian 2003; Baik et al. 2007; Mote et al. 2007); (4) enhanced aerosols in the urban environment

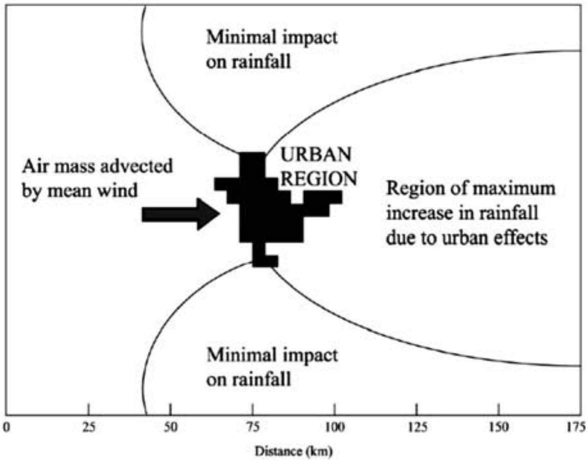



**Fig. 14.2** Conceptualization of spatial extent of urban rainfall effect (an adaptation from Shepherd et al. 2002 in Simpson 2006)

for cloud condensation nuclei sources (e.g., Molders and Olson 2004; Jin et al. 2005; Rosenfeld et al. 2008); or (5) bifurcation or diversion of precipitating systems by the urban canopy or related processes (e.g., Bornstein and Lin 2000; Loose and Bornstein 1977). Because there is no conclusive theory on which of these mechanisms, independently or synergistically, controls urban rainfall processes, we will hereafter refer to the "urban rainfall effect" or URE as some combination of the relevant processes. Shepherd (2005) and Lowry (1998) reviewed the literature documenting historical findings related to urban effects on precipitation and possible mechanisms.

Studies beyond 2005 continued to describe the influence of urbanization on rainfall throughout the world. Burian and Shepherd (2005) found that the downtown metropolitan and downwind regions of Houston, Texas had 59 and 30% respectively greater rainfall amounts between noon to midnight in the warm season compared to the upwind region. Shepherd (2006) found that, during the monsoon season, northeastern suburbs and exurbs of the Phoenix metropolitan area, an arid region, experienced statistically significant anomalies in mean precipitation of 12–14% from a pre-urban (1895–1949) to post-urban (1950–2003) period. He hypothesized that topography-urban interactions were the cause of the anomaly. Chen et al. (2007) suggested a similar urban-topographical interaction as the cause of a Taipei (Taiwan) nocturnal rainfall anomaly. Mote et al. (2007), using Doppler radar analysis, showed (Fig. 14.3a) that the eastern (e.g. downwind) suburbs of Atlanta, Georgia receive up to 30% more warm season rainfall during evening and early morning hours, possibly because of the URE. Diem (2008) also confirmed an anomaly in the northeast Atlanta suburbs.

Stallins and Rose (2008) discussed increased lightning activity downwind of Atlanta and in other cities in the US, Spain, and Brazil. Lightning is typically associated with convective rainfall. Rose et al. (2008) combined lightning data

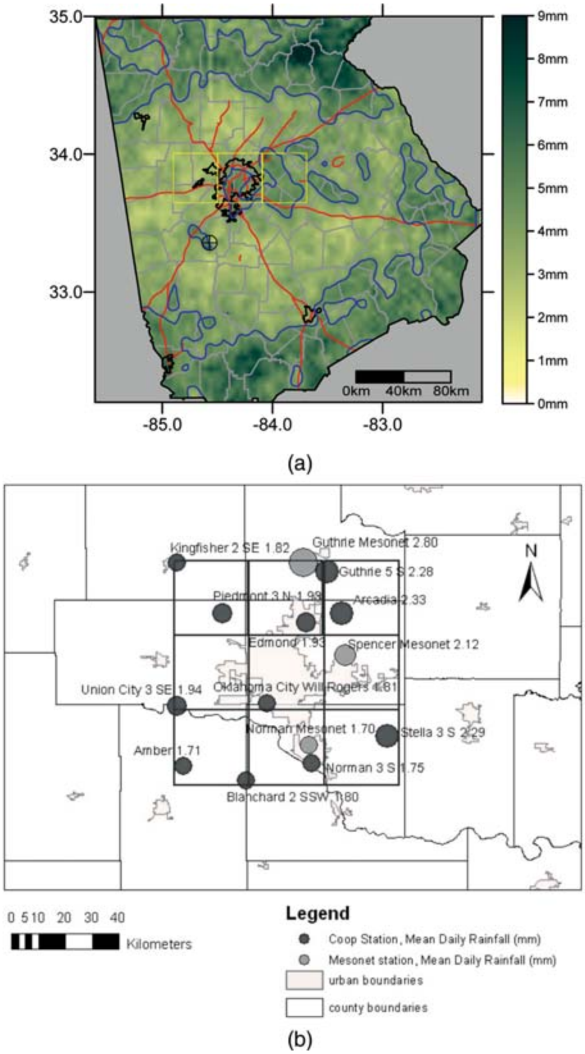



**Fig. 14.3** (**a**). Warm season radar-derived rainfall anomalies downwind (east) of Atlanta (following Mote et al. 2007). *Darker green* shades represent larger cumulative totals. (**b**). Warm season mean anomalies (1998–2006) in rainfall around Oklahoma city using a mesoscale gauge network and TRMM-based satellite data (following Hand and Shepherd 2009)

with the North American Regional Reanalysis dataset and showed enhancement of lightning and precipitation around Atlanta as a function of prevailing wind. Meng et al. (2007)'s radar-based analysis also showed that, due to the urban effects, thunderstorms associated with a tropical cyclone strengthened when moving over Guangzhou City (China), with maximum radar echoes observed directly over the urban area and precipitation located within the city. Hand and Shepherd (2009) used space-based and rain gauge rainfall estimates (Fig. 14.3b) to reveal a statistically significant rainfall anomaly in the north-northeast (e.g. downwind based on prevailing wind analysis) sections of Oklahoma City (USA) and suburbs.

Though many studies find urban-induced rainfall enhancement, studies have also presented alternative effects on rainfall. Trusilova et al. (2008) used a regional model to simulate European climate sensitivity to urban land cover. They found statistically significant increases (decreases) in winter (summer) rainfall in their urban simulations as compared to the pre-urban settlement runs. Kaufmann et al. (2007) argued that urbanization in the Pearl River Delta of China, a rapidly urbanizing part of the world, has reduced regional precipitation because of changes in surface hydrology. Along similar lines, Zhang et al. (2009, 2007) argued that replacement of natural surfaces with Beijing's urban land cover reduced ground evaporation and evapotranspiration, which reduced rainfall. Guo et al. (2006) also found decreased cumulative rainfall around Beijing. However, these studies, with the exception of a simple analysis in Zhang et al. (2009), did not consider the effects of pollution (aerosols).

Urban pollution has been shown to affect precipitation processes. Smaller cloud droplet size distributions and suppressed rainfall have been shown to occur due to increased aerosol concentrations from anthropogenic sources over and downwind urban areas (Rosenfeld 1999; Rosenfeld 2000; Givati and Rosenfeld 2004; Lensky and Drori 2007; Rosenfeld et al. 2008). Very recent studies (van den Heever and Cotton 2007; Rosenfeld et al. 2007) are beginning to shed light on the possible role of giant cloud condensation nuclei (CCN) (enhancement) and smaller CCN (suppression).

## 14.3 Modeling Studies of the Urban Rainfall Effect

In his comprehensive review of the subject, Lowry (1998) discussed several potential problems with methodology and inferences used in many historical studies of urban-induced precipitation. Lowry recommended that future studies should have at least the following characteristics: "(1) Designed experiments – especially legitimate controls and, where appropriate, stratification schemes – in which explicitly stated hypotheses are tested by means of standard statistical methods, (2) Replication of the experiments in several urban areas, (3) Use of spatially small, and temporally short, experimental units reflecting the discontinuous nature of precipitating systems, and (4) Dissaggregation of standard climatic data to increase sample size and avoid merging effects between dissimilar synoptic weather systems."

Physically based coupled atmosphere-land surface (CALS) modeling work is required to improve basic understanding of weather and climate impacts in the urban zone. CALS modeling also enables controlled, designed experiments with the ability to replicate various meteorological and land cover scenarios. CALS essentially solve an inter-related set of equations describing atmospheric motion, heating, and moisture and the interactions with land surface processes (Fig. 14.4). The land surface model represents various land cover and relevant attributes that couple to atmospheric processes (Fig. 14.5).

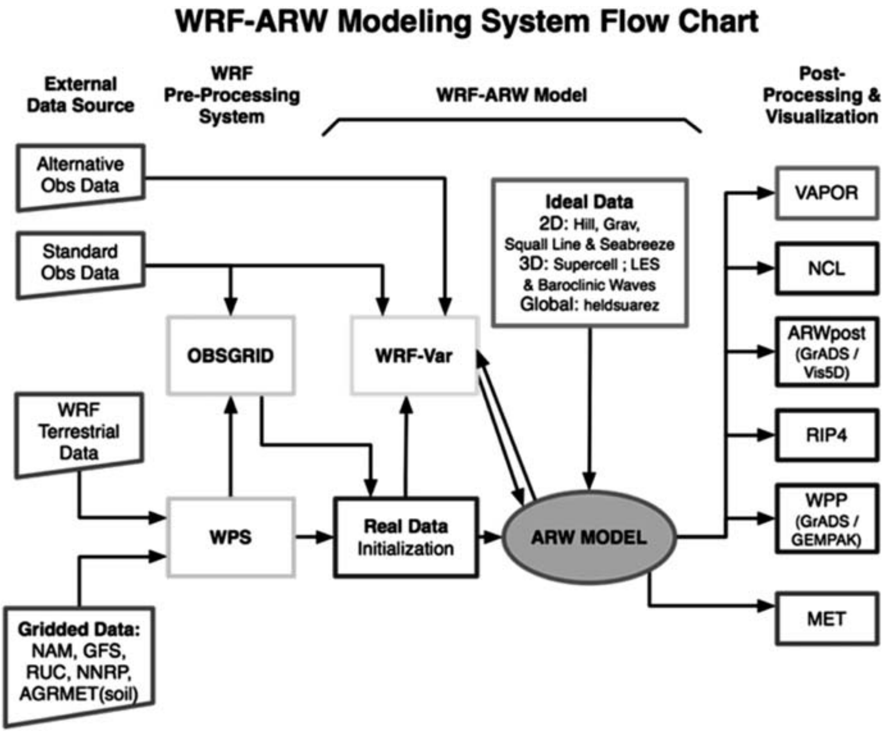



**Fig. 14.4** Typical schematic for a mesoscale weather model (Courtesy of UCAR)

Early numerical model simulations investigating urban rainfall effects can be traced back to the 1970s. Vukovich and Dunn (1978) employed a three-dimensional primitive equation model to show that UHI intensity and boundary layer stability have dominant roles in the development of heat island circulations. Huff and Vogel (1978) found that the urban circulation is primarily related to increased sensible heat fluxes and surface roughness in the urban area. Hjelmfelt (1982) simulated St. Louis' UHI and found positive vertical velocities downwind of the city. He suggested that the surface roughness convergence effect and the downwind augmentation of the UHI circulation by the synoptic flow were the cause. Shafir and Alpert (1990) applied a numerical model to argue that a Jerusalem (Israel) rainfall anomaly existed because of enhanced temperatures, humidity, and cloud condensation nuclei over the city.

Thielen et al. (2000) used a two-dimensional (2D) model to show that surface processes, particularly sensible heat flux, affected the development of precipitation around Paris, France. In their experiments, when UHIs were weak, surface sensible heat fluxes, convergence, and buoyancy variations that influence rainfall development were most effective at a distance from the central heat source. Craig and Bornstein (2002) reported on three-dimensional (3D) mesoscale

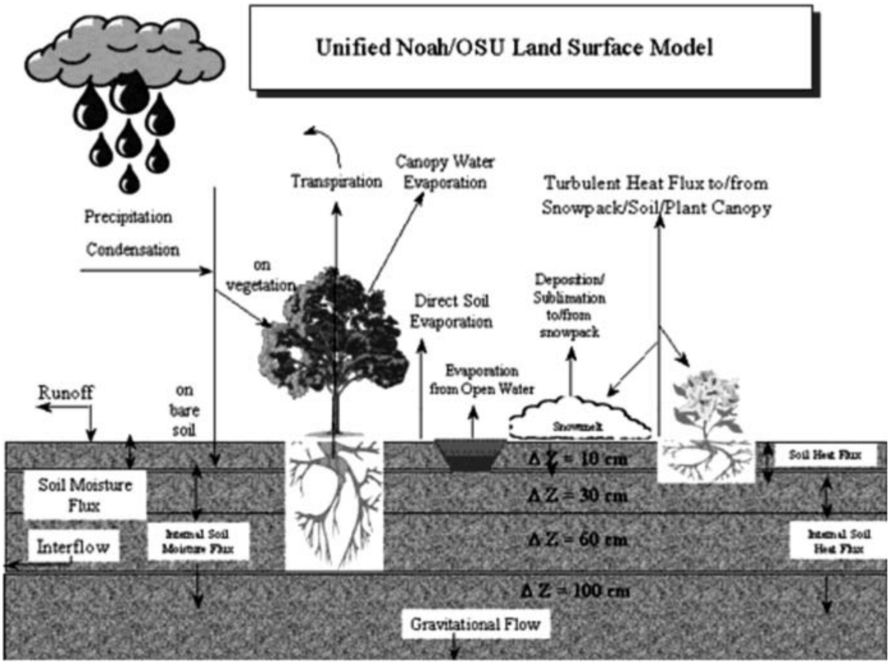



**Fig. 14.5** NOAH land surface model (Courtesy of UCAR)

simulations that showed how the UHI causes convergence that can lead to convection. Rozoff and Cotton (2003), using a storm-resolving model, examined a 1999 storm case in St. Louis to investigate the role of surface convergence mechanisms on initiating deep, moist convection. They found that surface convergence on the leeward side of the UHI played a key role in initiating convection.

Niyogi et al. (2006) simulated a mesoscale system in Oklahoma City, Oklahoma using a land surface model, which included an urban canopy model. Urban canopy models represent the actual three-dimensionality of buildings within an urban area rather than simple albedo, heat, and vegetation characteristics. They found that the city focused the precipitation downwind of the urban region. Ikebuchi et al. (2007) found that changes in urban land cover and anthropogenic heating significantly affected the location and intensity of rainfall for a heavy convective event in the Tokyo (Japan) metropolitan area. Baik et al. (2007) using a nonlinear numerical model and a two-layer linear analytical model found that as the atmospheric boundary layer becomes less stable, a downwind updraft cell induced by the urban heat island strengthens. Their results offered an explanation for possible daytime urban-induced rainfall anomalies by demonstrating that as the boundary layer destabilizes the height of the maximum updraft velocity and the depth of the downwind updraft circulation increase. During daytime nearly neutral or weakly stable conditions, UHI circulations can be relatively strong even though the UHI magnitude is relatively weak. Han and Baik (2008) extended this research to three-dimensions and resolved an internal gravity wave field with an upward branch just downwind of the theoretical heating center (e.g. the city). Zhang et al. (2009) used a coupled modeling system to describe how Beijing's expanding urban land cover promotes less evaporation, larger temperatures, and more sensible heat flux. The authors argue that these changes have led to a reduction in warm season rainfall. They also experimented with increased vegetation as a mitigation strategy. This result is contrary to many previous model results but is interesting to consider for future debates.

## 14.4 Atlanta and Houston Case Studies

What is clear from these modeling studies is that urban land cover alters the thermodynamic and dynamic process of the atmospheric surface and boundary layer and can certainly alter convective processes under certain conditions. What is not clear from the literature is what set of factors (convergence, fluxes, boundary layer destabilization, or aerosols) is most important and under what circumstances. To investigate these questions further, we investigated, using modeling approaches, two different urban settings in which the literature has identified possible UREs: Case 1: Atlanta (USA), a rapidly urbanizing inland city and Case 2: Houston (USA), a rapidly urbanizing coastal city.

### *14.4.1 Atlanta*

The Atlanta metropolitan region has been the focus of several studies aimed at understanding the impact of urbanization on warm season rainfall. Additionally, Yang and Lo (2003) suggested that if Atlanta, a typical post-modern North American city, continues to grow at the same rate, the net increase in urban land from 1999 to 2050 will be approximately 928,379 ha (~50 ha per day rate of increase), representing an increase of 254% for the entire period.

#### 14.4.1.1 Model Configuration and Approach

Previous studies have hypothesized that urban land cover initiates or alters convection over Atlanta on some case dates (Bornstein and Lin 1999; Shepherd et al. 2002; Dixon and Mote 2003). More recently, Mote et al. (2007) and Diem (2008) have continued to verify rainfall anomalies in the eastern and northeastern suburbs of Atlanta using rain gauge and radar data. All of these studies were observational and therefore lacked the ability to look into the hypothetical scenarios such as the behavior of convection in a different or modified land cover type. Bornstein and colleagues simulated a 1996 case and reported it in a conference proceeding. Recently, Shem and Shepherd (2009) simulated the impact of urban land cover on Atlanta convection using a couple mesoscale atmosphere-land surface model.

They employed the Weather Research and Forecast (WRF) Model-NOAH model to simulate convective precipitation for two cases. The Weather Research and Forecasting (WRF) model (see Fig. 14.4), as described in Shem and Shepherd (2008), "is a mesoscale model with advanced dynamics, physics and numerical schemes. It is a compressible, non-hydrostatic, Euler equation, mesoscale meteorological model. WRF features multiple dynamical cores, a 3-dimensional variational (3DVAR) data assimilation and a software architecture accommodating computational parallelism and system flexibility. It is viable across scales ranging from meters to thousands of kilometers." The Advanced Research WRF (ARW) core version used in this study has in several ways replaced the Pennsylvania State University and National Center for Atmospheric Research (NCAR) mesoscale numerical model (MM5) as a tool in the analysis and prediction of the urban-scale local weather. Although MM5 is still quite viable.

The NOAH land surface model (Ek et al. 2003) is a joint effort by the National Center for Atmospheric Research (NCAR), National Center for Environmental Prediction (NCEP) and the Air Force Weather Agency (AFWA). NOAH (Fig. 14.5) contains four soil layers, thirteen vegetation covers and nine different soil types. NOAH incorporates bulk parameterization for urban land cover via alterations in roughness, surface albedo, volumetric heat capacity, soil thermal conductivity and green vegetation fraction. The simulations incorporated a simple urban land cover parameterization (Tewari et al. 2004) which includes: (1) roughness length-0.8 m, to represent mechanical turbulence generated buildings roughness elements and drag due to buildings; (2) surface albedo-0.14, to represent radiation processes in urban canyons; (3) volumetric heat capacity-3.0 J m-3 K-1, for the urban surfaces; (4) soil thermal conductivity-3.24 W m-1 K-1, to parameterize urban heat storage; and (5) an overall reduction of green vegetation fraction. The reader is referred to Shem and Shepherd (2009) for the specific model configuration (e.g. boundary conditions, parameterizations, etc.).

Three distinct land cover scenarios were used as sensitivity experiments. URBAN represented land cover based on the 30 (s) 1994 United States Geological Survey (USGS) dataset (Fig. 14.6). LARGE-URBAN characterized the expansion of the city of Atlanta to its size circa 2005. NOURBAN was a simulation in which the city of Atlanta was removed and replaced by the dominant land cover type of the surrounding rural location, i.e., "dryland-cropland-pasture".

#### 14.4.1.2 Results

We summarize results from Shem and Shepherd (2009)'s simulation of the convective event on 26 July 1996. National Weather Service (NWS) surface

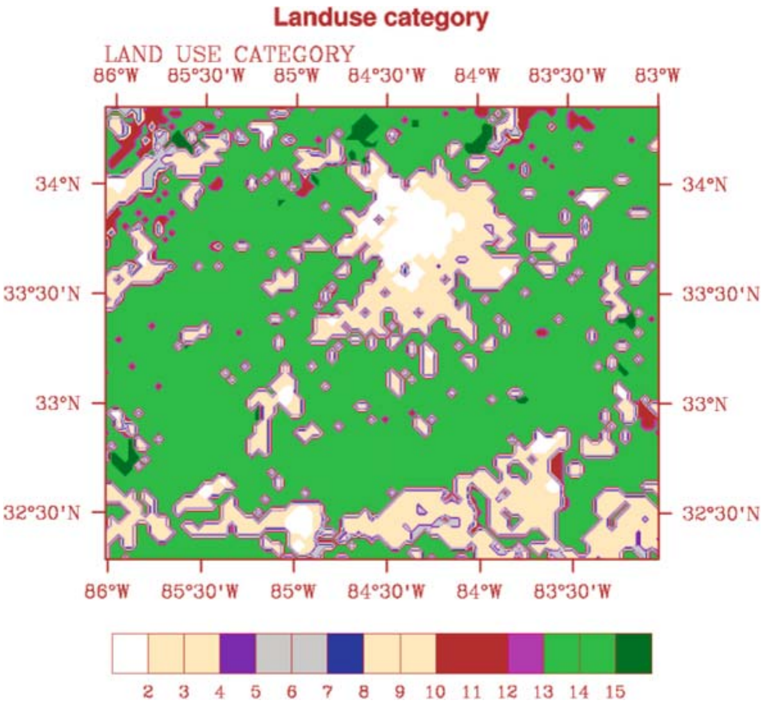



**Fig. 14.6** Atlanta's urban land cover circa 1994 (*white*). For the 2005 case, the *light brown* region around Atlanta is converted to urban (from Shem and Shepherd 2009)

maps indicated a weak quasi-stationary or cold front over the southern USA (Bornstein and Lin 2000). URBAN results are compared to NWS radar observations for the Atlanta area. The model initiates precipitation about 2 h earlier than observed rainfall in the southern tip of Atlanta, i.e. about 25 km from the city center. However, the model did resolve the approximately 6-h duration of the event. The rainfall intensity in the model is not as strong as in the radar data, however the model adequately resolves the spatio-temporal evolution with enough accuracy for analysis. A weak daytime UHI center was observed slightly northeast of Atlanta as a result of heat advection from the city by a southwesterly regional flow (Bornstein and Lin 2000).

Shepherd et al. (2009) recently suggested that convergence can result from urban-induced mechanical turbulence and the urban-induced pressure perturbation. Such convergence could lead to convection. Simulated low-level convergence (Fig. 14.7) revealed a notable convergence ring around the boundary of the city, which may have been caused by the increased surface roughness of the urban surface but more likely, the leading edge of an urban circulation.

Rainfall initiation time for both scenarios is similar for the URBAN and NOURBAN case, which suggests that enhancement (not initiation) by the urban land cover was important on this day. Analysis of cloud fields and

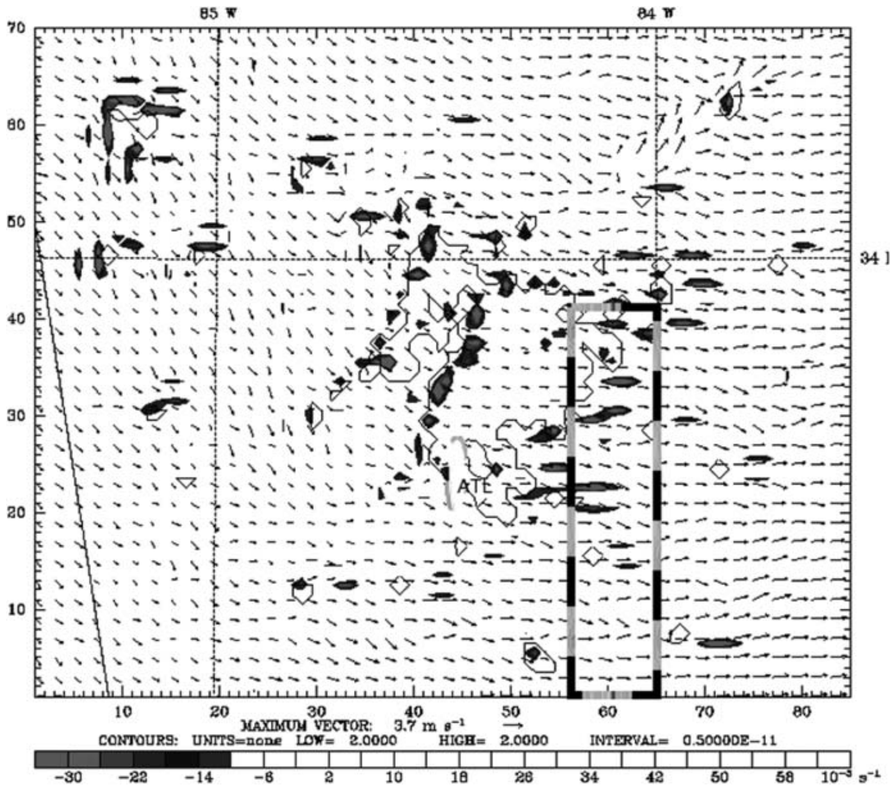



**Fig. 14.7** Enhanced total accumulated rainfall zone -eastern side (*selected rectangle*) of the city superimposed on the difference plot for convergence ($10^{-5}$ $s^{-1}$) between the URBAN and NOURBAN scenarios. The *selected rectangle* covers a strip approximately 20–50 km from the city center – marked ATL (from Shem and Shepherd 2009)

vertical velocity revealed a higher cloud depth in the NOURBAN scenario even though the lifetime of the clouds is almost the same in both scenarios. The shallower cloud depths in the URBAN case could be attributed to reduced moisture content over the urbanized area (see aforementioned studies by Zhang et al. 2009). Cumulative rainfall (Fig. 14.8) was enhanced in the downwind section (e.g. east-southeast) of the city. The increase in the total cumulative rainfall in the downwind section was 13%, which is consistent with findings from Shepherd et al. (2002) and Shepherd (2006).

The model simulation captured the evolution and propagation of the UHI center. The maximum UHI ranged from 2.0 to 2.5 K from 1200 to 1430 EDT after which convection was initiated (Bornstein and Lin 2000). Convection gradually weakened the UHI for about 2 h due to reduction in solar insolation.

A comparative analysis of the three scenarios indicated that convection on this day was likely modified rather than initiated by the urban land cover. Both

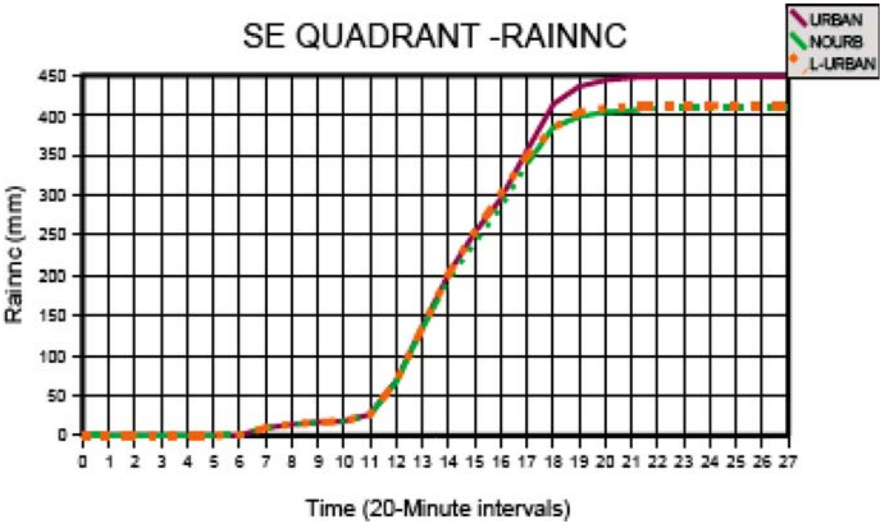



**Fig. 14.8** Simulated accumulated rainfall totals for the southeast quadrant of the study region (from Shem and Shepherd 2009)

the domain-wide and quadrant-by-quadrant analyses indicated that the initiation time for the thunderstorms were the same for both the URBAN and NOURBAN scenarios. However, notable differences in the temporal evolution of the thunderstorms begin after about 2 h. Shem and Shepherd (2009) discuss the possible mechanisms related to this urban enhanced storm.

This study represents an iterative process in continued numerical model case studies aimed at understanding the impact of urbanization on convective thunderstorm formation. The main finding was that convective initiation on this particular day was not likely due to urban land cover. However, urban land cover likely enhanced rainfall in the area, which was also reported in studies by Westcott (1995), Pyle et al. (2007), and Niyogi et al. (2006). The research, like others, confirmed that the UHI magnitude is relatively weak during the most convectively active part of the day. However, sensible heat flux and boundary layer dynamics do respond to the presence of urban land cover. Baik et al. (2007) found as the boundary layer destabilizes, the height of the maximum updraft velocity and the depth of the downwind updraft circulation increase. Under daytime nearly neutral or weakly stable conditions, UHI circulations can be relatively strong even though the UHI magnitude is relatively weak. Some of the major limitations of this study include the representation of the urban region via bulk modification of specific variables like roughness length, volumetric heat capacity, albedo and soil thermal conductivity. More research using a more detailed urban classification is necessary. Further, aerosol representation is required to fully understand land cover-aerosol interactions. Van den Heever and Cotton (2007) provide a nice framework for how to achieve this requirement.

## *14.4.2 Houston*

Houston, Texas sits on the 5,000 km$^2$ Gulf Coastal Plain with an elevation of 27 m above sea level. The eastern third of the state of Texas including Houston (upwind and downwind) is considered a sub-tropical humid climate. Southeast Texas receives roughly 140 cm of rain annually (Lyons 1990). Since Houston is located near Galveston Bay and the Gulf of Mexico, sea breeze (SB) circulations, particularly during the warm season, influence its weather. Banta et al. (2005) described the typical evolution and structure of the Houston area sea-bay breeze (SBB).

Burian and Shepherd (2005), Shepherd and Burian (2003), and Orville et al. (2001) showed that interactions between urban and sea breeze circulations may contribute to observed rainfall and lightning anomalies around Houston. McPherson (1970) recorded that a preferred location for convective initiation and enhancement is in the Galveston Bay area east of Houston. His results employed a modeling study of the effect of complex coastline on the sea breeze circulation. However, he did not explicitly account for urban land cover effects.

Other modeling studies have explored the interaction of the sea breeze-urban circulations (Yoshikado 1994; Kitada et al. 1998; Kusaka et al. 2000; Ohashi and Kida 2002; Lo et al. 2007). Yoshikado (1994) found that the daytime UHI persists under the influence of the sea breeze when large-scale flow is weak. Also, the sea breeze front was found to linger over the city as a result of UHI effects. Yoshikado (1994) noted that heavily urbanized regions greater than 10 km in width interact with thermally direct circulations. His model simulations showed maximum vertical velocity in the convergence zone of the sea breeze and UHI circulation growing over the inland side of Tokyo, Japan around 1,200 LST. Using pre-European settlement (pre-urban) and current satellite-derived (post-urban) data in the land surface model, Gero and Pitman (2006) found that a simulated storm near Sydney was not present in the pre-settlement run. However, their factorial analysis suggested that agricultural land, not urbanization was possibly the forcing mechanism for the storm. Lo et al. (2007) demonstrated that Hong Kong caused increased temperature gradient and enhancement of the mesoscale sea breeze circulation.

#### 14.4.2.1 Configuration and Approach

On 25 July 2001, weak convection initiated in the early afternoon as the SB front propagated inland. At 1,200 UTC, Houston was experiencing a typical southeasterly flow at the surface. A high-pressure system was situated east of Florida and its broad clockwise flow regime was the primary contributor to the prevailing surface level flow in the Gulf of Mexico.

The Lake Charles, Louisiana (LCH) site, the most representative upper-air sounding (not shown), indicated that the lower troposphere was relatively moist from the surface to approximately 700 hPa and that upper-level flow was generally easterly. A moist lower troposphere (along with a source of lift and instability) is a necessary condition for the development of convection. Common meteorological indicators suggested that the atmosphere would support convective development on this day. Convective Available Potential Energy (CAPE) was 1978 J/kg and the atmosphere was marginally unstable based on lifted index (–2.24), K-index (32.6), and Total Totals (38.7) for the Lake Charles-Houston area.

National Weather Service WSR-88D Doppler radar data (Fig. 14.9) at 2,037 UTC indicated that moderate convection developed over the Houston region on July 25 2001. Convection along the SB front started in the early afternoon hours around 1,700 UTC. By 2037 UTC, there is scattered convection along the SB front, including an area intense convection west-northwest of Houston.

The PSU/NCAR mesoscale model (MM5) release 3–7, used for this study, was developed in cooperation with The Pennsylvania State University and the University Corporation for Atmospheric Research (Grell et al. 1994). MM5 is a

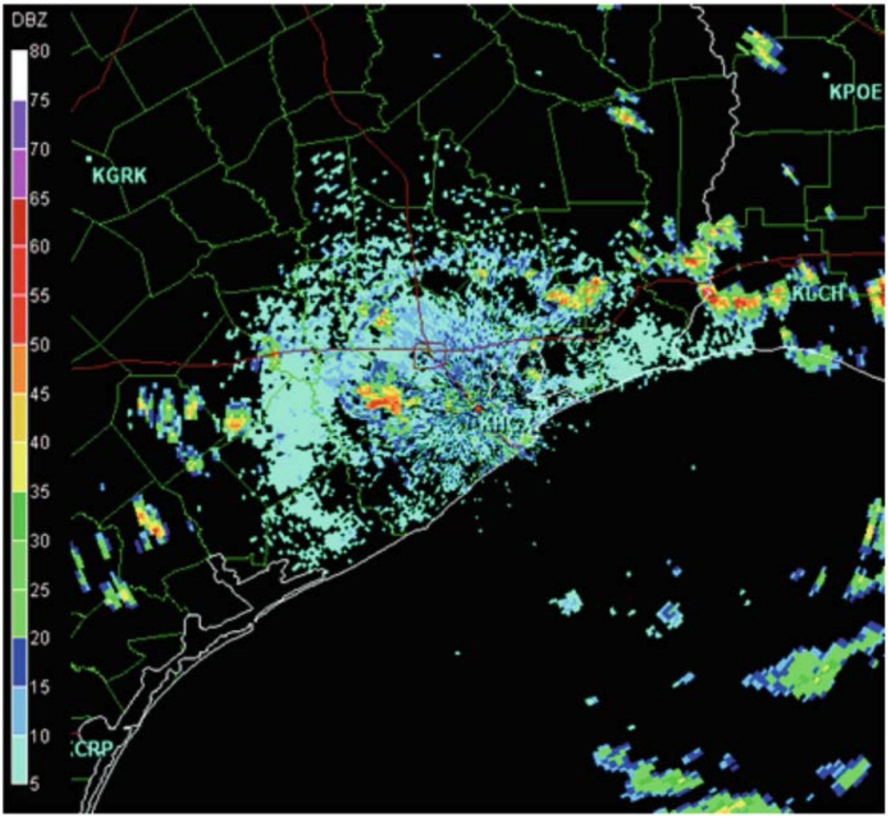



**Fig. 14.9** National weather service WSR-88D radar (KHGX) data at 2037 UTC (following Shepherd et al. 2009)

limited-area, non-hydrostatic, terrain-following sigma-coordinate model designed to simulate or predict mesoscale atmospheric circulation. It contains sophisticated microphysics, radiation, and turbulence capabilities and is compatible with several land surface models. MM5 is very similar, conceptually, to WRF (Fig. 14.4). The aforementioned NOAH land surface model (Fig. 14.5) was also used.

The Houston-Galveston Area Council (H-GAC, http://www.h-gac.com) projected Houston's urban land cover growth to the year 2025 using the UrbanSim model (Waddell 2002). UrbanSim is a model that predicts dynamic urban growth at land parcel or household level. The reader is referred to Teerarojanarat et al. (2004) for the historical origin of UrbanSim its characteristics. An important feature of UrbanSim is that it considers key actors in the urban development processes: (1) household and business actors-reflecting consumer preferences for place or location and (2) developer actors-reflecting where and what type of construction is built. Government, political, and environmental constraints are inputs designed to restrict development activity. In UrbanSim, urban development is represented as an interaction between market behavior and governmental projections. This feature is quite useful for assessing the impacts of alternative governmental plans and policies related to land use. UrbanSim differs from economic and spatial-interaction models that rely on cross-sectional equilibrium solutions using large geographic districts. However, it is very similar to the CUF-2 model (Landis and Zhang 1998a, b) because the approach is disaggregate and based on predicting changes over smaller time steps. However, it does vary in that it explicitly represents the demand for real estate at each location, the actors, and choice processes that influence patterns of urban development.

The approach was to simulate the July 25–26, 2001 case day (Shepherd et al. 2009). We conducted three experiments: URBAN, NOURBAN, and URBAN2025 (Fig. 14.10). The URBAN simulation was based on the United States Geological Survey (USGS) land cover and represents Houston's urban extent during the mid-1990s. Changes in land cover type are translated to the model through surface roughness, vegetation, albedo, soil moisture, and thermal properties. For example, roughness lengths in cities are significantly larger than dryland-cropland-pasture due to buildings. Urban regions also have less vegetation, smaller albedo, and larger heat storage. NOURBAN simulation replaced the urban land cover with the dryland-cropland-pasture category. For the URBAN2025 experiment, land cover projected to 2025 using UrbanSim is implemented in the MM5-NOAH modeling framework.

#### 14.4.2.2 Results

The 25 July 2001 case day was used to examine the factors that led to urban-initiated convection on the western fringe of Houston and to assess how such convection responded to a hypothetical future urban land use scenario. A

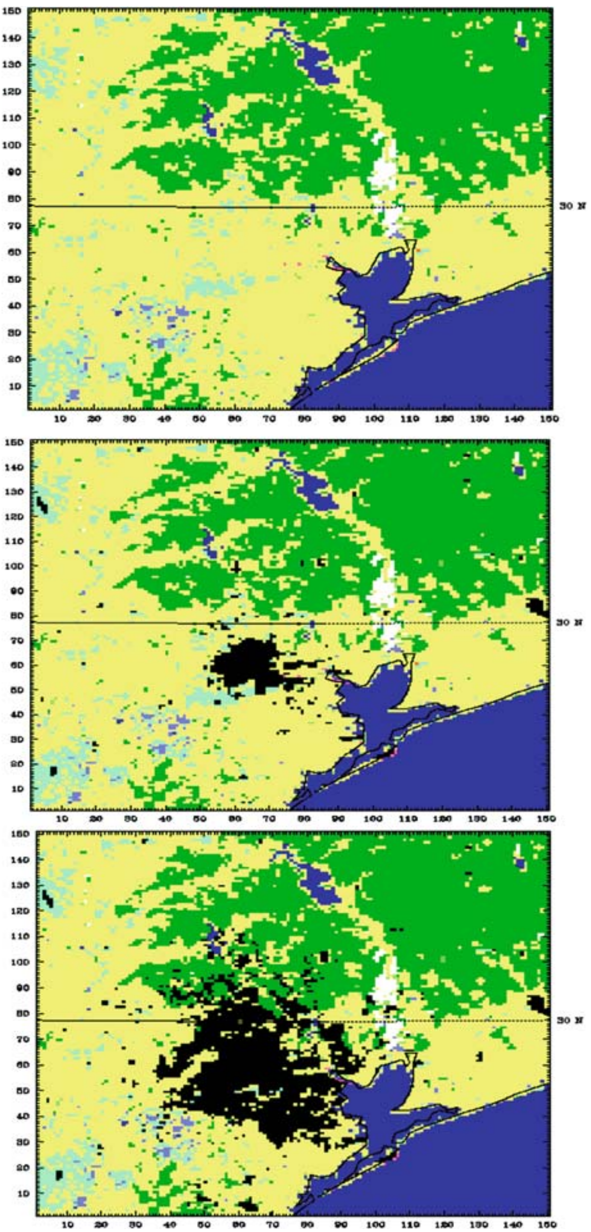



**Fig. 14.10** United States geological survey land use for experiments NOURBAN (*top*), URBAN (*middle*), and URBAN2025 (*bottom*). The land cover represented are urban (*black*), evergreen needleleaf (*dark green*), dryland-cropland pasture (*yellow*), deciduous broadleaf (*white*), crop-grass mosaic (*light green*), mixed dry-irrigated (*lavender*)

comparison of model generated and observed surface temperatures and wind indicate that the model is reproducing real conditions well (Fig. 14.11). Further results corroborated and validated recent observations that rainstorms can be enhanced or initiated over Houston.

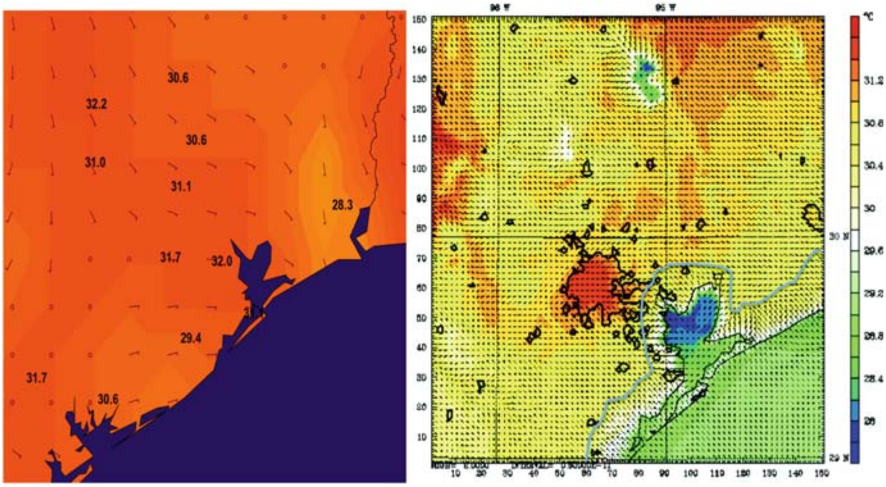



**Fig. 14.11** Observed (*left*) and simulated (*right*) temperature (°C) and surface wind fields (*left*) at 1,640 UTC on 25 July 2001. The *blue line* is the approximate location of the sea breeze front (following Shepherd et al. 2009). Temperature scale ranges from 27.5–32°C

Figure 14.12 shows the differences in model radar reflectivity (convection) on the northwest fringe of Houston in URBAN that is not present in NOURBAN at time 2037 UTC. The URBAN run generates a the sea breeze convection and the urban storm on the northwest fringe of the city. NOURBAN captures the sea breeze convection but not the urban storm. It should be noted that the model storm is slightly offset from the location of the observed storm. Slight mismatches in space and time are very typical for model-generated rainfall. This results in a more robust rainfall accumulation over the storm's history. Our results also confirmed the response of the boundary layer to urban surface processes and illustrated how surface fluxes and convergence can translate energy to urban boundary layer convective cells. Other results clearly indicated the enhanced (reduced) sensible (latent) heat flux over the city. The decreased latent heat flux reflects the relative lack of vegetation in the city. These signatures are indicative of the UHI and related land-atmosphere interactions. Convergence zones (green colors) along the fringe of the city (and along the sea breeze), the presence of the UHI as indicated in the potential temperature field (Fig. 14.13), and urban pressure low (Fig. 14.14) were critical factors that led to enhanced convection independently or in conjunction with the sea breeze. Further, we found supporting evidence that secondary outflow interactions with urban and non-urban convergence zones also contribute to more vigorous and widespread precipitation. As such, future urban land cover growth (using 2025 urban land cover) could lead to temporal and spatial precipitation variability in Houston and other coastal urban microclimates (Fig. 14.15). More work is needed to determine what the tipping point is for an urban area to affect

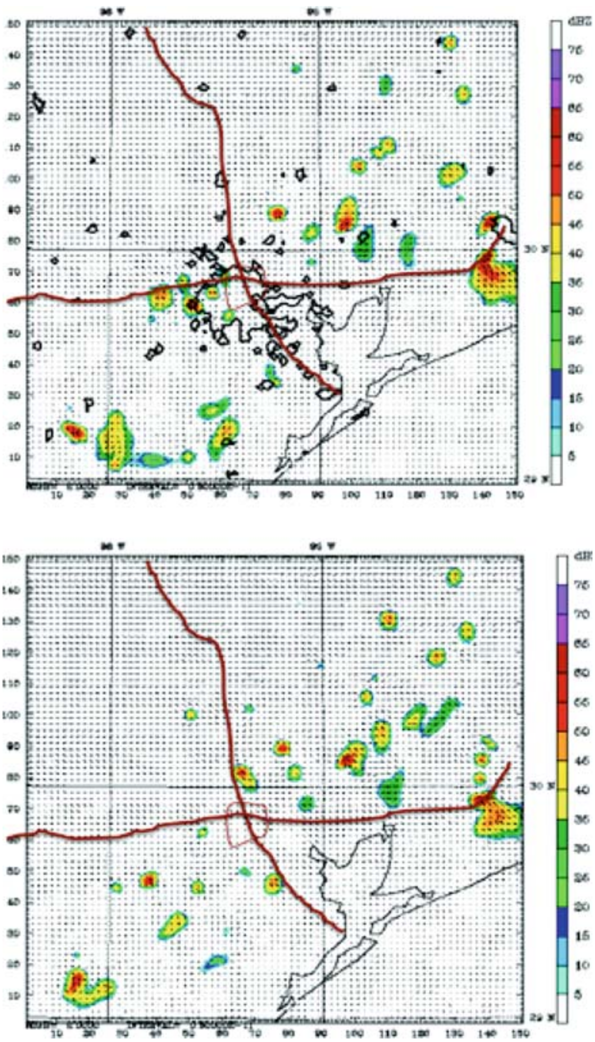


**Fig. 14.12** Simulated radar reflectivity (dBZ) fields at 2,040 UTC for URBAN (*top*) and NOURBAN (bottom). The *black line* outlines Houston urban land. The *red lines* represent the road network in Fig. 14.1 and provide a reference for comparison (following Shepherd et al. 2009)

convective processes differently and to further isolate what urban factor (temperature, roughness, or others) contribute and in what relative amounts (Shepherd et al. 2009)

It should be noted that this study has several weaknesses that must be addressed in future. First, the projected urban land use in 2025 should reflect more realistic heterogeneity of urban land classes and morphology. Second, aerosol populations and pollution need to be considered. It is becoming more apparent that urban-induced dynamic processes and aerosol populations (e.g. through cloud condensation nuclei) contribute to precipitation variability around cities (Shepherd 2005; van den Heever and Cotton 2007). Third, a

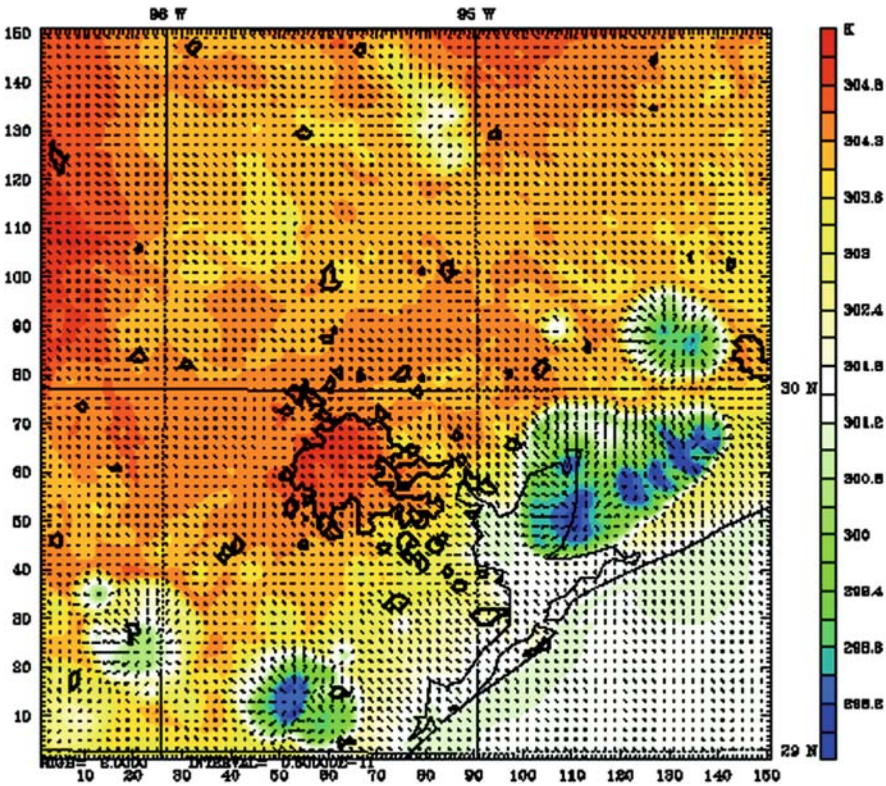



**Fig. 14.13** Potential temperature (K) at 0.5 m above the surface at 1,840 (UTC) *bottom*. *Black vectors* represent near surface wind flow (following Shepherd et al. 2009)

broader array of case days should be considered. Finally, future projection scenarios should attempt to reflect future trends in both air temperature and sea surface temperature (e.g. warming or cooling) to understand how the dynamic urban environment interacts with the background environment.

## 14.5 Recommendations to Improve Model Studies of the Urban Rainfall Effect

Shepherd (2005) offered a series of important recommendations that will improve our ability to model and understand the urban rainfall effect. They are worth reproducing here with some slight updates:

1. "New observing systems to monitor and track anthropogenic and natural aerosols, land cover/land use changes, cloud microphysics, and precipitation processes.

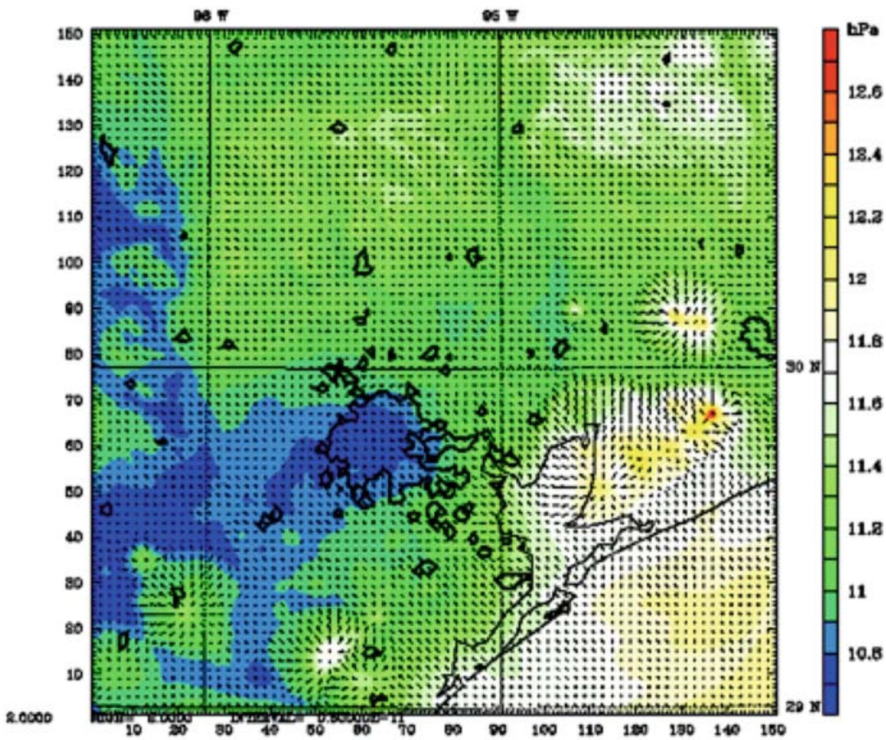



**Fig. 14.14** Perturbation pressure (hPa) field at 1,840 UTC for URBAN (following Shepherd et al. 2009)

2. Modeling systems that explicitly resolve aerosols, cloud microphysics, complex land surfaces, and precipitation evolution so that a more conclusive understanding of the feedbacks and interactions can be attained. Ongoing work by van den Heever and Cotton (2007) is advancing our ability to assess aerosol effects in mesoscale models. Khain et al. (2004) employed spectral (bin) microphysics to simulate aerosols in cloud-mesoscale modeling systems. Molders and Olson (2004) also demonstrated methodologies for introducing aerosols into mesoscale modeling systems. Additionally, high performance land surface modeling and data assimilation systems like the NASA Land Information System (Jin et al. 2007; Peters-Lidard et al. 2004) embody the need for improved representation of complex and heterogeneous land surfaces. The obvious next steps involve coupling these land surface, mesoscale, and aerosol-microphysics schemes.
3. Implementation of urban parameterizations at the local scale to resolve urban canyon, dynamics, and flux processes, particularly in terms of roughness, surface cover properties, low-level moisture associated with irrigation, and aerosols. Though it is not clear what role local urban canopy scale dynamics (UCD) play in larger scale precipitation processes, it is imperative

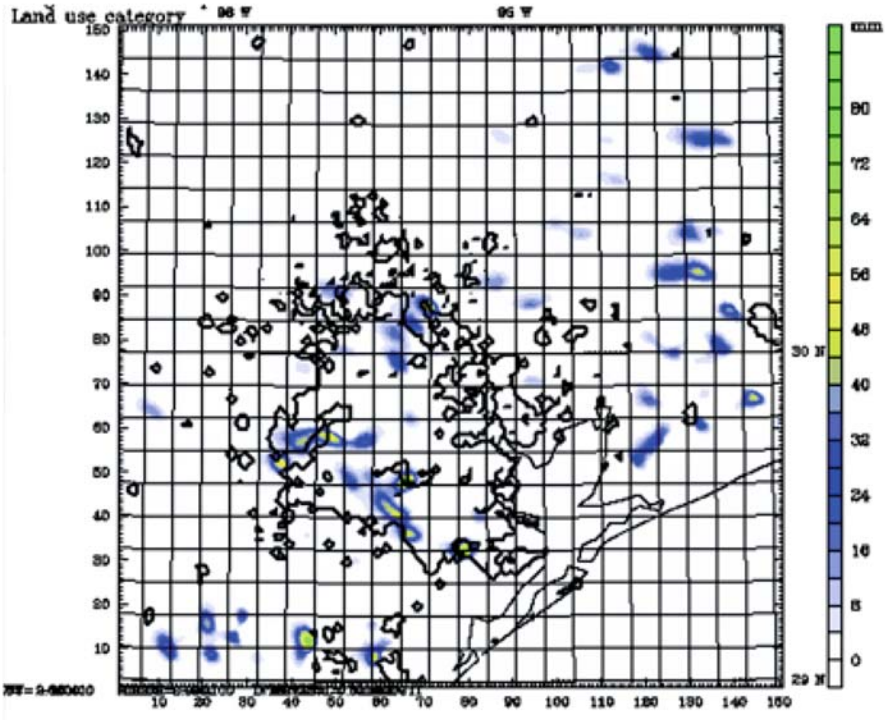



**Fig. 14.15** Accumulated rainfall difference (mm) between URBAN2025 and URBAN simulations over the 6-h period 1,740–2,340 (UTC) (following Shepherd et al. 2009)

to enable a capability to test UCD impacts. It is encouraging that more detailed urban parameterized modeling studies are now available (Zhang et al. 2008; Holt and Pullen 2007) although most did not evaluate precipitation effects. For example, Dupont et al. (2004) identified the importance of complex surface characterization and urban canopies on predicting several meteorological fields but precipitation was not included. Kusaka and Kimura (2004) embedded an urban canopy model in a mesoscale model and demonstrated that canyon structure was as significant as anthropogenic heating in producing a nocturnal heat island. Yet, this paper did not focus on wet microphysical (e.g. cloud or precipitation) processes either.

4. Field studies to validate satellite observations and modeling simulations of urban-precipitation processes and to extend basic understanding of the processes involved. The SPRAWL and Project Atlanta experiments reported in Shepherd et al. (2004) and Quattrochi et al. (1998) are the types of field experiment required to isolate specific synoptic flow regimes, mesoscale conditions, and precipitation cases. Field studies enable detailed interrogation of the underlying physical processes for specific cases and candidate case days for modeling studies.

5. Climate modeling systems that adequately characterize the urban environment. As the evidence that urban-areas impact precipitation mounts, the climate modeling community must begin to represent urban land surface and aerosol processes to better understand the aggregate influences of built-up land and urban aerosols on short and long term climate change (Jin and Shepherd 2005, Jin et al. (2007)."

## 14.6 Current and Future Advances in Modeling Urban Effects on Precipitation

Advances are emerging that address and extend some of the aforementioned recommendations. The single-layer urban canopy model (UCM) of Kusaka and Kimura (2004) and the Town Energy Budget (TEB) model (see Niyogi et al. 2006) are readily available and compatible with popular mesoscale modeling systems like WRF and RAMS and allow for explicit representation of urban morphological parameters and energy exchanges. Teams at the University of Georgia, University of Utah, and University of Maryland have evaluated the effect of remotely sensed urban morphological parameters on the WRF-NOAH-UCM simulations in Houston.

The National Center for Atmospheric Research (NCAR) is experimenting with linking WRF/NOAH/UCM with more advanced computational fluid dynamics (CFD) models. CFD models (Baik et al. 2003) simulate fluid dynamic processes at very high spatial resolution often needed to describe flow in complex urban terrains. Reynolds et al. (2008) coupled an atmosphere-land surface model's precipitation output with a hydrological model to assess the impact of urban land cover on atmospheric and terrestrial hydrologic processes.

These efforts will certainly move the science understanding forward. Additionally, knowledge of how urban environments impact precipitation will undoubtedly have implications for how urban areas are represented in future generations of weather and climate models; how urban planners and water resource managers plan cities; how agricultural activities carried out; and numerous other societal benefit applications.

**Acknowledgments** We would like to thank the NASA Precipitation Measurement Missions program for supporting the research that lead to this chapter.

## References

Baik, J.J., Kim, J.J. & Fernando, H.J.S. (2003). A CFD model for simulating urban flow and dispersion. *Journal of Applied Meteorology*, 42, 1636–1648

Baik, J.J., Kim, Y.H., Kim J.J. & Han, J.Y. (2007). Effects of boundary layer stability on urban heat island- induced circulation. *Theoretical and Applied Climatology*, 89, 73–81

Banta, R.M., Senff, C., Nielsen-Gammon, J., Darby, L., Ryerson, T., Alvarez, R., Sandberg, S., Williams, E., & Trainer, M. (2005). A bad air day in Houston. *Bulletin of the American Meteorological Society*, 86, 657–669

Bornstein, R. & LeRoy, M. (1990). Urban barrier effects on convective and frontal thunderstorms. *Preprints, Conference on Mesoscale Processes, Boulder, CO, American Meteorological Society*, 25–29

Bornstein, R. & Lin, Q. (2000). Urban heat islands and summertime convective thunderstorms in Atlanta: three case studies. *Atmospheric Environment*, 34, 507–516

Braham, R.R. (1981). Urban precipitation processes. *METROMEX: A review and summary, Meteorological Monograph*, 40, 75–116

Burian, S.J. & Shepherd, J.M. (2005). Effect of urbanization on the diurnal rainfall pattern in Houston. *Hydrological Processes*, 19, 1089–1103

Chahine, M. (1992). The hydrological cycle and its influence on climate. *Nature*, 359, 373–338.

Changnon, S. A. (1979). Rainfall changes in summer caused by St. Louis. *Science*, 205, 402–404

Changnon, S.A., Huff, F.A., Schickedanz, P.T. & Vogel, J.L. (1977). Weather anomalies and impacts. *METROMEX Vol. I. Illinois state water survey bulletin*, 62, 260

Changnon, S.A., Semonin, R.G., Auer, A.H., Braham, R.R. & Hales, J. (1981). *METROMEX: A Review and Summary*. *Meteorological Monograph*, 40, 181

Changnon, S.A., Shealy, R.T. & Scott, R.W. (1991). Precipitation changes in fall, winter and spring caused by St. Louis. *Journal of Applied Meteorology*, 30,126–134

Chen, T.C., Wang, S.Y. & Yen, M.C. (2007). Enhancement of afternoon thunderstorm activity by urbanization in a valley: Taipei. *Journal of Applied Meteorology and Climatology*, 46, 1324–1340

Craig, K. & Bornstein, R. (2002). MM5 simulation of urban induced convective precipitation over Atlanta. Preprint Vol., *4th AMS symposium on the Urban Environment*, Norfolk, VA, pp. 5–6

Diem, J.E. (2008). Detecting summer rainfall enhancement within metropolitan Atlanta, Georgia USA. *International Journal of Climatology*, 28(1), 129–122

Dixon, P.G. & Mote, T.L. (2003). Patterns and causes of Atlanta's urban heat island-initiated precipitation. *Journal of Applied Meteorology*, 42(9), 1273–1284

Dupont, S., Otte, T. & Ching, J. (2004). Simulation of meteorological fields within and above urban and rural canopies with a mesoscale model. *Boundary-Layer Meteorology*, 113, 111–158

Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grummann, P., Koren, V., Gayno, G. & Tarpley, J.D. (2003). Implementation of Noah land surface model advances in the national centers for environmental prediction operational mesoscale eta model. *Journal of Geophysical Research*, 108, 8851, doi:10.1029/2002JD003296

Fujibe, F. (2003). Long-term surface wind changes in the Tokyo Metropolitan area in the afternoon of sunny days in the warm season. *Journal of Meteorological Society of Japan*, 81, 141–149

Fujibe, F. & Asai, T. (1980). Some features of the surface wind system associated with the Tokyo heat island. *Journal of Meteorological Society of Japan*, 58, 149–152

Gero, A.F. & Pitman, A.J. (2006). The impact of land cover change on a simulated storm event in the Sydney basin. *Journal of Applied Meteorology and Climatology*, 45, 283–300

Givati, A. & Rosenfeld, D. (2004). Quantifying precipitation suppression due to air pollution. *Journal of Applied Meteorology*, 43, 1038–1056

Grell, G.A., Dudhia, J. & Stauffer, D.R. (1994). A description of the fifth-generation Penn State/NCAR Mesoscale Model (MM5). *NCAR Technical Notes NCAR/TN-398+STR*, p. 122

Guo, X., Fu, D. & Wang, J. (2006). Mesoscale convective precipitation system modified by urbanization in Beijing City. *Atmospheric Research*, 82, 112–126

Han, J.Y. & Baik, J.J. (2008). A theoretical and numerical study of urban heat island–induced circulation and convection. *Journal of the Atmospheric Sciences*, 65, 1859–1877

Hand, L. & Shepherd, J.M. (2009). An investigation of warm season spatial rainfall variability in Oklahoma City: Possible linkages to urbanization and prevailing wind. *Journal of Applied Meteorology and Climatology*, 48, 251–269

Hjemfelt, M.R. (1982). Numerical simulation of the effects of St. Louis on mesoscale boundary layer airflow and vertical motion: simulations of urban vs. non-urban effects. *Journal of Applied Meteorology*, 21, 1239–1257

Holt, T. & Pullen, J. (2007). Urban canopy modeling of the New York City metropolitan area: a comparison and validation of single- and multilayer parameterizations. *Monthly Weather Review*, 135, 1906–1930

Horton, R.E. (1921). Thunderstorm breeding spots. *Monthly Weather Review*, 49, 193

Huff, F. A. (1986). Urban hydrological review. *Bulletin of the American Meteorological Society*., 67, 703–712

Huff, F.A. & Changnon, S.A. (1972a). Climatological assessment of urban effects on precipitation at St. Louis. *Journal of Applied Meteorology*, 11, 823–842

Huff, F.A. & Changnon, S.A. (1972b). Climatological assessment of urban effects on precipitation St. Louis: Part II. Final Report. NSF Grant GA-18781. *Illinois state water survey*, p. 237

Huff, F.A. & Vogel, J.L. (1978). Urban, topographic and diurnal effects on rainfall in the St. Louis region. *Journal of Applied Meteorology*, 17, 565–577

Ikebuchi, S., Tanaka, K., Ito, Y., Moteki, Q., Souma, K. & Yorozu, K. (2007). Investigation of the effects of urban heating on the heavy rainfall event by a cloud resolving model CReSiBUC. *Annuals of disaster Prevention Research Institute, Kyoto University*, 50, 7

Jin, M. & Shepherd, M. (2005). Inclusion of urban landscape in a climate model: how can satellite data help? *Bulletin of the American Meteorological Society*, 86(5), 681–689

Jin, M., Shepherd, M. & King, M.D. (2005). Urban aerosols and their variations with clouds and rainfall: a case study for New York and Houston. *Journal of Geophysical Research*, 110, D10S20, doi:10.1029/2004JD005081

Jin, M., Shepherd, M. & Peters-Lidard, C. (2007). Development of a parameterization for simulating the urban temperature hazard using satellite observations in a climate model. *Journal of Natural Hazards*, DOI 10.1007/s11069-007-9117-2

Kaufmann, R.K., Seto, K.C., Schneider, A., Liu, Z., Zhou, L. & Wang, W. (2007). Climate response to rapid urban growth: Evidence of a human-induced precipitation deficit. *Journal of Climate*, 20(10), 2299

Khain, A., Pokrovsky, A., Pinsky, M., Seifert, A. & Phillips, V. (2004). Simulation of effects of atmospheric aerosols on deep turbulent convective clouds using a spectral microphysics mixed-phase cumulus cloud model. Part I: Model description and possible applications. *Journal of the Atmospheric Science*, 61(24), 2963–2982

Kitada, T., Okamura, K. & Tanaka, S. (1998). Effects of topography and urbanization on local winds and thermal environment in the Nohbi Plain, Coastal Region of Central Japan: a numerical analysis by mesoscale meteorological model with a k− turbulence model. *Journal of Applied Meteorology*, 37, 1026–1046

Kusaka, H. & Kimura, F. (2004). Thermal effects of urban canyon structure on the nocturnal heat island: numerical experiment using a mesoscale model coupled with an urban canopy model. *Journal of Applied Meteorology*, 43, 1899–1910

Kusaka, H., Kimura, F., Hirakuchi, H. & Mizutori, M. (2000). The effects of land-use alteration on the sea breeze and daytime heat island in the Tokyo metropolitan area. *Journal of Meteorological Society of Japan*, 78, 405–420

Landis, J. & Zhang, M. (1998a). The second generation of the California urban futures model: Part 1: model logic and theory. *Environment and Planning B: Planning and Design*, 25, 657–666

Landis, J. & Zhang, M. (1998b). The second generation of the California urban futures model: Part 2: specification and calibration results of the land-use change submodel. *Environment and Planning, B: Planning and Design*, 25, 795–824

Landsberg, H. (1956). The climate of towns. In W.L. Thomas (Ed.), *Man's role in changing the face of the earth*. (Chicago: The University of Chicago Press)

Landsberg, H.E. (1970). Man-made climate changes. *Science*, 170, 1265–1274

Lensky, I.M. & Drori, R. (2007). A satellite-based parameter to monitor the aerosol impact on convective clouds. *Journal of Applied Meteorology and Climatology*, 46, 660–666

Lo, J.C.F., Lau, A.K.H., Chen, F., Fung, J.C.H. & Leung, K.K.M. (2007). Urban modification in a mesoscale model and the effects on the local circulation in the Pearl River Delta Region. *Journal of Applied Meteorology and Climatology*, 46, 457–476

Loose, T. & Bornstein, R.D. (1977). Observations of mesoscale effects on frontal movement through an urban area. *Monthly Weather Review*, 105, 563–571

Lowry, W.P. (1998). Urban effects on precipitation amount. *Progress in Physical Geography*, 22(4), 477–520

Lyons, S. (1990). Spatial and temporal variability of monthly precipitation in Texas. *Monthly Weather Review*, 118, 2634–2648

McPherson, R.D. (1970). A numerical study of the effect of a coastal irregularity on the sea Breeze. *Journal of Applied Meteorology*, 9, 767–777

Meng W., Yen, J. & Hu, H. (2007). Urban effects and summer thunderstorms in a tropical cyclone affected situation over Guangzhou city. *Science in China Series D Earth Sciences*, 50(12), 1867–1876

Molders, N. & Olson, M.A. (2004). Impact of urban effects on precipitation in high latitudes. *Journal of Hydrometeorology*, 5(3), 409–429

Mote, T.L., Lacke, M.C. & Shepherd, J.M. (2007). Radar signatures of the urban effect on precipitation distribution; A case study for Atlanta, Georgia. *Geophysical Research Letters*, 34, L20710, doi:10.1029/2007GL031903

Niyogi, D., Holt, T., Zhong, S., Pyle, P.C. & Basara, J. (2006). Urban and land surface effects on the 30 July 2003 mesoscale convective system event observed in the southern Great Plains. *Journal of Geophysical Research*, 111, D19107, doi:10.1029/2005JD006746

Ohashi, Y. & Kida, H. (2002). Local circulations developed in the vicinity of both coastal and inland urban areas: numerical study with a mesoscale atmospheric model. *Journal of Applied Meteorology*, 41, 30–45

Oke, T.R. (1981). Canyon geometry and nocturnal urban heat island: comparison of scale model and field observations. *Journal of Climatology*, 1, 237–254

Oke, T.R. (1987). *Boundary layer climates* (2nd Edn). (London/New York: Methuen Co.)

Orville, R., Huffines, G., Gammon, J.N., Zhang, R., Ely, B., Steiger, S., Phillips, S., Allen, S. & Read, W. (2001). Enhancement of cloud-to-ground lightning over Houston Texas. *Geophysical Research Letters*, 28, 2597–2600

Peters-Lidard, C. D., Kumar, S., Tian, Y., Eastman, J.L. & Houser, P. (2004). Global urban-scale land-atmosphere modeling with the land information system, symposium on planning, nowcasting, and Forecasting in the urban zone, *84th AMS Annual Meeting 11–15 January 2004 Seattle*, WA, USA

Pyle, P., Niyogi, D., Arya, S., Shepherd, M., Chen, F. & Wolfe, B. (2007). An observational and modeling-based storm climatology assessment for the Indianapolis urban region. *Journal of Applied Meteorology and Climatology*. (conditionally accepted)

Quattrochi, D., Luvall, J., Estes, M., Lo, C., Kidder, S., Hafner, J., Taha, H., Bornstein, R., Gillies, R. & Gallo, K. (1998). Project Atlanta (Atlanta land use analysis: Temperature and air quality): A study of how urban landscape affects meteorology and air quality through time. *Preprint Volume. Second AMS Urban Environment Conference*, Albuquerque, NM, pp. 104–107

Reynolds, S., Burian, S., Shepherd, J.M. & Manyin, M. (2008). Urban induced rainfall modifications on urban hydrologic response. W. James, Editor in Chief, co-edited by Irvine, K.N., McBean, E.A., Pitt, R.E., and Wright, S.J. In Reliable modeling of urban water systems, Computational Hydraulics International, Guelph, Ontario, CA, pp. 99–122.

Rose, L.S., Stallins, J.A. & Bentley, M. (2008). Concurrent cloud-to-ground lightning and precipitation enhancement in the Atlanta, Georgia (USA) urban region. *Earth Interactions*, 12, 1–30

Rosenfeld, D. (1999). TRMM observed first direct evidence of smoke from forest fires inhibiting rainfall. *Geophysical Research Letters*, 26, 3105–3108

Rosenfeld, D. (2000). Suppression of rain and snow by urban air pollution. *Science*, 287, 1793–1796

Rosenfeld, D., Dai, J., Yu, X., Yao, Z., Xu, X. Yang, X. & Du, C. (2007). Inverse relations between amounts of air pollution and orographic precipitation. *Science*, 315, 9

Rosenfeld D, Lohmann, U., Raga, G.B., O'Dowd, C.D. & Kulmala, M. (2008). Flood or drought: how do aerosols affect precipitation? *Science*, 321(5894), 1309–1313

Rozoff, C.M. & Cotton, W.R. (2003). Simulation of St. Louis Missouri, land use impacts on thunderstorms. *Journal of Applied Meteorology*, 42, 716–738

Shafir, H. & Alpert, P. (1990). On the urban orographic rainfall anomaly in Jerusalem – A numerical study. *Atmospheric Environment B: Urban Atmosphere 24B*, 3(1990), 365–375

Shem, W. & Shepherd, J.M. (2009). On the impact of urbanization on summertime thunderstorms in Atlanta: two numerical model case studies. *Atmospheric Research*, 92, 172–189.

Shepherd, J.M. (2005). A review of current investigations of urban-induced rainfall and recommendations for the future. *Earth Interactions*, 9, 1–27

Shepherd, J.M. (2006). Evidence of urban-induced precipitation variability in arid climate regimes. *Journal of Arid Environments*, 10.1016/j.jaridenv.2006.03.022.

Shepherd, J.M. & Burian, S.J. (2003). Detection of urban-Induced rainfall anomalies in a major coastal city. *Earth Interactions*, 7, 1–14

Shepherd, J.M., Carter, W.M., Manyin, M., Messen, D. & Burian, S. (2009). The impact of urbanization on current and future coastal convection: A case study for Houston. *Environment and Planning B*. (in press)

Shepherd, J.M., Pierce, H. & Negri, A.J. (2002). Rainfall modification by major urban areas: observations from spaceborne rain radar on the TRMM satellite. *Journal of Applied Meteorology*, 41, 689–701

Shepherd, J.M., Taylor, L. & Garza, C. (2004). A dynamic multi-criteria technique for siting the NASA-Clark Atlanta rain gauge network. *Journal of Atmospheric and Oceanic Technology*, 21, 1346–1363

Shreffler, J.H. (1978). Detection of centripetal heat-island circulations from tower data in St. Louis. *Bound.-Layer Meteorology*, 15, 229–242

Simpson, M.D. (2006). Role of urban land use on mesoscale circulations and precipitation. Dissertation. North Carolina State University, p. 294

Stallins, J.A., & Shea Rose, R. (2008). Urban lightning: current research, methods, and the geographical perspective. *Geography Compass*, 2(3), 620-639 doi:10.1111/j.17498198.2008.0-0110.x

Teerarojanarat, S., Fairbairn, D. & Chunithipaisn, S. (2004). Urban growth simulation with UrbanSim. Proceedings of *the FOSS/GRASS Users Conference*. Bangkok, Thailand. September 2004

Tewari., M., Chen, F., Wang, W., Dudhia, J., LeMone, M., Mitchell, K., Ek, M., Gayno, G. Wegiel, J. & Cuenca, R. (2004). Implementation and verification of the unified Noah land surface model in the WRF model. *20th Conference on Weather Analysis and Forecasting/16th Conference on Numerical Weather Prediction*. 11–15 January, 2004, Seattle, Washington.

Thielen, J., Wobrock, W., Gadian, A., Mestayer, P.G. & Creutin, J.D. (2000). The possible influence of urban surfaces on rainfall development: A sensitivity study in 2D in the meso-gamma scale. *Atmospheric Research*., 54, 15–39

Trenberth, K. (2005). Uncertainty in hurricanes and global warming. *Science*, 308, 1753–1754

Trenberth, K.E., Jones, P.D., Ambenje, P., Bojariu, R., Easterling, D., Klein Tank, A., Parker, D., Rahimzadeh, F., Renwick, J.A., Rusticucci, M., Soden, B. & Zhai, P. Contributing Authors: R. Adler (USA), L. Alexander (UK, Australia, Ireland),

H. Alexandersson (Sweden), R. Allan (UK), M.P. Baldwin (USA), M. Beniston (Switzerland), D. Bromwich (USA), I. Camilloni (Argentina), C. Cassou (France), D.R. Cayan (USA), E.K.M. Chang (USA), J. Christy (USA), A. Dai (USA), C. Deser (USA), N. Dotzek (Germany), J. Fasullo (USA), R. Fogt (USA), C. Folland (UK), P. Forster (UK), M. Free (USA), C. Frei (Switzerland), B. Gleason (USA), J. Grieser (Germany), P. Groisman (USA, Russian Federation), S. Gulev (Russian Federation), J. Hurrell (USA), M. Ishii (Japan), S. Josey (UK), P. Kållberg (ECMWF), J. Kennedy (UK), G. Kiladis (USA), R. Kripalani (India), K. Kunkel (USA), C.-Y. Lam (China), J. Lanzante (USA), J. Lawrimore (USA), D. Levinson (USA), B. Liepert (USA), G. Marshall (UK), C. Mears (USA), P. Mote (USA), H. Nakamura (Japan), N. Nicholls (Australia), J. Norris (USA), T. Oki (Japan), F.R. Robertson (USA), K. Rosenlof (USA), F.H. Semazzi (USA), D. Shea (USA), J.M. Shepherd (USA), T.G. Shepherd (Canada), S. Sherwood (USA), P. Siegmund (Netherlands), I. Simmonds (Australia), A. Simmons (ECMWF, UK), C. Thorncroft (USA, UK), P. Thorne (UK), S. Uppala (ECMWF), R. Vose (USA), B. Wang (USA), S. Warren (USA), R. Washington (UK, South Africa), M. Wheeler (Australia), B. Wielicki (USA), T. Wong (USA), D. Wuertz (USA), (2007). Observations: Surface and atmospheric climate change. In: Climate change 2007: The physical science basis. In S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor & H.L. Miller (Eds.), *Contribution of working group i to the fourth assessment report of the intergovernmental panel on climate change*. (London: Cambridge University Press)

Trusilova, K., Jung, M., Churkina, G., Karstens, U., Heimann, M. & Claussen, M. (2008). Urbanization impacts on the climate in Europe: Numerical experiments by the PSU–NCAR Mesoscale Model (MM5). *Journal of Applied Meteorology and Climatology*, 47, 1442–1455

UNFPA (2007). The state of world population 2007: United Nations Population Fund. *United Nations Publications*, p. 1

Unger, J. (2004). Intra-urban relationship between surface geometry and urban heat island: Review and new approach. *Climate Research*, 27, 253–264

van den Heever, S.C. & Cotton, W.R. (2007). Urban aerosol impacts on downwind convective storms. *Journal of Applied Meteorology and Climatology*, 46, 828–850

Vukovich, F. M. & Dunn, J.W. (1978). A theoretical study of the St. Louis heat island: Some parameter variations. *Journal of Applied Meteorology*, 17, 1585–1594

Waddell, P. (2002). UrbanSIM: modeling urban development for land use, transportation and environmental planning. *Journal of the American Planning Association*, 68(3), 297–314

Webster, P.J., Holland, G.J., Curry, J.A. & Chang, H.R. (2005). Changes in tropical cyclone number, duration, and intensity in a warming environment. *Science*, 309, 1844–1846

Westcott, N. (1995). Summertime cloud-to-ground lightning activity around major Midwestern urban areas. *Journal of Applied Meteorology*, 34(7), 1633–1642

Yang, X. & Lo, C.P. (2003). Modeling urban growth and landscape changes in the Atlanta metropolitan area. *International Journal of Geographical Science*, 17(5), 463–488

Yoshikado, H. (1994). Interaction of the sea breeze with urban heat islands of different sizes and locations. *Journal of Meteorological Society of Japan*, 72, 139–142

Zhang, C., Chen, F., Miao, S., Li, Q., Xia, X. & Xuan, C. (2009). Impacts of urban expansion and future green planting on summer precipitation in the Beijing metropolitan area. *Journal of Geophysical Research*, 114, D02116, doi:10.1029/2008JD010328

Zhang C., Miao, S., Li, C. & Chen, F. (2007). Incorporation of offline-resolution land use information of Beijing into numerical weather model and its assessing experiments on a summer severe rainfall. *Chinese Journal of Geophysics*, accepted

Zhang, H., Sato, N., Izumi, T., Hanaki, K. & Aramaki, T. (2008). Modified RAMS-Urban canopy model for heat island simulation in Chongqing, China. *Journal of Applied Meteorology Climatology*, 47, 509–552

# Chapter 15
# Interpolating a Consumption Variable for Scaling and Generalizing Potential Population Pressure on Urbanizing Natural Areas

**Dalia Varanka**

**Abstract** Measures of population pressure, referring in general to the stress upon the environment by human consumption of resources, are imperative for environmental sustainability studies and management. Development based on resource consumption is the predominant factor of population pressure. This paper presents a spatial model of population pressure by linking consumption associated with regional urbanism and ecosystem services. Maps representing relative geographic degree and extent of natural resource consumption and degree and extent of impacts on surrounding areas are new, and this research represents the theoretical research toward this goal. With development, such maps offer a visualization tool for planners of various services, amenities for people, and conservation planning for ecologist. Urbanization is commonly generalized by census numbers or impervious surface area. The potential geographical extent of urbanism encompasses the environmental resources of the surrounding region that sustain cities. This extent is interpolated using kriging of a variable based on population wealth data from the U.S. Census Bureau. When overlayed with land-use/land-cover data, the results indicate that the greatest estimates of population pressure fall within mixed forest areas. Mixed forest areas result from the spread of cedar woods in previously disturbed areas where further disturbance is then suppressed. Low density areas, such as suburbanization and abandoned farmland are characteristic of mixed forest areas.

**Keywords** Population pressure · Urbanization · Kriging

---

D. Varanka (✉)
Mid-Continent Geographic Science Center, United States Geological Survey, Rolla, MO 65401, USA
e-mail: dvaranka@usgs.gov

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_15,

## 15.1 Introduction

A city is generally defined as a relatively densely-inhabited settlement with higher order economies and administration that supplies and demands resources from a broader region of variable extent. The term urban refers to the characteristics of a city, and urbanization refers to the process of becoming urban or urbanized. Urbanization trends are characterized by population and demographic transitions and production/consumption patterns. Urbanization is commonly represented by geographical extents of either population aggregations or urban land use and land cover. These perspectives, however, do not allow for a more accurate accounting of urban resource consumption, its spatial extent, and its corresponding pressure on the environment, because urbanization involves factors operating at scales extending into the periphery of the city. One of these processes that extend population pressure into natural areas is urbanism; urbanism is a fluid concept used to mean characteristic ways of life experienced within a city (Wirth 1938). Conventionally, maps offer static representations of demographic or developed land surface at various cartographic scales, but resource consumption requires representation of an "operational scale" of physical processes and flows driven by urbanism (Lam et al. 2005). The operational scale would be delimited at points of maximum variability in the extent of the processes involved and the homogeneity in its smallest units of analysis (Lam and Quattrochi 1992). Methods for analyzing the operational scale of urbanism, and related environmental consumption, are not yet widely developed or used.

Places that look natural, with extensive biological amenities, are commonly regarded as less spoiled by legacies of manufacturing and industrialization and their by-products, as well as less restricted by codes of conduct. In fact, natural areas within certain proximities of urban areas may be degraded and support largely exotic and invasive species or other environmental impacts not easily recognized. Nevertheless, nature is considered to be an amenity, especially to suburban lifestyles (Mueller-Wille 1990). Natural areas offer an alternative to urban hardscapes; for example, in terms of aesthetic perception, as structures of sociological identities, or in the types of materials available for living. The reproduction of certain lifestyle practices is carried into the natural environment to serve human preferences; for example, food or physical comfort, hobbies such as photography, music, hunting, and fishing.

This paper addresses the need for improved spatial representations of urban relations to the environment in the United States by presenting a spatial trend surface of an interpolated variable. This variable is a proxy of human consumption of material resources based on per capita income and population density to represent population pressure on the environment. The objective of the spatial model is to frame and approximate the relative spatial patterns of potential environmental resource consumption because of urbanism. In this way, the scale of urbanization extends from population numbers and impervious surface land cover to include the region around cities. This extension would enable urban consumption to be increasingly acknowledged as a geographic factor when integrated with other environmental data.

The science needed to analyze scales of urbanism, and thus population pressure and its environmental burdens, must generalize geographical extents of cohesive resource consumption variables, such as water, air, food, and materials. The importance of improving procedures for determining geospatial scale and generalization of urban environmental consumption data for scientists and land managers is to facilitate new methods to estimate an urban ecological footprint, which is a measure of the terrestrial area needed to provide the resources demanded by a population for its standard of living (Lambin et al. 2001). Although many variables make up calculations of the urban ecological footprint, population and land use are common to all of them.

The following sections of this paper discuss urbanization, consumption, and geographical scale, the analytical approaches of interpolation of a consumption variable and its overlay with land-use/land-cover data, and ends with a discussion of the application results and of using the spatial model for these purposes.

## 15.2 Scales of Urbanization

Urban demographic transitions and production/consumption patterns are represented by varying scales of spatial extent. At its most compact geographical scale, urbanism is characterized by population and infrastructure concentration. These are represented by U.S. Census Bureau definitions of urban areas and clusters based on population density and by data derived from those definitions (U.S. Census Bureau 2005). Urban areas since the late twentieth century have been characterized by low-density occupancy, where the expanding land area classified as urban disproportionately exceeds the increasing urban population in a metropolitan area (Hobbs and Stoops 2002). Data calculated from U.S. Census Bureau enumeration are biased since counts are for places of residence and exclude places of daytime activity that expand the sphere of urban area. Although important consumption measures are based on household patterns captured by the census, population pressure is often more distributed across the land through networks of interaction, especially highways, where people go to places of employment, recreation, shopping, and services. These interactive networks enable the extension of urbanism to low-density areas. In the latter half of the twentieth century in the United States, the average density of a metropolitan area (including its suburbs), ranged from 299 to 407 people per square mile, which is much closer to the range of non-metropolitan densities (19–24 people per square mile) in the same period than to central city densities of 2,716–7,517 per square mile (Hobbs and Stoops 2002). Low population densities can still be recognized as urbanization for studies of resource consumption and levels of impact on natural resources.

A broader level of urbanization is measured by impervious surface and enclosed urban green space. Impervious surface measures capture many places of daytime activity that are non-residential. Similarly, cemeteries and recreational facilities such as urban parks and golf courses have natural cover, but have increased levels of urban land use compared to natural areas and are considered part of the city. These limited natural areas, however, support an ecology that is accustomed to harsher environmental circumstances, such as aggressive, generalist species, or plants and animals that are intensively supported by human beings. Beyond urban land cover, areas with natural cover and high population pressure, such as parks on the periphery of cities and basins from which cities draw resources, are urbanized land uses that are rarely characterized as urban except where cities have drawn boundaries of wide extent to accommodate future growth. At international scales, remittance landscapes are "… sites for the investment of [urban] remittances and …also a source for the capital outlay needed for overseas [and rural to urban] migration" (McKay 2005).

The physical, social, and economic needs of cities consume the resources of the larger urban area, sometimes called the urban ecological footprint. Increasing scales of urbanization have accompanying scales of environmental burdens. For example, at the scale of population aggregation, the most common risks concern public health and sanitation conditions. At the scale of the built environment, urban impervious surfaces create problems of auto emissions, water quality, and other environmental effects (Lo et al. 1997). At regional and global scales, systematic effects, such as global warming, disperse the environmental burden beyond the concentration of human-induced change. Though isolated examples of the relation of urban factors and their ecological extent are poorly understood, techniques of mapping the relation of urban extents and their interrelated ecological footprints are not well known.

Although not all consumption patterns are regional, the regional scale is important for understanding urban environmental challenges for a number of reasons (Marcotullio and McGranahan 2007). The global market network depends on local production, labor, and distribution. Though the markets may be global, the demand is local. To focus on the regional scale of urban consumption captures local development processes and their links to global systems by including local buying power.

As personal wealth and opportunities to purchase products increase, so does consumption. The rise of consumption in western societies is partially because of the growth of the diversity of products supported by globalization. Sociologists have noted that the value of products in post-industrial economies is no longer related as much to the use value, the value of the practical application of their materials, but to the association of identities with the image of the product (Jayne 2006). Associations of identity with images drive consumer demands for a complex array of products and services. Local and regional urban rejuvenation projects offer more products for consumption with greater selection by attracting businesses from around the world. Communities compete for global investment in the global urban network hierarchy by developing projects that promote their image and consumption opportunities. The prominent image and attraction potential of these developments requires a consumer society to sustain them and are targeted to a smaller, wealthier percentage of the general public, with little support for local cultural activity, resulting in a stratification of social groups and use of urban space (Zukin 2006). These globalized consumption forms occur less regularly than everyday consumption, and their effects are less evenly distributed across the city.

In the post-industrial private sector, wealth-based consumption is most likely to have increased by the proliferation of highly visible images, products, and lifestyles that form and support identity. Yet, everyday consumption needs such as education, public health, food, and energy uses are more level, persistent, and socially evenly distributed. Regional urban consumption is supported by everyday consumption. Global images increase retail purchases, but these are balanced by considerations of comfort and relaxation; factors such as familiarity of family relations, moralities of spending and waste, and reuse and recreation of used objects mediate consumption that is exotic and distinctive in visual impression. In the post-industrial/post-modern city, the second-hand economy, such as vintage, discounted, and appropriated goods are locally based. Such production ties back to local sourcing and supply chains (Crewe and Lowe 1995). Impacts of the local services can be economically and culturally successful and innovative compared to homogenized global production because they are strongly tied to local identity and not to globalized, hierarchical competition.

Differences in global and local cultural economies have different economic and material measures. Global consumption relies on quantifiable measures of material processes. Local consumption involves the roles that culture, environment, and symbolic processes play in consumption, in addition to material and capital processes, and these measures may exist only in qualitative measures, or not at all. The explanation of local retail location may be based in terms of city amenities; a unique local identity, social reputation, restricted growth capitalizing on lower rents, and historic/cultural associations of smaller neighborhoods. Regional consumption trends exist that will resist global homogenization (Ley 1986).

Qualitative aspects of urbanism, at multiple scales, primarily begin on the ground in relatively local settings. This will be true of changes in urban lifestyles for resource conservation and quality of life concerns as well (Kates and Torrie 1998). Cumulative environmental impacts at regional scales are counted in global change assessments as initiatives that are implemented at the regional scale. Governance set at global scales will require local and regional implementation for monitoring. Consumption is a two-way process, based on individual choice and constrained by broader economic and political forces. Despite that political policy, regulation, and governance play a part in shaping the spatial extent of consumption burdens, particularly by transferring risks to selective places, many important resource consumption demands cannot be geographically shifted because of prohibitive time and energy costs or consumer behavior. Major sources of ecological services, such as air, land, and water are drawn from within the region.

## 15.3 Approach for the Analysis

For the purposes of the model presented in this paper, population pressure refers to the consumption of useable environmental services at the regional scale. Ecosystem services include the benefits derived from air, land, ground water, land cover, and extracted industrial minerals, such as gravel and crushed rock for infrastructure (roads, utilities, and structures) (Daily 1997). These are used in varying ways and degrees depending on cultural characteristics of the city. In addition to the direct use of natural resources, natural resource consumption is also indirectly involved in global or local products available through retailing. Products of globalization are represented in the model by the interaction between the retailer and consumer at the point of purchase. Their inclusion represents the demand for manufactured and/or packaged materials that were extracted either within the region or bought elsewhere and used locally.

The approach for interpolating population pressure addresses two spatial assumptions; that people use natural resources as they move and act across the landscape, and that consumption of resources will be greater close to places of residence rather than farther away. Consumption decreases away from home, but extends through trips to work, for shopping, during recreation and leisure time, and for special services, such as health care and education. Movement across the landscape mostly is by automobile, so the road network is a supporting structure of the spatial depiction of population pressure, but the environmental burden of air and water pollution, material extractions, and waste sites are generally pushed to areas of lower human density, as continuous variables. The urban ecological footprint is based on economic and spatial activity, either locally or distantly connected by transportation that results in landscape fragmentation; the interstitial areas are reserved for recreation, local agriculture, and natural resource conservation. The effect of trade on environmental population pressure is represented in this study by income earned by local residents, who then purchase commodities, even if the actual resource environmental impact is outside the study area. The environmental burdens of imported resources consumed in the study area are only indirectly considered as outcomes of transport within the study area.

This study uses kriging to create a population trend surface of a consumption variable that represents these assumptions. Kriging is a method of interpolating attribute values for unknown point locations based on attributes for known point locations and includes spatial dependence. It also generates statistical values for accuracy and reliability of the unknown point locations. Kriging is an interpolation technique based on weighted values derived from regionalized variable theory. Regionalized variable theory is recognized as one of the most promising approaches to determining operational scale over geographical extent (Journal and Huijbregts 1978; Oliver et al. 1989). Regionalized variable theory lends insight into the spatial scale of environmental variables by plotting the variation of point sampling values within a geographical extent and calculating their spatial autocorrelation from the attributes of those sample points (Atkinson and Tate 2000). The plot of semi-variance (half the differences between all possible points spaced a constant distance apart) against distance among sample points, called the variogram, indicates the limits to the spatial autocorrelation and thus the scale of the environmental process under study (Burrough and McDonnell 1998).

Kriging assumes a continuous field of values; it is applied to categorical population data in this analysis because population pressure, affecting natural systems, is continuous data. That urban data are commonly regarded to be categorical in nature often is based on political or enumeration boundaries, or the generalization of certain structural variables, such as buildings or impervious surfaces. Kriging has been used for land-cover analysis (Lister et al. 2000) and can be used with discrete spatial objects, which are common in the urban context, by defining the characteristics of the neighborhood, making reasonable assumptions about geographic processes, and using supporting data (Haining 2003). The method has also been applied to urban issues; for example, Luo and Wei (2004) used variogram and kriging to model land value distribution in Milwaukee, Wisconsin. Wu and Murray (2005) estimated urban population density with cokriging, which used impervious surface fraction as a secondary variable to improve estimation accuracy. To categorize urbanism as an object fails to reflect that some aspects of urbanism resemble continuous data at certain scales, such as the aesthetic value of scenery, as do factors of urban habitat, such as climate or soil, or effects on those variables. Volumes of exchange, another defining characteristic of the flow of urbanism, can be represented as network data. These data forms can be combined when expanding urbanism beyond census numbers and impervious surface to new scales of accounting for the urban use of environmental resources. Though based on population data that normally are ratio-type data, population pressure can be considered to have aspects of network and continuous properties.

The successful application of variograms and geostatistical methods, supported by other geographic information science techniques, would link the scaling of urbanization and environmental problems to modeling and decision support. In addition to regionalized variable theory, geographical scale could be analyzed by employing hierarchical systems or multi-level approaches that rank the geographic instances that are generalized by modeling through uneven relations, such as greater than or less than. The simultaneous modeling of multiple factors at different levels and at selected sites sheds light on their interrelations. This approach can successfully analyze relations between scales, but it cannot be used for identifying or defining scales of extent (Easterling and Polsky 2004).

In the model presented in this paper, the aim is not to arrive at precise values, as much as to develop the logic of the model. Modeling simplifies complex and mediating dynamics and options with various facets for the purpose of grasping general trends. The conceptual simplifications behind the spatial model of population pressure on the environment are represented by averaging and interpolating data values over areas likely to be affected by multiple points of influence and network connections. Consumer preferences and consumption levels associated with specific social classes are biased toward specific places, but places have a multiplicity of meanings that coexist at the level of the individual and change with context (Crewe and Lowe 1995). Multiplicity of meanings place concepts of absolute location outside the grasp of quantification and the scope of this paper. Averaging is a simplification of the complexity of those associations across social classes.

After the trend surface was interpolated, it was compared with the land cover to draw correlations between interpolated population pressure and ground surface effects. The kriging surface is continuous, interpolated data. To compare the correspondence between the interpolated trend surface and categorical land-use data, the kriging surface was grouped into classes using the Jenks-Caspall Optimal Classification method (Jenks and Caspall 1971). As an optimal classification method, the Jenks method provides the most accurate intensity values for specific places. As a result, the classification breaks are not comparable with other places and would be inappropriate for comparative studies. The classification, however, was applied to the entire state, and cities within the study area (the state) are all classified the same way. Classification schemes with absolute values would also alter the size of these interpolated population levels.

### *15.3.1 Related Literature*

In some literature, increasing population pressure is linked to greater poverty and environmental degradation (Ehrlich 1968). Plentiful labor indicates greater human consumption, less investment, lower wages, and slower growth. Alternatively, population pressure has been attributed with the achievement of labor-intensive development and market complexity, higher land productivity and agricultural intensification, and technological innovation and development (Boserup 1965, 1981). Models of environmental burdens include the Environmental Kuznets Curve (Stern 2004), which hypothesizes that economic growth initially increases burdens, but after a certain point, continued growth eventually decreases burdens. The Urban Environmental Transition hypothesis (McGranahan et al. 2001) proposes that burdens follow the same scale as affluence - low income countries have local burdens that decline with continued growth, middle income countries produce regional burdens that initially rise then fall, and affluent countries produce global scale burdens that increase, but these are all modified by policy and environmental governance. Higher affluence produces more spatially displaced burdens because of policies, not economics. City-regional burdens, such as ambient air pollution, are determined to have a more ambiguous relation to affluence.

Research on population trend surfaces has estimated population using dasymetric mapping, where population numbers have been assigned to land-use class polygons (primarily those of residential and employment categories) or interpolated trend surfaces. Dasymetric mapping uses land use as the determining factor (Holt et al. 2004; Mennis 2003; Eicher and Brewer 2001; Langford and Unwin 1994). In contrast to dasymetric mapping, this study determined the interpolated population, and examined how it relates to land use. Bracken and Martin (1989) generated population surface models using the inverse distance weighting (IDW) method to interpolate point values and develop a population surface from census enumeration districts. Martin (1996) assessed surface and zonal models of population. Langford et al. (1991) interpreted population counts to raster cells. Though similar in their objectives, daytime population density databases and models focus more on temporal scale to precisely identify distributions at various time periods, than on the spatial scale of the urban environment.

The surface model presented in this paper applies to studies of potential environmental change in a descriptive way. Most urban environmental change studies use either the gradient approach or landscape fragmentation metrics. The gradient paradigm hypothesizes that land at a greater distance from the city center has less disturbed area (McMahon and Cuffney 2000). This method involves selecting sampling sites over the landscape. The fragmented landscape observation assumes that land-use changes with time form patches of decreasing size and discontinuous, disjointed land cover. Land-use zoning tends to fragment the urbanized landscape, though practices such as impact zoning and agricultural zoning may halt the expanding urban gradient. Disturbance indexes of urban expansion on certain environments take a number of various factors and externalities into account instead of focusing on discrete variables (van Beynen et al. 2007). A general surface resembles both of these approaches by representing spatial lag or distance decay, and population is assumed to represent a combination of impacts.

### *15.3.2 Study Area*

The study area is the state of Missouri. Missouri has two major cities, St. Louis in the eastern part and Kansas City in the western part of the state, as well as many urban centers located along the Interstate 70 (I-70) and Interstate 44 (I-44) corridors. I-70 roughly follows the course of the Missouri River across the center of the state; I-44 runs southwest from St. Louis to Tulsa, Oklahoma and beyond. I-70 connects Interstates 29 and 35 that run north from Kansas City, and Interstate 55, that runs south from St. Louis (Fig. 15.1). Naturalized areas of Missouri are predominantly small-scale agriculture, forests, grasslands, and recreation areas. All these land uses support low or medium levels of development; urbanism, income, and production are relatively stable.


**Fig. 15.1** Missouri urban/rural block groups and main roadways

The form of the cities reflects that consumption is not only the processing of materials, but also includes the choices involved in the selection and modification of urban spaces and land areas. In the industrial era, cities were planned for the delivery of material consumption. For this reason, the original sites of Missouri cities often reflect advantages of transportation or water. Collective social consumption needs were met to support labor and basic human survival. In the post-industrial era, the relation of economic activity to the physical environment is more chaotic. In contrast to collective consumption, consumer culture is a way of building identity, social class, and distraction away from job boredom and is more eclectic than collective consumption. In both the industrial and post-industrial periods of urban development, however, collective consumption is unequally distributed within the city, often according to social class and status.

### *15.3.3 Data Sources and Analysis*

The steps to create a kriged surface are depicted in Fig. 15.2. The selected census block groups had a designated urban population greater than 50%. Census block groups may be combinations of rural and urban designated blocks, but the defining characteristics of a city (such as population density) alone are not urbanism, the metric the surface aims to capture, but formative influences on the sociology of urbanism. A predominantly urban character of the group was ensured by selecting only those groups with a greater than 50% overall urban population. A wealth index was calculated by multiplying population by per

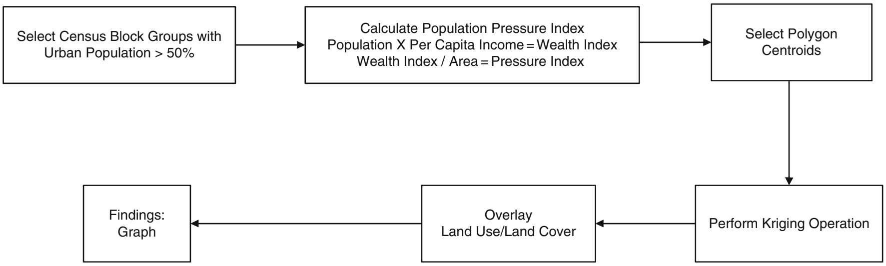



**Fig. 15.2** Analysis flowchart

capita income; dividing the wealth index by geographical area for the pressure index; selecting polygon centroids; performing the kriging operation; and overlaying the results with the land-use polygons. The advantage of using land-use data is that it provides a link to environmental economics and ecosystem services studies that are a way of quantifying the added value of natural resources that, together with capital and labor, make up consumption (Costanza et al. 1997).

Population pressure is usually measured as a ratio, such as population density or population per arable land unit. Efforts to refine the population pressure measurement might involve a measure relative to carrying capacity, or as it is modified by standards of living, social class, technology, and inter-ecosystem trade (Daly 1996). The IPAT equation, where Impact = Population × Affluence × Technology, is one way to quantify impact (Commoner 1972). In this study, the population numbers were weighted by income as an indicator of living standard and social class, as it would be in the IPAT equation, but to add a spatial component, the values are normalized by area rather than technology, because technology is assumed to stay constant. Consumer culture in post-modern, post-industrial cities blurs traditional social organization, based on class, gender, or race, but is sharply divided by access to consumption between the poor and the wealthy, or wealth (Jayne 2006). A consumption variable was derived from the weighted population data for census block groups by multiplying total population values by the per capita income. Population and income values were obtained from 2000 U.S. Census Bureau block group data, and block group polygons are derived from the 2004 U.S. Census Bureau Topologically Integrated Geographic Encoding and Referencing System (TIGER) line data. A set of data sample points for kriging was obtained by converting polygon values to point data for interpolation. Technological development and associated standard of living is assumed to be a constant during the period of this study, with few environmental conservation initiatives contributing a difference.

Unlike census tracts, which are categorized as urban or rural by urban areas and urban cluster definitions (U.S. Census Bureau 2002), the block groups were assigned an arbitrary urban or rural value by selecting block groups whose "percent" urban population value was greater than 50% (Fig. 15.1). Selecting individual urban tracts was an option, but the block groups have associated per capita income data.

The irregular pattern of population aggregate counts and enumeration units are difficult to integrate with other data types, but by interpolating a raster surface of the consumption variable from the centroid point that was assigned to each census unit, the consumption values were overlayed with the geospatial land-use data, and pixel sizes were adjusted to 30 m to match the data extracted from Landsat 7 in 2005. The edges of the kriging surface are anomalous because they do not entirely extend to the northwest or southeast boundaries of the state where insufficient data created an edge effect skewing the data. Inside the map of Missouri, however, the higher values of the surface appearing in white compare well visually with the population distribution evident in the input urban block group data (Fig. 15.3). The urban block groups are the colored polygons on the grey scale of interpolated population values. The pixels represent the lower limits of resolution for the analysis.

The surface that results from the kriging indicates some possible trends for closer examination.

1. Some areas of high urbanism may extend in a roughly concentric pattern from the outer limits of the city. St. Louis and Kansas City exhibit this trend.
2. Urbanism trend such as in Springfield exhibits some properties of the classic sector model of urban morphology, by which activities expand outward from the city center (Hoyt 1939). The highway network at Springfield may account for this pattern.
3. Cities along interstate highways, such as Columbia, on Interstate 70, and St. Robert, on Interstate 44, show elongated extents toward the larger

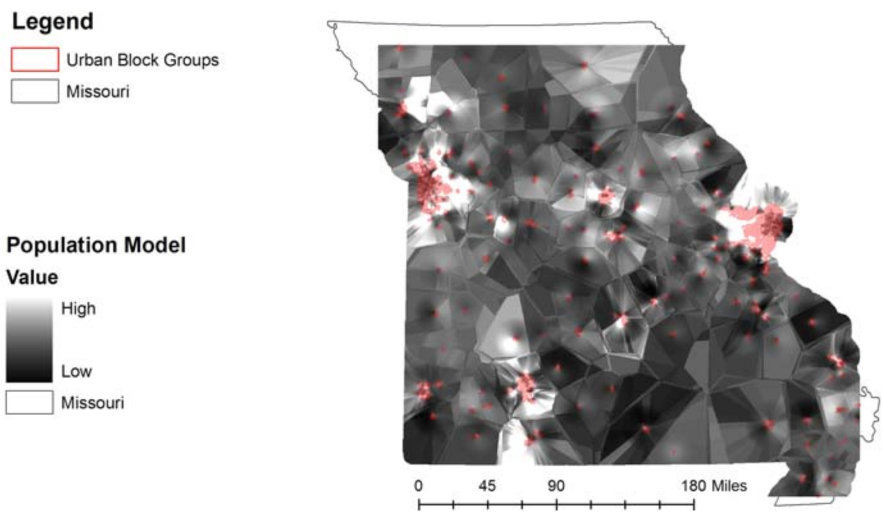



**Fig. 15.3** Kriging surface compared to urban census block groups

regional cities along the transportation corridor (St. Louis and Kansas City in the case of Columbia and Springfield in the case of St. Robert).

4. Some cities, such as Rolla, seem tightly contained with minimal expanded limits of urbanism.
5. Some cities have extended areas of high values with sharp limits, such as St. Joseph, and other cities have gradually declining values evenly spread away from them, such as Kirksville or Sikeston.

These results are related to the modifiable area of block groups and thus are skewed because of block group size.

Land-use data (Fig. 15.4) indicate that Missouri has predominant land-use types of deciduous forest in the southern Ozark region, grasslands in the

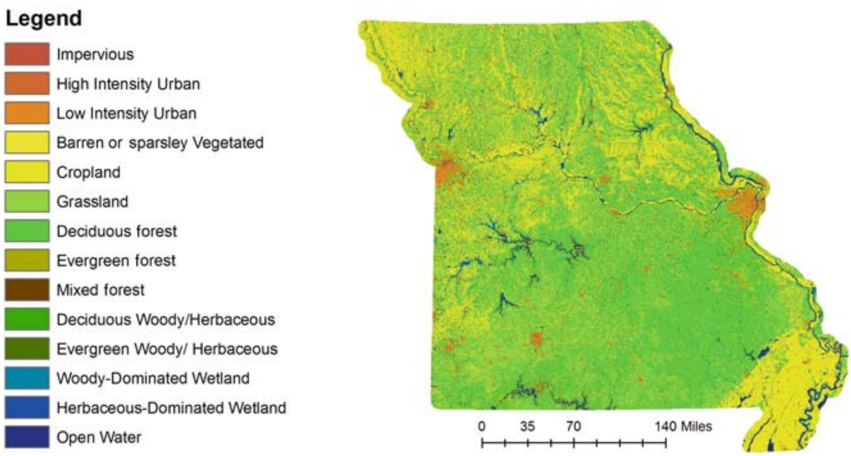



**Fig. 15.4** Missouri land use

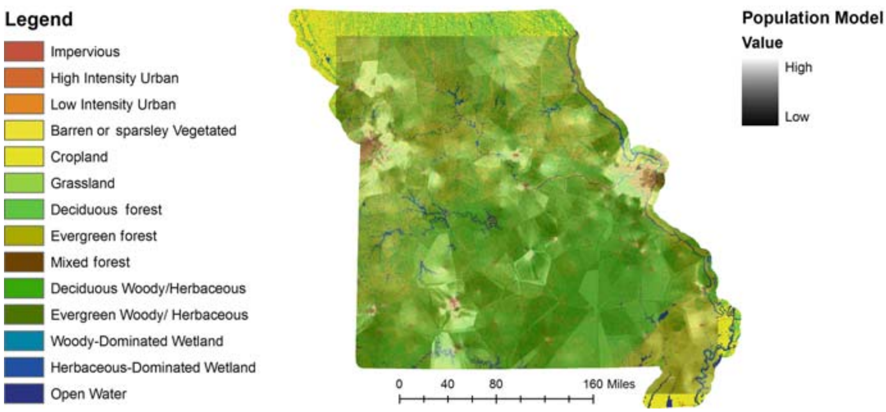



**Fig. 15.5** Kriging values overlayed with land use

glaciated north, cropland predominately in the southeastern “Bootheel” region, and low-density urban land use. The classification of this land-use/land-cover typology has several associated ecosystem values associated with them.

The Jenks-Caspall Optimal Classification method was used to reduce the data to five classes ranging from low to high. Most values align in the first category shown in dark blue (Fig. 15.6). The remaining trend surface values are shown as red and yellow as highest values, and medium and medium low values as green and light blue.

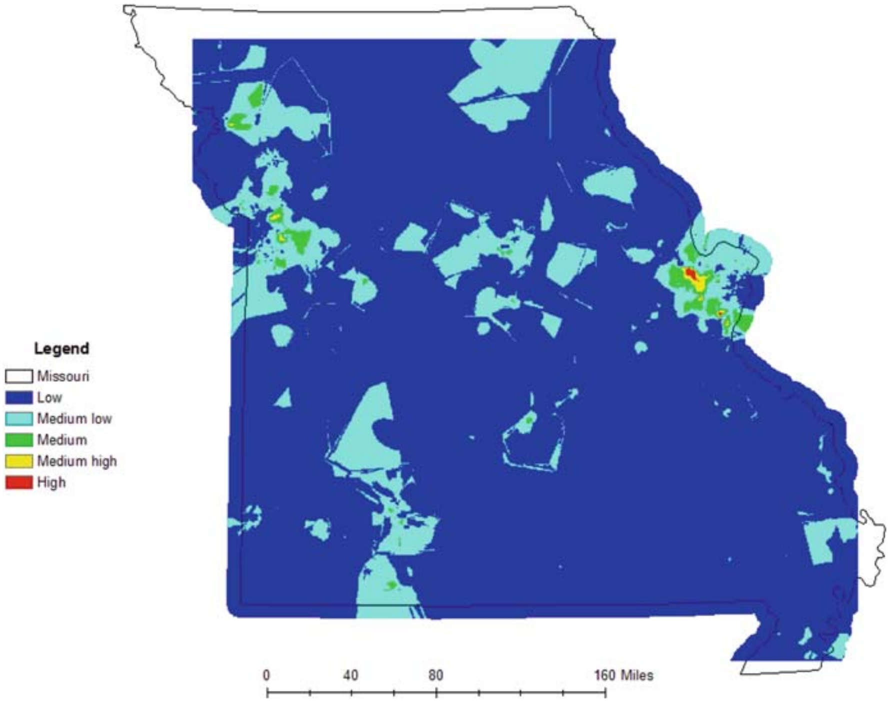
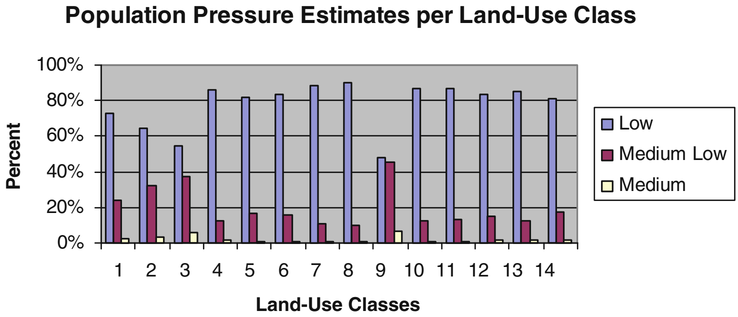



**Fig. 15.6** Levels of land-use pressure derived from grouped kriging values

## 15.4 Results

To explore correspondences between interpolated population pressure levels and land-use classes, an overlay process in the form of a cross-tabulation technique that combines all the attributes from one layer with all the attributes of a second layer and produces a union and intersection of the data sets. The overlay shows relatively low class levels of pressure across all land-use categories of the state, with higher levels in urban land-use classes (Classes 1, 2, and 3) (Fig. 15.7). The highest level of population pressure in the state falls in mixed forest areas (Class 9), though it is the smallest land-use class in the state.




| Class | Class Name | Percent | Class | Class Name | Percent |
|---|---|---|---|---|---|
| 1 | Impervious | 1.77 | 8 | Evergreen Forest | 1.13 |
| 2 | High Urban | 0.16 | 9 | Mixed Forest | 0.01 |
| 3 | Low Urban | 2.03 | 10 | Deciduous/Open | 3.43 |
| 4 | Barren or Sparse | 0.38 | 11 | Evergreen/Open | 0.04 |
| 5 | Cropland | 23.29 | 12 | Woody Wetland | 1.79 |
| 6 | Grassland | 32.53 | 13 | Shrub Wetland | 0.39 |
| 7 | Deciduous Forest | 30.53 | 14 | Open Water | 2.52 |

**Fig. 15.7** Population pressure estimates per land-use class and their area as percent of total

Mixed forest land-use classes in Missouri are the result of suppression of disturbance processes, especially fires, in previously disturbed areas (Nelson 2005). The increased levels of interpolated population pressure that correlate with mixed forests are explained by the general diffusion trend toward lower-density settlement, such as suburbanization with close proximity to wooded areas. Other possible reasons for this pattern are land that has been cleared for land speculation, or farms where agricultural practices have been abandoned. The population pressure includes the presence of aggressive species of plants, but the environmental quality of native habitats is compromised.

## 15.5 Conclusions

The application of the model indicates that interpolating a consumption variable overlayed with land-use data initially shows that regions that normally are considered natural experience low to medium population pressure levels as a result of the diffusion of human demands over the landscape. The model and the data represent the regional scale of urbanism. The model is meant to demonstrate the logic of population pressure only. Actual levels of pressure obtained through overlay associations are misleading because varying classification schemes exhibit problems of the modifiable areal unit problem (MAUP). The size and classification limits of the surface pixels alter the results of the spatial overlay. Relative correspondences, however, can be made through a careful evaluation of the data preparation.

Further research is needed on the spatial extent and environmental preferences of the movement of people and consequent pressure on the environment at the regional scale. New personal mobility technologies that allow the study of location and movement over time can surely advance the study of this problem. The model could be further refined using transportation network data and be validated by integrating the general regional model with other science data and models. Similar to the way that ecosystem service applications of land-use data require a typology for the data that indicates ecosystem service value, land-cover data for population pressure could be classified with values that are characteristic of urbanism and their corresponding types of consumption (Troy and Wilson 2006). Linking the practices of consumption and ecosystem services begins to quantify the balance of demand and supply that sustainable development aims to achieve. Though issues of population pressure are more complex than what is represented by a population pressure surface, the surface as a spatial model is helpful as a preliminary indicator of population pressure extent, if not misinterpreted.

**Acknowledgments** The urban block group data was produced by the Missouri Census Data Center. The land-use/land-cover data was produced by the Missouri Resource Assessment Partnership and made available by the Missouri Spatial Data Information Service.

## References

Atkinson, P.M. & Tate, M.J. (2000). Spatial scale problems and geostatistial solutions: a review. *Professional Geographer*, 52, 607–623

Boserup, E. (1965). *The conditions of agricultural growth: the economics of agrarian change under population pressure*. (New York: Aldine Press)

Boserup, E. (1981). *Population and technological change: a study of long-term trends*. (Chicago: University of Chicago Press)

Bracken, I. & Martin, D. (1989). The generation of spatial population distributions from census centroid data. *Environment and Planning A*, 21(4), 537–543

Burrough, P.A. & McDonnell, R.A. (1998). *Principles of geographical information systems*. (Oxford: Oxford University Press)

Commoner, B. (1972). The environmental cost of economic growth. In R.G. Ridker (Ed.). *Population, resources and the environment* (pp. 339–363). (Washington, DC: Government Printing Office)

Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Kimburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P. & van den Belt, M. (1997). The value of the world's ecosystem services and natural capital. *Nature*, 387, 253–260

Crewe, L. & Lowe, M. (1995). Gap on the map? Towards a geography of consumption and identity. *Environment and Planning A*, 27, 1877–1898

Daily, G.C. (1997). *Nature's services: societal dependence on natural ecosystems*. (Washington, DC: Island Press)

Daly, H.E. (1996). *Beyond growth: the economics of sustainable development*. (Boston: Beacon Press)

Easterling, W.E. & Polsky, C. (2004). Crossing the divide, linking global and local scales in human-environment systems. In E. Sheppard & R.B. McMaster (Eds.), *Scale and geographic inquiry – nature, society, and method* (pp. 66–85). (Malden, MA: Blackwell Publishing Ltd.)

Ehrlich, P. (1968). *The population bomb*. (New York: Ballantine Books)

Eicher, C. & Brewer, C. (2001). Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartography and Geographic Information Science*, 28, 125–138

Haining, R. (2003). *Spatial data analysis, theory and practice*. (Cambridge: Cambridge University Press)

Hobbs, F. & Stoops, N. (2002). Demographic trends in the 20th century, U.S. Census Bureau, Census 2000 Special Reports, Series CENSR-4. (Washington, DC: U.S. Government Printing Office)

Holt, J.B., Lo, C.P. & Hodler, T.W. (2004). Dasymetric estimation of population density and areal interpolation of census data. *Cartography and Geographic Information Science*, 31(2), 103–121

Hoyt, H. (1939). *The structure and growth of residential neighborhoods in American cities*. (Washington, DC: Federal Housing Administration)

Jayne, M. (2006). Cities and consumption. In M. Miles & J. Short (Eds.), *Routledge critical introductions to urbanism and the city* (244 p). (London: Routledge)

Jenks, G.F. & Caspall, F.C. (1971). Error on choroplethic maps: Definition, measurement, reduction. *Annals, Association of American Geographers*, 61(2), 217–244

Journal, A.G. & Huijbregts, C.J. (1978). *Mining geostatistics* (600 p). (New York: Academic Press)

Kates, R.W. & Torrie, R.D. (1998). Global change in local places. *Environment*, 40(2), 39–41

Lam, N., Catts, D., Quattrochi, D., Brown, D. & McMaster, R. (2005). Scale. In R.B. McMaster & E.L. Usery (Eds.), *A research agenda for geographic information science* (pp. 93–128). (Boca Raton, FL: CRC Press)

Lam, N. & Quattrochi, D.A. (1992). On the issues of scale, resolution, and fractal analysis. *Mapping Sciences: Professional Geographer*, 44(1), 88–98

Lambin, E.F., Turner, B.L., II, Geist, H.J., Agbola, S.B., Angelsen, A., Bruce, J.W., Coomes, O., Dirzo, R., Fischer, G., Folke, C., George, P.S., Homewood, K., Imbernon, J., Leemans, R., Li X., Moran, E.F., Mortimore, M., Ramakrishnan, P.S., Richards, J.F., Skånes, H., Steffen, W.L., Stone, G.D., Svedin, U., Veldkamp, T.A., Vogel, C. & Xu, J. (2001). The causes of land-use and land-cover change: moving beyond the myths. *Global Environmental Change*, 11, 261–269

Langford, M., Maguire, D.J. & Unwin, D.J. (1991). The areal interpolation problem: estimating population using remote sensing within a GIS framework. In I. Masser & M. Blackmore (Eds.), *Handling geographical information: methodology and potential applications* (337 p). (London: Longman)

Langford, M. & Unwin, D.J. (1994). Generating and mapping population density surfaces within a geographical information system. *The Cartographic Journal*, 31, 21–25

Ley, D. (1986). Alternative explanations for inner-city gentrification: a Canadian assessment. *Annals of the Association of American Geographers*, 76, 521–535

Lister, A., Riemann, R. & Hoppus, M. (2000). User of regression and geostatistical techniques to predict tree species distributions at regional scales: 4th International Conference on Integrating GIS and Environmental Modeling (GIS/EM4). Problems, Prospects and Research Needs, Banff, Alberta, Canada, Sept. 2–8, 2000

Lo, C.P., Quattrochi, D.A. & Luvall, J.C. (1997). Applications of high-resolution thermal infrared remote sensing and GIS to assess the urban heat island effect. *International Journal of Remote Sensing*, 18, 287–304

Luo, J. & Wei, Y.H.D. (2004). A geostatistical modeling of urban land values in Milwaukee, Wisconsin. *Geographic Information Sciences*, 10(1), 49–57

Marcotullio, P. & McGranahan, G. (Eds.) (2007). *Scaling urban environmental challenges: from local to global and back*. (London, Earthscan)

Martin, D. (1996). An assessment of surface and zonal models of population. *Geographical formation systems*, 10(8), 973–989
McGranahan, G., Jacobi, P., Songsore, J., Surjadi, C. & Kjellen, M. (2001). *Citizens at risk: from urban sanitation to sustainable cities*. (London: Earthscan)
McKay, D. (2005). Reading remittance landscapes: Female migration and agricultural transition in the Philippines. *Danish Journal of Geography*, 105(1), 89–99
McMahon, G. & Cuffney, T.F. (2000). Quantifying urban intensity in drainage basins for assessing stream ecological conditions. *Journal of the American Water Resources Association*, 36(1), 247–261
Mennis, J. (2003). Generating surface models of population using dasymetric mapping. *The Professional Geographer*, 55(1), 31–42
Mueller-Wille, C. (1990). Natural landscape amenities and suburban growth, Metropolitan Chicago, 1970–1980. *Geography Research Paper* (No. 230). (Chicago: The University of Chicago Press)
Nelson, P.W. (2005). *The terrestrial natural communities of Missouri* (Jefferson City, MO: Missouri Natural Areas Committee)
Oliver, M., Webster, R. & Gerrard, J. (1989). Geostatistics in physical geography, *Part I: Theory: Transactions-Institute of British Geographer*, 14, 259–269
Stern, D.I. (2004). The rise and fall of the environmental Kuznets curve. *World Development*, 32(8), 1419–1439
Troy, A. & Wilson, M.A. (2006). Mapping ecosystem services: practical challenges and opportunities in linking GIS and value transfer. *Ecological Economics*, 60(2), 435–449
U.S. Census Bureau (2002). Census 2000 Urban and Rural Classification, April 30, 2002: Washington D.C., [electronic version]. Retrieved from http://www.census.gov/geo/www/ua/ua-_2 k.html, accessed July 28, 2008.
U.S. Census Bureau (2005). Metropolitan Statistical Areas and Components, December 2005: Washington D.C., [electronic version]. Retrieved from http://www.census.gov/population/-www/estimates/aboutmetro.html, accessed July 28, 2008.
van Beynen, P., Feliciano, N., North, L. & Townsend, K. (2007). Application of a karst disturbance index in Hillsborough County, Florida. *Environmental Management*, 39(2), 261–277
Wirth, L. (1938). Urbanism as a way of life. *The American Journal of Sociology*, 44(1), 1–24
Wu, C. & Murray, A.T. (2005). A Cokriging method for estimating population density in urban areas. *Computers, Environment and Urban Systems*, 29, 558–579
Zukin, S. (2006). David Harvey on cities. In N. Castree & D. Gregory (Eds.), *David Harvey, a critical reader* (pp. 102–120). (Malden, MA: Blackwell Publishing)

# Chapter 16
# Modeling Cities as Spatio-Temporal Places

**Xiaobai Yao**

**Abstract** A place exists in a space (footprint) which evolves over a temporal extent (lifeline). This study focuses on those places such as cities or component places within cities (hereafter referred to as city places or simply places). It concerns the modeling and analysis of places from the integrative spatio-temporal perspective. Drawing on recent research development related to spatio-temporal ontology, this study discusses a spatio-temporal ontology specific to city places. The ontology distinguishes between the static view of places (Static-Place) and the dynamic view of spatial-temporal regions led by spatio-temporal processes (ST-Place). The study considers places as spatio-temporal constructs. It proposes the concept of primitive ST region for the construction of spatio-temporal regions and ultimately spatio-temporal places. The study then presents a spatio-temporal data model to link spatial features and the spatio-temporal processes. In this way, the study models an ST-place from a spatio-temporal perspective by forming high dimensional spatio-temporal regions. Based on this conceptual data model, the study further discusses the potential use of this model in spatio-temporal analysis in general and in reasoning spatio-temporal topological relations in particular.

**Keywords** Place · Spatio-temporal modeling · Spatio-temporal region · Ontology · Spatio-temporal topology

## 16.1 Introduction

Place is a fundamentally important geographical concept while it carries significant meanings in almost every field of human inquiry and activity. It is the holder of activities and information regarding various aspects of human lives, as nothing we do is unplaced (Casey 1997). A place-centered information

X. Yao (✉)
Department of Geography, University of Georgia, Athens, GA 30605, USA
e-mail: xyao@uga.edu

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_16,

system will thus be extremely useful in our understanding of our environment and human activities. Georeferencing by place names is the most common form of referencing geographical features either in geographical information systems (GIS) or cognitively in people's minds (Hill 2006). The past couple of decades have witnessed the revolutionary shift of geographic data collection and dissemination from a limited number of centralized databases to a plethora of distributed open source documentations containing place information. A typical example is the large volume of web-based documents. In addition to existing spatial databases, huge amounts of place-related data are now available in various forms including texts, tables, as well as non-spatial databases. Many agencies and individuals are providing or using geographic information through various Internet GIS applications. To take advantage of these rich data sources and to exploit the great potential hidden in them, it is important to make a suitable modeling strategy of places available for the development of a place-based information system. This modeling strategy should allow the recording and managing of the spatio-temporal linkages and relationships among places in databases. Such a place-based information system can be particularly useful to decision makers and planners for intelligence, cultural, engineering, socio-economic, as well as education applications, to name a few examples. Researchers have been making efforts to better understand the concept of place and to find ways of incorporating them in GIS (Jordan et al. 1998; Agarwal 2005). These efforts can be furthered by modeling the innate space-time interconnections of places.

From a spatio-temporal perspective, information about places can be found in various types of documents but typically in gazetteers and maps. Maps show locations of selected places while a gazetteer is the traditional form of a place-centered database, which can be considered a dictionary of places. Gazetteers were originally in printed copies before the Internet era. Modern gazetteers have quickly taken advantage of information technologies and become increasingly available as digital systems available on the World Wide Web. Examples include the Alexandria Digital Library Gazetteer (ADL 2007), the Geographic Names Databases by the National Geospatial-Intelligence Agency (GNS 2009), Geographic Names Information Systems (GNIS) by the United States Geological Survey and the United States Board on Geographic Names (GNIS 2009), the Gazetteer for Scotland (2009), and many others. Many of the gazetteers are considered a type of system that organizes a variety of descriptive information of the named places. The unique strength of contemporary gazetteers is the knowledge organization tied by place names. However, the spatial analysis and modeling capabilities are quite limited in gazetteers. First, the geospatial location information in most gazetteers is usually limited to a pair of longitude/latitude coordinates for a place such as a county or a city. This leaves out necessary details of spatial boundaries of places, whereas such details are important to spatial analysis of relationships between places. Secondly, gazetteers are generally kept as non-spatial databases and thus do not provide spatial analysis capabilities on their own. In response to these limitations, efforts are being made to couple gazetteers with GIS technologies to take advantage of the advanced spatial data management and handling capabilities. For instance, footprints of places can be recorded or geocoded in GIS with higher spatial granularity. ADL (ADL 2007) is such a digital gazetteer service with participating GIS components.

A more thorny challenge is modeling place from the integrative spatio-temporal perspective. Many researchers in a variety of discourses have contested that place is not a static concept, but rather a continuous whole of past, present, and the future associated with it (Tuan 1975; Relph 1976). Parkes and Thrift (1978) maintain that the realization of place is through the temporal structuring of space. Many recent studies also consistently suggest that such realization is through a series of events or occurrences. For instance, Tuan (2001) conceptualizes place as a function of time. From the spatio-temporal perspective, place is regarded as not only the geographic space and associated descriptive properties of the place, but also the changes of the space and the properties over time. In a geometric sense, a place exists as a volume in the high-dimensional space-time vector space. The footprint we usually see in a gazetteer or GIS is just an approximation of the cross-section of the volume at any given time. Thus, the contemporary place-centered systems cannot model dynamic processes as integral parts of places. Instead, these systems accommodate the snapshot view of places. This makes the current place-centered systems very limiting in regard to modeling and analysis of places. Against this backdrop, the study is motivated to advance our understanding and modeling of place as a dynamic spatio-temporal construct. In the light of ontological underpinnings, the study attempts to present a spatio-temporal place model that supports integrative spatio-temporal analysis of places.

Although most of the discussions will be generally applicable to all places, this study concerns a specific type of places – the so-called "city places" which, in this chapter, generally refers to city, towns or city-like settlements, as well as component places inside a city such as landmarks, districts, and others. A city is a relatively large and permanent settlement, usually with certain administrative, legal, or historical status (Goodall 1987). Cities are typically places where much denser population distributions are seen and much more intensive human activities take place. According to the United Nations Population Fund (UNFPA), the world's urban population rapidly grew from 220 million to 2.8 billion over the 20th century and the next few decades will see an unprecedented scale of urban growth in the developing world (UNFPA 2007). The UNFPA also predicted that in 2008, for the first time in history, more than half of the world population would be living in cities. Indeed, history has witnessed enormous spatial transformations of cities along with their socio-economic growth. With the growth, many cities have expanded and encroached onto neighboring lands. In addition to gradual changes, some city places may also experience significant changes due to, for instance, sudden political or natural incidences. As a result, spatial territories of a city and spatial relations among city places have been typically evolving over time, along with the attribute changes inside urban spaces. Tracing the lifelines of cities and investigating the temporal changes of spatial relationships among these places will be instrumental to the understanding of these places as spatio-temporal constructs. In the remainder of the chapter, considerations will be specifically made for city places and examples will be taken from cities.

The rest of the chapter is organized as follows. The next section reviews prior studies of spatio-temporal modeling of geographical information. What follows is a section on the discussion of spatio-temporal ontology. In the light of the spatio-temporal ontology, Section 16.4 proposes the concept of spatio-temporal region, explains the construction of it, and develops a spatio-temporal place model using this concept. Section 16.5 discusses the potential of this proposed concept and place model for spatio-temporal analysis and modeling, particularly the reasoning of spatio-temporal topology between places. The concluding section discusses limitations and topics for future research.

## 16.2 Time and Spatio-Temporal Modeling

Time is a unique dimension in the spatio-temporal world. People have various ways of conceptualizing it. Two complementary views of time, namely Newton's absolute view and Leibniz's relative view, are discussed in the literature (Peuquet 1994). With regard to the direction of time, it can be linear, cyclical, or branching depending on the viewpoint and context of using/recording time (Worboys 1998; Gould 1987). For example, historic records using world time is an example of linear time and employs the absolute view. The seasons and days/nights are examples of the cyclic direction and are measured using the relative view of time. Similar to spatial resolution, time can also be measured at different levels of granularity. Peuquet (1994) maintains that time is essentially continuous but can be broken into discrete units of coarser granularity if the modeled phenomena evolves in steps instead of continuously.

Most contemporary geospatial information systems apply the static view of space-time. This approach records space for each of the selected timestamps. The series of datasets thus implicitly represent the spatial conditions in time series. The seminal work of integrative spatio-temporal modeling is the space-time prism proposed by Hägerstrand (1970). The work proposed a 3-Dimensional (3-D) space with 2-D for spatial locations and 1-D for time. The 2-D space continuously evolves along the time axis and thus a 3-D prism is formed. The idea was only conceptual when it was proposed due to computational limitations at the time. However, it has been influential to many later implemented approaches. Langran (1992) is one of the first who drew on Hägerstrand's work and proposed several space-temporal models. Among other innovative efforts, there are two general groups of approaches that are consistent with the object and field views of the geographical world respectively. The first group models identifiable geographical objects over space and time. These objects keep their identity while progressing spatially and/or thematically over time. Examples of these include, but are not limited to, object-based (Wachowicz 1999), identity-based (Hornsby and Egenhofer 2000), and event-oriented (Worboys 2005) approaches. The other group opts to be more consistent with the field-based view of space. In this later group of approaches, the locations (cells) do not change over time, but their status (or attribute) changes over time driven by events or processes. Examples of this group include event-based (Peuquet and Duan 1995) and process-based (Yuan 1997) models. Particularly in studies of human travel activities, many efforts have been made to address the space-time interplay in analyzing accessibility (Kwan 1998; Miller 1999; Miller and Wu 2000), activity and mobility (Wang and Cheng 2001; Yu and Shaw 2008), and visualization (Kwan 2004). The space-time view has also been addressed in data mining and other geocomputation research (Hamilton et al. 2006; Worboys and Duckham 2006; Compieta et al. 2007; Neutens et al. 2007; Yu and Shaw 2008).

## 16.3 Spatio-Temporal Ontology for Places

Ontology of a specific domain is a logical theory which gives an explicit, partial account of a conceptualization in the domain (Guarino and Giaretta 1995). In the geospatial domain, an ontology produces an account of concepts, categories, processes, and relations between them (Mark et al. 2000). The ontology of space-time deals with the nature of space and time and the interactions between them (Peuquet 2001). As discussed in previous sections, place is a complex spatio-temporal construct. In order to develop a viable common conceptual framework for modeling and analyzing places, it is essential to have a formal spatio-temporal ontology as the foundation for further modeling.

Many researchers have made a lot of efforts on the formalization of spatio-temporal ontology. These previous works recognized or emphasized interactions among space, time, and processes (including events, occurrences, etc.). After surveying the variety of kinds of phenomena within the realm of geography and GIScience, Galton (2003) argues that a fully temporal geo-ontology is urgently needed. The author surveyed literature on the conceptual views of space and expanded on the very interesting temporal analogy of the objects-based and fields-based spatial views. The temporal analogues of objects are processes of various kinds while the temporal fields are mappings from time to values. Galton (2003) further suggests that the concept of location probably needs to be expanded from a purely spatial or a purely temporal sense to the spatial-temporal sense. In another effort, Grenon and Smith (2004)'s work addresses the formal ontology of spatio-temporal beings. The work presents two main kinds of ontologies (SNAP and SPAN), each for one type of existence in the spatio-temporal world. The SNAP ontology is for the type of entities that "have continuous existence and a capacity to endure though time even while undergoing different sorts of changes" (Grenon and Smith 2004, p. 139). Examples include the earth, a building, or a city. Such an entity exists with its identity through its duration. A city is the same city as it was yesterday and probably tomorrow although it is going through some seen or unseen changes. A SPAN ontology is for the second type of existence, which refers to processes, events, activities, and changes. Examples include the landing of an aircraft, earthquakes, and so on. The major difference between the two types of entities lies in their distinct ways of existence in time. A SNAP ontology is formed through the depiction of entities at a given time. Thus it conceptualizes a static, snapshot view of the spatio-temporal world. Interestingly, this snapshot view is shared by the current commercial GISs. SPAN entities, however, exist only in their successive temporal parts or phases. The SPAN ontology conceptualizes entities through changes, processes, and events over time.

Starting from the previous pioneer research on spatio-temporal ontology, this study focuses on the spatio-temporal ontology for city places. Drawing on the SNAP–SPAN divide of the spatio-temporal ontology, this study defines the Static-Place and ST-Place (spatio-temporal-place). The Static-Place takes a static view of a place at any time instance. Following the Static-Place ontology, a city has snapshot recordings of the city's space and attributes at specified time points. In an ST-Place ontology, however, the city has a continuous lifeline. For instance, the place of Atlanta in the United States has evolved from a small rail terminus some 150 years ago to a major metropolitan city through a series of spatio-temporal processes. Over time, the city has maintained its identity as a place but has expanded spatially and changed its roles and properties. Therefore, the place identified as Atlanta refers to the existence in the geographical and historic context.

Figure 16.1 is a schematic representation of Static-Place components. Figure 16.2 is a schematic representation of ST-Place. In both figures, an

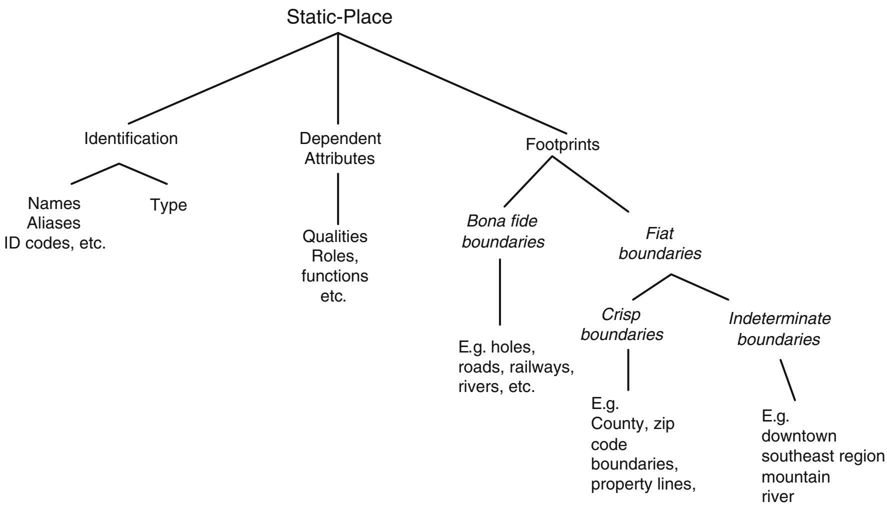



**Fig. 16.1** A schematic representation of the constituents of static-place

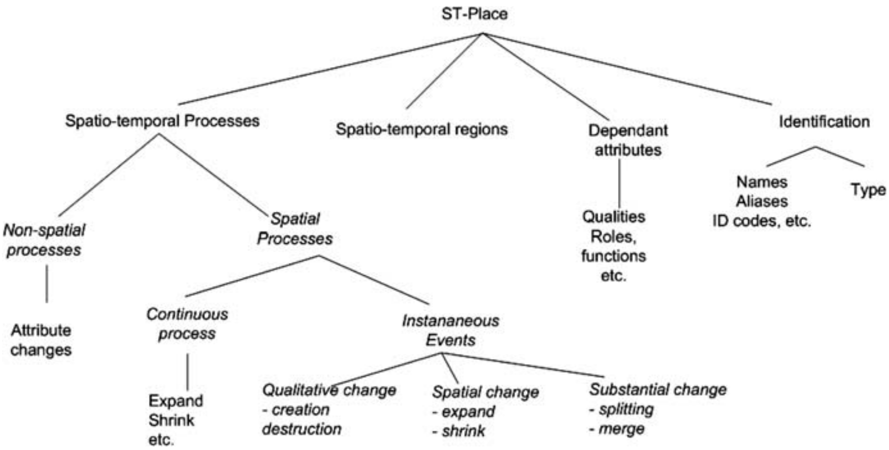



**Fig. 16.2** A schematic representation of constituents of ST-place

item in regular font means it is a component of its parent node concept. Italicized items are categories of the parent node concept.

In the ontology of ST-place, the spatio-temporal region (ST region) is a key component. The concept was first seen in Grenon and Smith (2004), yet no specific definition is available. In this study, ST region is specified as a high-dimensional volume, analogous to the idea of the space-time prism.

**Definition 1. ST region**: a 3-dimensional (3D) region in a XY-Time space. The X-Y plane georeferences the footprint of an ST-place at any given time. The footprints progress along the time dimension to form the 3D region.

## 16.4 A Spatio-Temporal Model for Places

### *16.4.1 Some Modeling Issues*

A data model is a representation of reality in a computer system (Goodchild 1987). A representation is not the reality itself but rather a simplified or generalized abstraction of reality. After such abstraction, something about the reality will be missing or somewhat twisted. A good data model suits the particular nature of the phenomena which it intends to represent. For instance, the vector model, which corresponds to the objects view of the geographic world, is better suited for geographic phenomena where spatial features with clear boundaries can be identified. The raster model, corresponding to the fields view, is more suitable for those phenomena with continuous distributions such as that of elevation and precipitation. In searching for a data model that is suitable for spatio-temporal modeling and reasoning of places, it is necessary to consider particularly the nature of place.

#### 16.4.1.1 Vagueness and Identity

The concept of place is inherently vague. Agarwal (2005) analyzes that the vagueness comes from a lack of understanding the linkages between the cognitive semantics and the world, and from the ambiguous ontological distinctions between place and other similar spatial concepts. Indeed, the vagueness presents itself in many different ways, particularly in the identity of a place.

One source of vagueness is the semantic ambiguity. The same name can be used for different types of places or sometimes even for different places of the same type. For example, the place name "Santa Barbara" is the name of a county as well as the name of a city in the same area in California, USA. At least three cities named "London" exist on this planet. This type of ambiguity is relatively easy to deal with, for instance, by proper geocoding and specifying the type of place. Another source is the spatial vagueness, which is trickier to deal with. In ancient China, cities were circumvallated and so it was easy to delineate the boundary of a city. However, in the majority of the cases, even though a city's limit exists in a legal or technical sense, there is nothing physical (like the city wall) to block interactions across the city proper. Thus such a city limit, if any, has little impact on city growth and human behavior. As a result, it is often simply not possible to clearly delineate the boundary of a city.

The most sophisticated source of vagueness is the conceptual ambiguity of place identities. As a simple example, the lifeline of a place starts with the appearance of the place with a certain identity and ends with the disappearance of this identity. However, as Galton (2003) pointed out, a spatio-temporal construct is not just a spatial object with a history. Instead, the four-dimensional view provides a wider variety of portions of space-time. By what criteria can we determine whether two portions in space-time belong to the same place identity? Should we use the place name, or place location? Neither is proven a good solver, according to the example below. To illustrate the problem, I use Beijing as an example. Beijing has been called over twenty different names in its three-thousand-year history. In Fig. 16.3(a), the background patches indicate

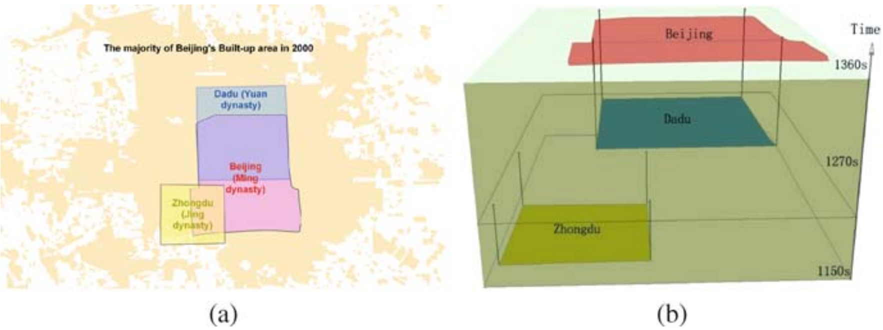



**Fig. 16.3** Identity of places: the case of Beijing (**a**) A plan metric view of the city locations, (**b**) a 3D view showing temporal durations of the cities

the majority of built-up areas of Beijing as in 2002. The three polygons in the middle of the figure indicate locations of three cities named Zhongdu, Dadu, and Beijing; each was the capital city at the time. Fig. 16.3(b) shows the location and duration of each city in history. Should we consider these cities sharing the same identity (Beijing) while presenting three different portions in the lifeline of Beijing? Maybe we would. However, they cannot be of the same identity if we use either city name or location. This is because the three city names are all different and two cities (Zhongdu and Dadu) are completely disjoint in space.

#### 16.4.1.2 Granularity of Space, Time, and Processes

Granularity refers to the fineness of recorded details. Spatial granularity is represented by spatial resolution in the field-view and level of aggregation in the object view. In the time dimension, granularity refers to the frequency of recording. In a precipitation observation system, for instance, it may be a day. In a place-based system over a long period of history, how, the temporal granularity will be much coarser such as a year or multiple years. Process granularity here refers to the finest size of process that is recorded in a system. The size of a process is measured by the changes the process brings about. Examples of processes include redistribution of spatial boundaries, expansion of a city (e.g. suburbanization), and so on. If the changes resulting from a process are smaller than the pre-defined granularity, it will not be recorded in a city place system.

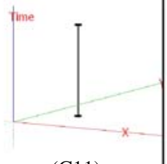
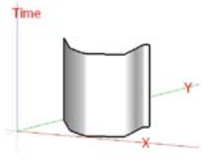
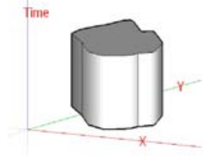
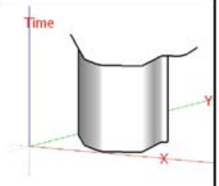
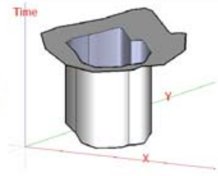
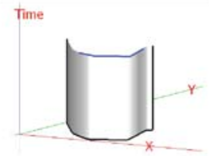
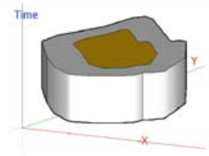
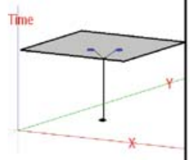
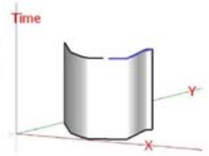
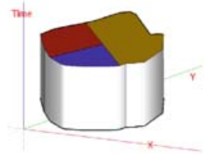
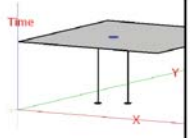
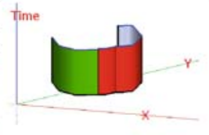
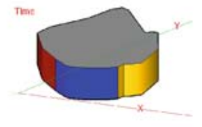
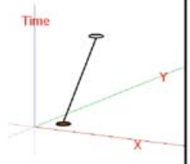
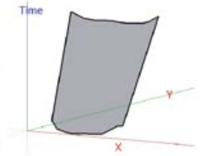
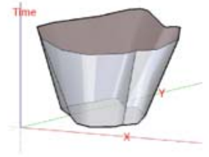
### *16.4.2 Spatio-Temporal Place Model*

Following the spatio-temporal ontology for places, the proposed spatio-temporal place model needs to be able to encode the identity, attributes, ST regions, processes and the interconnections among them. The definition of the ST region in Section 16.3 suggests that spatio-temporal processes drive the formation of an ST region. However, this definition did not specify how many processes may exist and in which way these processes mediate between space and time. Thus, the proposed spatio-temporal model employs the primitive ST region as the basic building block.

**Definition 2. Primitive ST Region:** an ST region in which only one process occurred during the entire time interval of the ST region. The process could be either instantaneous or continuous.

The construction of a primitive ST region is relatively simple as will be further discussed and illustrated in Table 16.1. An ST region can be a composite of multiple consecutive primitive ST regions. It is possible to model a complex ST region by linking multiple primitive ST regions, which is analogous to the way a curve is formed in a coordinate system through adding vertexes at every turning point.

**Table 16.1** Constructing primitive ST regions by type of process

| **Type of process** | | **Point** | **Polyline** | **Polygon** |
|---|---|---|---|---|
| Attribute(s) change | |  (C11) |  (C12) |  (C13) |
| Instantaneous Spatial Events | **Creation** | New Point | New Polyline | New Polygon |
| | **Destruction** | Same as C11, C12, and C13 respectively, except that the cross section of the ST region at $t_1$ is empty. | | |
| | **Expand** | Either C11 or a special case of C23 depending on the spatial granularity |  (C22) |  (C23) |
| | **Shrink** | Same as (C11) |  (C32) |  (C33) |
| | **Splitting** |  (C41) |  C42 |  (C43) |
| | **Merging** |  (C51) |  (C52) |  (C53) |
| **Continuous spatial change** | |  (C61) |  (C62) |  (C63) |

Suppose we need to construct a primitive ST region resulting from an ST process which spans between $t_0$ and $t_1$. The 3D shape of the primitive ST region is determined by the type of the ST process and the type of geometry of the footprints at $t_0$ and $t_1$. ST region can be constructed by fitting the volume between the two sets of footprints. Table 16.1 is a cross-tabulation of primitive ST regions by the ST process type and the geometry of the place footprints. There are generally three major types of ST processes: non-spatial attribute changes, continuous spatial processes, and instantaneous spatial events. There are also three types of basic geometries in GIS, point, polyline, and polygon.

*Non-spatial attribute change:* Non-spatial attribute changes refer to the changes of associated properties but not the space of a place. Because no spatial changes occur, there is no need for extra computation. The space will remain the same as the footprints at time $t_{0.}$

*Instantaneous Spatial Changes:* Instantaneous changes take effect at the event time ($t_1$). For illustration, let us use $t_0$ to denote the event time of the preceding ST process (if there is one) or the starting time of the place (if there is no previous event). Because the footprint of the place does not change until $t_1$ since $t_0$, the ST region is similar to that of non-spatial attribute change before time $t_1$.

*Continuous Changes:* Some events, such as expand and shrink, can take place and progress gradually during the time interval. This type of process is called a continuous change. Even if the footprints at $t_0$ *and* $t_1$ are the same as that of instantaneous change, the ST region will still be different. The volume can be constructed with certain spatio-temporal surface fitting or spatio-temporal interpolation techniques. The surface fitting and interpolation in spatio-temporal space is an interesting topic on its own and will not be discussed in this chapter.

ST-Place model considers a place being composed of a series of spatio-temporally referenced primitive ST regions. It is represented in Equation (16.1) as follows

$$\mathrm{B} = \{<p_1, p_2, ..., p_n> | p_1, p_{2,...,}p_n \in \mathrm{PST}\} \qquad (16.1)$$

where B is a place in the ST-Place model, PST is the collection of primitive ST regions. A primitive ST region can be expressed as

$$\mathrm{p} = \{< t_0, t_1, s_0, a_0, a_1 e, > | \mathrm{t_0, t_1} \in \mathrm{T},\ s_0, s_1 \in \mathrm{S}\ e \in \mathrm{E},\ \mathrm{a} \in \mathrm{A}\} \qquad (16.2)$$

where p is a primitive ST region, S refers to regions of space, T is the collection of time intervals, E denotes events and processes that drive the changes of the place, and A is the collection of attributes associated with the place. $t_0$ and $t_1$ are the starting and ending time points of the primitive ST region. $s_0$ and $s_1$ are the respective space of the place at these two timestamps. It should be noted that elements in S, T, E, and A are interconnected. Similar to the space S, A is a multiple-dimensional array of vectors with one dimension be the time dimension. Driven by *e*, elements in *S* and A change over T, following the way as illustrated in Table 16.1. In a special case, when $t_0 = t_1$, *e* will not be applicable as it exists only in time successions. Therefore at any given point of time, ST-Place is reduced to Static-Place which could be expressed as $\{<s, a>|s \in S,\ a \in A\}$. In other words, the Static-Place is a special case of ST-Place. Equation (16.2) is equivalent to Equation (16.3), which basically means a primitive ST region consists of two instances of Static-Place and a process.

$$p = \{<sp_0, sp_1, e, >|sp_0, sp_1 \in SP, e \in E\} \quad (16.3)$$

where SP is the collection of Static-Place and other notations are the same as previously defined.

For implementation of the ideas in geographic information systems, Fig. 16.4 presents a conceptual architecture. Four types of object classes are defined, namely the ST region, Primitive ST-Region, spatio-temporal (ST) processes, and Static-Place. Contemporary GIS software is consistent with the Static-Place ontology, thus the Static-Place entities can be easily managed by popular GIS software. The Static-Places at different time instances become components of the primitive ST-regions.

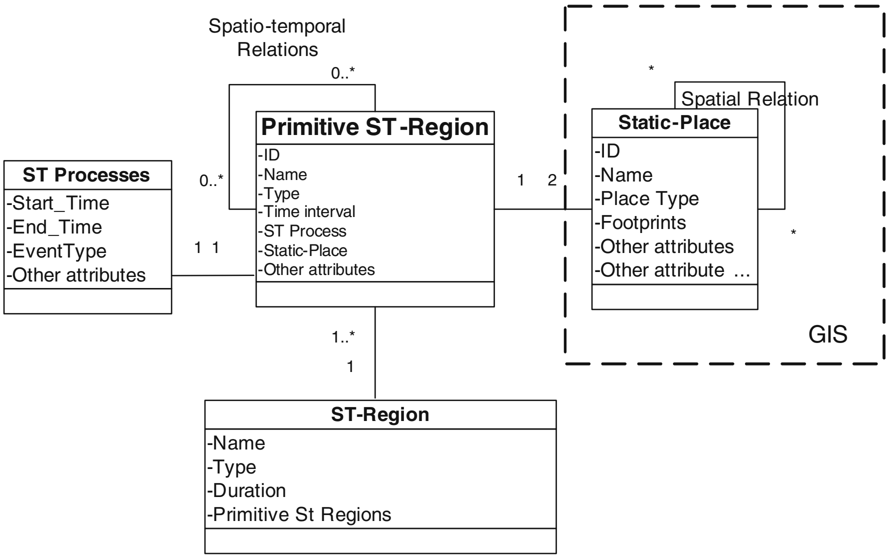



**Fig. 16.4** Conceptual framework of the spatio-temporal place model

## 16.5 Analyzing Spatio-Temporal Relationship

One of the most useful and interesting aspects of modeling data in GIS is the sophisticated analytical and reasoning potential. This section discusses such analytical capabilities of the proposed model. Changes of a place may occur spatially, temporally, characteristically, and typically combinations of the three perspectives. The spatial perspective concerns the geometric changes. The temporal perspective deals with processes and paces of changes over time. The characteristic perspective refers to those transformations on attributes of a place. Clearly, place changes with at least two and typically all three perspectives involved. For instance, the spatial and characteristic changes have to be connected with time, mediated by associated processes. More spatio-temporal analysis based on this model can be explored in further research. To start, this study will briefly explain a method of spatio-temporal topological reasoning based on the model.

A large literature on topological modeling can be found using primitive sets of spatial features. Well-known examples include the intersection models (Egenhofer and Franzosa 1991) and Region Connection Calculus models (Randell et al. 1992). Proposed are also models to reason about the topological relations between features with indeterminate boundaries (Cohn and Gotts 1996; Clementini and Felice 1996) or indeterminate locations (Galton and Hood 2005). Similarly, in the pursuit of analyzing topological relations from the integrative spatio-temporal perspective, new challenges arise due to the embedded interconnections between space and time. In his seminal work, Allen (1984) presents the formal temporal logic to model temporal topology. The formalism has been widely followed by many researchers for reasoning about temporal or spatio-temporal topology (e.g. Howarth and Couclelis 2005; Claramunt and Jiang 2001). Drawing on the prior work, this chapter will present another attempt in this line of research. Unlike previous studies, the presented method further analyzes the topological relations between two primitive regions and then combines it with temporal topology. To analyze the spatio-temporal topological relationship between two primitive ST regions A and B, this study defines a two-step approach as follows:

1. Identify two temporal primitive sets for each primitive ST region.

An ST region A has a time interval t. The start and ending points, which are collectively called the boundary of the time interval, are denoted as $t_0^A$ and $t_1^A$ respectively. The interior of the time interval is denoted as $t^A$, which refers to all time points within the interval except the two boundary points. The temporal topology between two A and B can be determined by intersections of the primitive sets, as expressed in Equation (16.4).

$$TR(A,B) = \begin{pmatrix} t^A \cap t^B & t^A \cap t_0^B & t^A \cap t_1^B \\ t_0^A \cap t^B & t_0^A \cap t_0^B & t_0^A \cap t_1^B \\ t_1^A \cap t^B & t_1^A \cap t_0^B & t_1^A \cap t_1^B \end{pmatrix} \tag{16.4}$$

Suppose A and B are two time intervals, and A is the shorter than B. Figure 16.5 lists some possible temporal relations, for illustration purpose.

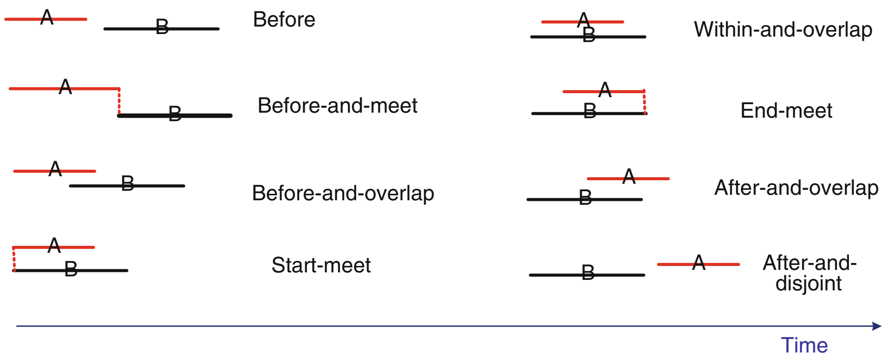



**Fig. 16.5** Temporal topology between a shorter time interval A and a longer time interval B

2. The topological intersection model can be extended to the spatio-temporal dimension. First, let us define the four types of spatial primitive sets of a primitive ST region A. They are also illustrated in Fig. 16.6.

   - the boundary of the footprint of the region at starting time $t_0^A$, denoted as $\partial A_0$
   - the boundary of the a footprint of the region at $t_1^A$, denoted as $\partial A_1$
   - the interior of A (the projection of the region on the X-Y space plane) at $t_0^A$, denoted as $A_0^o$
   - the interior of A (the projection of the region on the X–Y space plane) at $t_1^A$, denoted as $A_1^o$

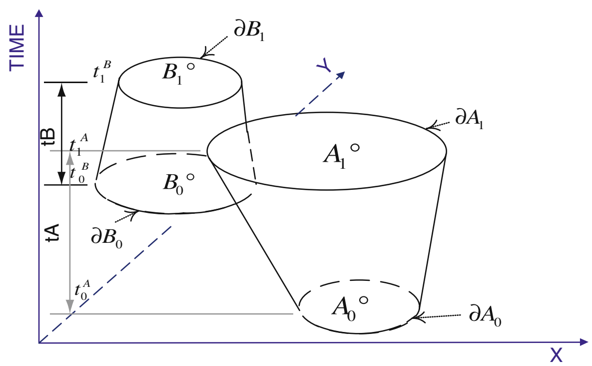



**Fig. 16.6** The primitive sets of an ST-region for topological analysis

Then the topological relationship between two primitive ST regions A and B boils down to the values of the sixteen intersections of primitive sets in Equation (16.5). The model considers values (i.e., empty or non-empty) of the sixteen intersections.

$$SR(A,B) = \begin{pmatrix} A_0^\circ \cap B_0^\circ & A_0^\circ \cap B_1^\circ & A_0^\circ \cap \partial B_0 & A_0^\circ \cap B_1 \\ A_1^\circ \cap B_0^\circ & A_1^\circ \cap B_1^\circ & A_1^\circ \cap \partial B_0 & A_1^\circ \cap \partial B_1 \\ \partial A_0 \cap B_0^\circ & \partial A_0 \cap B_1^\circ & \partial A_0 \cap \partial B_0 & \partial A_0 \cap \partial B_1 \\ \partial A_1 \cap B_0^\circ & \partial A_1 \cap B_1^\circ & \partial A_1 \cap \partial B_0 & \partial A_1 \cap \partial B_1 \end{pmatrix} \quad (16.5)$$

The final analysis result is TR×SR, which combines results from the above two steps to concisely describe the spatio-temporal topological relationship between the two primitive ST regions.

Let us demonstrate how it works using the real world example of Beijing shown in Fig. 16.3 and the hypothetic example in Fig. 16.6. Table 16.2 lists the analysis results.

**Table 16.2** Analysis example: ST topological relationship analysis between primitive ST regions shown in Figs. 16.3 and 16.6

| Primitive ST regions | Zhongdu (A) and Dadu (B) | Zhongdu (A) and Beijing (B) | A and B in Fig. 16.6 |
|---|---|---|---|
| TR matrix | $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ | $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ | $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$ |
| Temporal relation | Zhongdu is "before and meet" Dadu | Beijing is "after" Zhongdu | A is "before and overlap " B |
| SR matrix | $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ | $\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$ | $\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$ |
| Spatial relation | Disjoint all the time | Overlap all the time | B disjoints A at the beginning of its lifespan but later expanded or shifted to overlap A |
| Spatio-temporal topology | Temporally, Zhongdu is before-and-meet Dadu while the two cities remain spatially disjoint all the time | Temporally, Beijing is after-and-disjoint Zhongdu, while the two cities spatially overlap all the time | A is before-and-overlap B; while they are spatially disjoint at the beginning of B, but B grew to overlap A |

## 16.6 Discussions and Future Research

Place is a complex spatio-temporal construct. While previous studies of modeling places assume the static view of place, this study concerns the modeling of city places from the integrative spatio-temporal perspective. Starting with an analysis of the ontological underpinnings, the study presents a spatio-temporal place model. The key components of the place model are the primitive ST regions and spatio-temporal processes. The chapter illustrates the process of constructing a primitive ST region. Based on the model, the chapter also attempt to reason about spatio-temporal topological relationships between places.

While the proposed model has great potential for sophisticated spatio-temporal analysis functions, several issues remain open for further research. First, although discussions in the chapter recognize the innate fuzzy nature of places concerning both the identity of a place and the spatial boundary of it, the proposed model still assumes clear-cut boundaries in operation. Previous studies in the literature have already developed methods dealing with representation and reasoning of uncertain boundaries (e.g. Cohn and Gotts 1996; Clementini and Felice 1996; Galton and Hood 2005), future research is necessary to incorporate these methods into the spatio-temporal place model so that applying fuzzy boundaries can be facilitated in the place model. Secondly, temporal interpolation is an indispensable step in the construction of primitive ST regions with the type of process that result in continuous spatial changes. The step requires efficient spatio-temporal surface fitting and interpolation techniques. This is an underexplored research area. Thus developing a suitable method for spatio-temporal interpolation is an important topic for further research along the line. Finally, this chapter did not discuss the issue of analyzing characteristic changes over space and time, which implies an integrative triplet perspective of space-time-attribute. It will make another interesting research avenue.

## References


ADL (Alexandria Digital Library) (2007). Gazetteer Server. http://webclient.alexandria.ucsb.edu/. Last Accessed on 12/10/2007.

Agarwal, P. (2005). Operationalising ‘sense of place’ as a cognitive operator for semantics in place-based ontologies. In A.G. Cohn & D.M. Mark (Eds.), *Spatial information theory, 7th international conference COSIT* (pp. 96–114). (Berlin: Springer)

Allen, J.F. (1984). Towards a general theory of action and time. *Artificial Intelligence,* 23(2), 123–154.

Casey, E.S. (1997). *The fate of place*. (Berkeley: University of California Press)

Claramunt, C., & Jiang, B. (2001). An integrated representation of spatial and temporal relationships between evolving regions. *Journal of geographical systems* 3(4), 411–428

Clementini, E. & Felice, P.D. (1996). An algebraic model for spatial objects with indeterminate boundaries. In P.A. Burrough & A.U. Frank (Eds), *Geographic objects with indeterminate boundaries* (pp. 155–170). (Bristol, PA: Taylor & Francis)

Cohn, A.G. & Gotts, N.M. (1996). The ’egg-yolk’ representation of regions with indeterminate boundaries. In P.A. Burrough & A.U. Frank (Eds.), *Geographic objects with indeterminate boundaries* (pp. 171–187). (Bristol, PA: Taylor & Francis)

Compieta, P., Di Martino, S., Bertolotto, M., Ferrucci, F. & Kechadi T. (2007). Exploratory spatio-temporal data mining and visualization. *Journal of Visual Languages and Computing,* 18(3), 255–279

Egenhofer, M.J., and Franzosa, R. (1991). Point-set Topological Relations. *International Journal of Geographical Information Systems*, 5(2), 161–174

Galton, A. (2003). Desiderata for a spatio-temporal geo-ontology. In W. Kuhn, M. F. Worboys & S. Timpf (Eds.), *Spatial information theory. Foundations of geographic information science. 6th International Conference COSIT* (pp. 1–12). (Berlin: Springer)

Galton, A. & Hood, J. (2005). Anchoring: a new approach to handling indeterminate location in GIS. In A.G. Cohn & D.M. Mark (Eds.), *Spatial information theory, 7th international conference COSIT 2005* (pp. 1–13). (Berlin: Springer)

Gazetteer for Scotland. (2009) http://www.geo.ed.ac.uk/scotgaz/gaztitle.html. Last accessed on May, 20, 2009

GNIS (Geographic Names Information System). (2009). Database for feature name query. http://en.wikipedia.org/wiki/Geographic_Names_Information_System. Last accessed on May, 20, 2009

Goodall, B. (Ed.) (1987). *The penguin dictionary of human geography*. (Harmondsworth: Penguin Books Ltd.)

Goodchild, M.F. (1987). A spatial analytical perspective on geographic information systems. *International Journal of Geographic Information Systems,* 1, 327–334

Gould, S.J. (1987). *Time's arrow and time's cycle: myth and metaphor in the discovery of geological time*. (Cambridge, MA: Harvard University Press)

Grenon, P. & Smith, B. (2004). SNAP and SPAN: towards dynamic spatial ontology. *Spatial Cognition and Computation,* 4, 69–104

Guarino, N. & Giaretta, P. (1995). Ontologies and knowledge bases: towards a terminological clarification. In N.J.I. Mars (Ed.), *Towards very large knowledge bases* (pp. 25–32). (Trento, Italy: IOS Press)

Hägerstrand, T. (1970). What about people in regional science? *Paper of the Regional Science Association,* 24(1), 7–21

Hamilton, H.J., Geng, L., Findlater, L. & Randall. D.J. (2006). Efficient spatio-temporal data mining with GenSpace graphs. *Journal of Applied Logic,* 4(2), 192–214

Hill, L.L. (2006). *Georeferencing: the geographic associations of information*. (Cambridge, MA: The MIT Press)

Hornsby, K. & Egenhofer, M.J. (2000). Identity-based change: a foundation for spatio-temporal knowledge representation. *International Journal of Geographic Information Science,* 14(3), 207–224

Howarth, J.T., & Couclelis, H. (2005). A linguistics-based framework for modeling spatio-temporal occurrences and purposive change. In A.G. Cohn & D.M. Mark (Eds.), *Spatial information theory, 7th international conference COSIT* (pp. 316–329). (Berlin: Springer)

Jordan, T., Raubal, M., Gartrell, B. & Egenhofer, M.J. (1998). An affordance-based model of place in GIS. In T. Poiker & N. Chrisman (Eds.). *8th international symposium on spatial data handling* (pp. 98–109). (Vancouver, Canada. International Geographic Union)

Kwan, M.-P. (1998). Space-time and integral measures of Individual accessibility: a comparative analysis using a point-based framework, *Geographical Analysis,* 30(3), 191–216

Kwan, M.-P. (2004). GIS methods in time-geographic research: geocomputation and geovisualization of human activity patterns. *Geografiska Annaler, Series B: Human Geography,* 86(4), 267–280

Langran, G. (1992). *Time in geographic information systems*. (London: Taylor & Francis)

Mark, D.M., Egenhofer, M.J., Hirtle, S. & Smith, B. (2000). *Ontological foundations for geographic information science* 2000 [cited. Available from http://www.ucgis.org/priorities/research/research_white/2000whitepapersindex.htm]

Miller, H.J. (1999). Measuring space-time accessibility benefits within transportation networks: basic theory and computational procedures, *Geographical Analysis,* 31, 187–212

Miller, H.J. & Wu, Y.-H. (2000). GIS software for measuring space-time accessibility in transportation planning and analysis, *GeoInformatica,* 4(2), 141–159

National GeoOnet Names Server (GNS). by National Geospatial-Intelligence Agency (NGA). http://earth-info.nga.mil/gns/html/index.html. Last accessed on 5/20/2009.
Neutens, T., Witlox, F. Van De Weghe, N. & De Maeyer, P.H. (2007). Space-time opportunities for multiple agents: a constraint-based approach, *International Journal of Geographical Information Science,* 21(10), 1061–1076
Parkes, D. & Thrift, N. (1978). In T. Calstein, D. Parkes & N. Thrift (Eds.), *Putting time in its place* (pp. 119–129). (London: Edward Arnold)
Peuquet, D.J. (1994). It's about time: a conceptual framework for the representation of temporal dynamics in geographic information systems, *Annals of the Association of American Geographers,* 84 (3), 441–461
Peuquet, D.J. (2001). Making space for time: issues in space-time data representation. *GeoInformatica,* 5(1), 11–32
Peuquet, D.J. & Duan, N. (1995). An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data. *International Journal of Geographic Information Science,* 9, 7–24
Randell, D.A., Cui, Z. & Cohn, A.G. (1992). A spatial logic based on regions and connection. In B. Bebel, W. Swarthout & C. Rich (Eds.), *Proceedings of 3rd international conference on knowledge representation and reasoning* (pp. 165–176). (San Francisco: Morgan Kaufmann.)
Relph, E. (1976). *Place and placelessness*. (London: Pion Limited)
Tuan, Y.-F. (1975). Place: an experiential perspective. *The Geographical Review*, 65(2), 51–165
Tuan, Y.-F. (2001). *Space and Place: the perspective of experiences*. (Minneapolis: University of Minnesota Press)
UNFPA. (2007). State of world population 2007: unleashing the potential of urban growth. Published by UNFPA. 108 pages. Also available on http://www.unfpa.org/public/publications/pid/408 (accessed on 12/1/2009)
Wachowicz, M. (1999). *Object-oriented design for temporal GIS*. (London: Taylor & Francis)
Wang, D. & Cheng, T. (2001). A spatio-temporal data model for activity-based transport demand modeling, *International Journal of Geographical Information Science,* 15(6), 561–585
Worboys, M. (1998). A generic model for spatio-bitemporal geographic information. In M.J. Egenhofer & R.G. Golledge (Eds.), *Spatial and temporal reasoning in geographic information systems* (pp. 25–39). (Oxford: Oxford University Press)
Worboys, M. (2005). Event-oriented approaches to geographic phenomena, *International Journal of Geographical Information Science,* 19(1), 1–28
Worboys, M., & Duckham, M. (2006). Monitoring qualitative spatiotemporal change for geosensor networks. *International Journal of Geographical Information Science,* 20(10), 1087–1108
Yu, H., & Shaw, S.L. (2008). Exploring potential human activities in physical and virtual spaces: a spatio-temporal GIS approach. *International Journal of Geographical Information Science,* 22(4), 409–430
Yuan, M. (1997). Use of knowledge acquisition to build wildfire representation in geographical information systems. *International Journal of Geographic Information Science,* 11, 723–745

# Part VI
# Studying Other Urban Problems Using Geospatial Analysis and Modeling

# Chapter 17
# Geospatial Analysis and Living Urban Geometry

**Pietro Pagliardini, Sergio Porta, and Nikos A. Salingaros**

**Abstract** This essay outlines how to incorporate morphological rules within the exigencies of our technological age. We propose using the current evolution of Geographical Information Systems (GIS) technologies beyond their original representational domain, towards predictive and dynamic spatial models that help in constructing the new discipline of "urban seeding". We condemn the high-rise tower block as an unsuitable typology for a living city, and propose to re-establish human-scale urban fabric that resembles the traditional city. Pedestrian presence, density, and movement all reveal that open space between modernist buildings is not urban at all, but neither is the open space found in today's sprawling suburbs. True urban space contains and encourages pedestrian interactions, and has to be designed and built according to specific rules. The opposition between traditional self-organized versus modernist planned cities challenges the very core of the urban planning discipline. Planning has to be re-framed from being a tool creating a fixed future to become a visionary adaptive tool of dynamic states in evolution.

**Keywords** Urban morphology · Network city · Urban space · Evolutionary design · Urban seeding

## 17.1 Introduction

It is easy to understand what determines optimal urban morphology. By means of recently derived geometrical rules (Alexander 2001, 2005; Krier 1998; Salingaros 2005; Salingaros et al. 2006; Steil et al. 2008), we know what can be done to fix the inhuman urban form found both in today's suburban sprawl, and in dysfunctional housing projects. These rules are contained, in

N.A. Salingaros (✉)
Department of Mathematics, University of Texas at San Antonio, San Antonio, TX 78249, USA
e-mail: yxk833@my.utsa.edu

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_17,

part, in traditional urbanist knowledge (Caniggia and Maffei 2001; Krier 1998; Marzot 2002), but that is now unfortunately considered “out of fashion”. Most importantly, minimal interventions are possible in order to recover a large part of failed projects. Whereas the form of some structures is indeed so inhuman that demolition is necessary, many cases could be significantly improved by much cheaper interventions on the existing urban fabric (Steil et al. 2008).

Prominent politicians declare an interest in building for low-income residents, yet blocks of public housing by famous architects actually resulted in inhuman living environments (Porta 2008). We believe that solutions offered up until now have been, for the most part, unsuccessful (Salingaros et al. 2006). In particular, the usual modernist urban planning typologies do not go anywhere near towards understanding the urban geometry needed for resident satisfaction. Politicians and architects now declare that we must demolish existing structures and reconstruct new quarters for social housing. This admirable sentiment does not take into account, however, that the present generation of architects connected to the political system could likely create yet another failed urban experiment. An ever-increasing number of high-profile urban projects directed by “star” architects are designed according to the same criteria as before, only that the buildings look more “sparkling” externally.

We are not speaking about the failure of a set of theories, or even single or a group of architects and planners: it is the failure of an entire discipline, which originated at the end of the nineteenth century around ideas of top-down control. Urban phenomena have now been recognized as enormously complex and therefore inherently uncontrollable from the top–down. We do not just need better architects and planners: we actually need architects and planners of an entirely different kind, who take the challenge of self-organization in cities seriously enough to investigate new forms of description, prediction, and intervention. Especially, we need a broader awareness that all this has nothing to do with *style* and everything to do with *structure*. The process of spatial evolution in traditional cities has always been unplanned, and this freedom must be mimicked to generate good future cities. Planners have to focus on the structural drivers of such evolution in order to manage the seeds of change, and not try to control its final state.

## 17.2 Networks of Urban Space in a Living City

We (together with many other authors) have now developed sufficient information and tools to be able to design new cities with the correct human-scale geometry, whose purpose is to enhance urban life. This characteristic ribbon geometry of urban pedestrian space follows very simple rules (Salingaros 2005):

1. A city’s life is the direct result of pedestrians using its public urban spaces.
2. Urban space is an open container for crisscrossing footpaths, protected from, but at the same time connected to all other forms of transportation.

3. Urban space also provides the setting for the crucial human contact with nature.
4. The function of building fronts is to enhance the enclosure and informational properties of urban space.
5. All urban space is connected in a pedestrian network: sidewalks simply widen out into plazas.
6. A street is urban space that allows itself to be traversed by vehicular traffic, sacrificing pedestrian space locally in exchange for connecting pedestrian space globally.
7. Where the pedestrian network crosses another transport network, pedestrians must be protected by the physical structure itself.
8. When a city doesn't provide living urban space, private developers will do so, but then it is disconnected from the urban fabric.

Networks are flows defined by movements of people, vehicles, goods, information, etc. Any physical flow imposes linearization on a city, thus city flows tend to be linear. That does not necessarily mean a rectangular grid, although an initial rectangular grid does help to establish a good linear flow after the built environment has developed. In older settlements, we see a more organic urban fabric that has been linearized afterwards so that flow can occur, sometimes by cutting roads into older urban fabric. Both these methods are entirely valid.

There is also a third very relevant model: cities founded formally that later evolve an organic urban fabric. Those morphologies originated from a formal top-down plan (mostly grid layout) that then developed historically into an organic settlement through time. A very common pattern of so many medieval towns in Italy and elsewhere in the former Roman Empire is their origin as roman military camps with a regular grid, and then their evolution into "organic" cities because of endless bottom-up changes and modifications (Bertuglia and Staricco 2000). The ensuing organic growth is contingent upon relaxation of the initially rigid geometry, in these cases due to drastic changes in society.

The main misunderstanding with today's urban form is that planners mistakenly believe that priority must be given to the fastest automotive traffic (Hall 2008). This error in thinking precludes planning for all the myriad small-scale movements and slow flows necessary for a living city. Another casualty of this approach is that, as a general principle, flows are made to erase stationary places such as plazas and parks that combine tangential pedestrian flow with pedestrian nodes. Those spaces must be protected from street traffic (Salingaros 2005).

Modernist planning is by its very nature devoted to separation. That practice draws back upon outdated scientific thinking (from which modernism claimed its roots), which in fact cannot deal with complex systems (Porta 1999). Cities, like organisms, are the prime examples of complex systems. Separation is nevertheless the gospel in every aspect of modernist theories on cities; therefore, separation of urban space users is just an application of this attitude to over-regulating urban life. One example of this, in addition to squares and parks, is the boulevard. Boulevards successfully combine rapid mechanical flows, slow mechanical flows, pedestrian flows, pedestrian stationary nodes, etc. (Jacobs et al. 2003). Boulevards were legally banned at the beginning of the twentieth century because they were complex spaces that gave place to different kinds of networks altogether.

A healthy mix of social classes and uses is obtained first of all by having an urban design that allows such a mix to occur. There exist distinct approaches, all converging towards a type of urban form that brings us back to the great historical city tradition. We prioritize establishing living urban fabric, before getting into any specific architectural style. Once the first-order problems of urban morphology have been resolved, then the city can support any architectural mix that contributes to the richness and variety of its urban environment.

Adaptive urbanism is not a stylistic problem, and is not tied exclusively to architecture, but rather to the complexity of the city as a whole. The disasters of today's periphery are not only degraded buildings, but also the inhuman quality of its urban environment. We believe that good urbanism can support some ugly buildings, whereas poor urbanism with even a few good buildings is not repairable. For this reason, traditional urbanism represents a genuine novelty in its global vision, similar in scope to that of the modernists but opposite in its aims, in which the street becomes the ordering element and a block is no longer a lot with a building in its center. The street and street fronts should be reconstructed so as to guarantee a strong and viable pedestrian realm.

Any city that looks too regular or that exists only on the large scale must be treated with suspicion: it may be alive, but there are primary reasons to warrant checking if human activity and movement are indeed taking place on all scales. A living geometry encourages movement and the utilization of space (not to be confused with people putting up with a hostile geometry because they have no other choice). If that is not happening, then the city is dead. Although hundreds of millions of persons have been forced to live in dead urban regions that does not make those places any more alive in an essential sense. Those dead regions actually look very neat and ordered from the air.

There are some who say that since society has changed, it is necessary that the city must also adapt by changing its form radically. Others believe instead that urban values remain valid because human biology has not changed. Those timeless urban values, however, are no longer resident in our consciousness, since they belong to the lost part of our collective intelligence. They now have to be rediscovered as a proper "discipline" that architects, urbanists, and historians must study and carefully apply, adapting it to the times. Since we boast of living in a mature and evolved society, we can make an intelligent effort of recovering rules that have determined beautiful and functional cities. We can then apply them to our present situation. The question is not if these rules are old but only that they are valid.

## 17.3 Geospatial Analysis and "Urban Seeding"

At the heart of our thesis is the distinction between living and dead urban fabric. There appears to be a strict relation between the *quality* of city spaces (i.e. living or dead) and the kind of process that lies behind their formation and evolution. That is, whether the processes are self-organized at the small scale from the bottom-up, or planned at the large scale from the top−down. What we are depicting here is the failure of a discipline that for the entire span of the last century has brought an anti-human realization wherever it acted on the ground, to the extent that it was allowed to by other external forces equally at work. In order to inject the positive quality that Christopher Alexander named "wholeness" (Alexander, 2001-2005; Alexander et al. 1987), which can only emerge and cannot be designed in cities, we need new practices for the description, prediction, and transformation of urban spaces of an entirely different genre. This different approach, which we term "urban seeding" instead of "urban planning", should be characterized by the following processes:

1. A distinction about what is *structural* (common to entire macro-cultural regions and changing at a slower pace of time) and what is *super-structural* (specific of every micro-cultural enclave and changing at a faster pace in time).
2. An awareness that while the super-structural dimension should be left to mainly self-organized processes of change and therefore does not regard planners at all, the structural dimension can and should be managed by formal processes of control, which in turn should be led by a mainly participative organizational rule.
3. An enhanced understanding of the structural dynamics of change that characterize the evolution of self-organized urban settlements, according to which proper policies can be put into place for targeting achievements that are shared by urban communities.

It is at this latter level that GIS technologies can help enormously. GIS is itself evolving from a set of computer-based technologies for describing spatially-related entities, into a very diverse set of procedures that increasingly have to do with modeling and predicting spatially-related dynamics. While embedding time variation in GIS applications and in network spatial analysis is still a matter of vision and research (Batty 2002, 2008), GIS scientists are increasingly working with interdisciplinary approaches that are at the heart of self-organization in spatial structures. GIS technologies are now intersecting studies in self-organized flows by means of cellular automata (Wagner 1997); fractal geometry and emerging urban form (Batty and Longley 1994); *e*-government, ubiquitous computing and the Internet as means for electronically-mediated public participation in decision-making (Carver 2001); community envisioning and urban design (Thwaites and Simkins 2007); information management for sustainable development (Campagna 2005); Virtual Reality and community involvement (Faust 1995); remote sensing (Donnay et al. 2001); pedestrian behavior and public life (Golicnik 2007); the physics of complex networks (Boccaletti et al. 2006); spatial networks, not limited to street networks, as affecting crucial urban dynamics such as land-use, community building, and crime (Jiang and Claramunt 2002; Porta et al. 2006).

In short, GIS helps to develop a much deeper understanding of key factors that rule the emerging spatial order at the structural level of city evolution. Up until very recently, this extremely important factor has been elusive because of our limited methods of measurement. Gathering and processing data on human behavior in the past required very costly video cameras set up for weeks in a particular spot (Whyte 1988). Now, we can display an enormous amount of data in visual form, processed in various ways that reference geographical locations in space by means of remote sensing techniques associated with GIS (Senseable City Lab 2006). Discovered patterns of use that confirm earlier theoretical results can be used as the basis for a radical re-organization of urban use and government policy. Where strong connections are concentrated only into a few channels, and if those channels are exclusively long-distance, then the urban morphology must clearly change to encourage shorter connections.

A basic topic of investigation that we cannot pursue in depth here is the analogy with Finite-Element Methods of analysis, a technique developed in Mathematics and Engineering. A dynamic complex system is decomposed into components that have distinct time variations. As a first approximation, those elements that vary more slowly are treated one way, whereas those that vary more rapidly are treated in another. The methods used for analysis are thus related to our model of city-as-computer mentioned later: *structural* features correspond to hardware, while *super-structural* features correspond to software. Continuously varying urban processes are decomposed and studied in the appropriate time frame. The finite-element analysis analogy underlines that even the structural features are evolving with time, but this is happening on a different time scale (Harris 2008).

## 17.4 Useful and Useless Urban Space

The two principal typologies generating dead contemporary left-over urban space are: the unenclosed open plazas between modernist tower blocks, and the green space between isolated suburban houses whose garage doors face each other across the street GIS analysis confirms our predictions that the vast concrete plazas between residential tower blocks, and the exterior space in the sprawling suburb, are characteristic of very low people utilization. Both are in fact purely decorative elements, hence from their very inception, wasted space. Some open plazas between office blocks may be used during working hours, but are usually empty in the evening. There has been a tremendous confusion on these points, for two separate reasons.

First, as far as dead modernist urban space is concerned, architects, urbanists, and politicians have believed modernist design dogma, which claimed that the city of the future needed those vast unenclosed spaces, the products of formalist design. A lot of political baggage was tied to formal abstract geometries, with attractive promises such as "freedom", "individual liberty", "political breathing space", etc. None of these slogans ever took into account the scientific nature of city form, nor paid attention to actual people use. The iconic power of modernist images guaranteed their continued application, a regrettable practice that continues to this day because those wasteful urban typologies have been legalized into the planning system.

Second, the vast open space found in the sprawling suburb has escaped intelligent analysis because it looks so good in a photo. Or rather, only when there are no cars parked in the street, which is usually how the ideal suburb is photographed. In reality, the strip of green defined by road and open suburban front lawns is next to useless because it is too exposed (Salingaros 2005). Lawn and trees in front yards simply do not define useful urban space because of the fundamentally irresolvable conflicts about their open/enclosed and private/public qualities (Alexander et al. 1977). The urban character of this typology, now occupying the majority of surface area in suburbia, was never properly thought out. The urban space of suburbia is primarily occupied with what amounts to a giant parking lot (curbside parking on the excessively wide road, and on driveways in front of the house garages, because those are now full of consumer junk so the cars don't fit inside).

American suburbanites accepted the burden of buying and maintaining a front lawn that they themselves never use, and which is meant strictly for the visual promotion of the suburban icon for private development companies. The front lawn is a fiercely private realm that cannot be used by others. As a result, no one uses this wasted space on any regular basis. In Europe, the same wasteful philosophy, replacing natural connections by formal, visual typologies, has led to the useless garden. Bearing no relation to older gardens that give pleasure to pedestrians, we see in Europe pieces of green that are inaccessible, isolated, and usually constrained by an unnatural formalistic geometry (as seen both in their overall shape, and in the shape of what they contain such as the planters built for trees).

People are so thoroughly confused about urban space after decades of seductive modernist images in the media that they no longer have any certainty of their own innate spatial feelings. Those educated in a Western-influenced system cannot distinguish between useful and useless open space, or between what should be private and what should be public open space. Residents of the developing world, who are forced to rely upon their basic intuitions, usually can, however. Added to this confusion, we have a legislated set of post-war regulations that usually forbid one to enclose any legally owned space so as to make it useful.

As we accumulate data on pedestrian presence and movement, we can pick out and classify those urban regions where people can be found. Then, we can follow movement to plot frequency and length of pedestrian trips. For example, the front entrance to a suburban house is rarely used, remaining an expensive and stubbornly decorative architectural element. On the other hand, traditional urban space in historic cities, and open spaces in owner-built informal settlements (favelas) both attract an incredibly high density of human presence. Studies establish correlations between human presence and the shape of urban space (Salingaros 2005). Those spaces are alive, providing paradigmatic urban environments of a living city. There is a growing interest in technologies for the remote sensing of people in urban spaces and for tracking the movement of persons in sectors of cities. There is more experience for small groups of people in small places. These conclusions raise the question of whether post-war planners created the "city of the future" deliberately to keep people confined inside buildings (Oliva-i-Casas 2001).

Architects routinely engage in deception (or just wishful thinking) by showing unrealized projects with pedestrians all over their dead open spaces. It's very easy to draw figures in the open space of projects in a presentation, but those attractive (and deceptively realistic) images usually bear no relation to the eventual use of the spaces they show. In more cases than not, those spaces are useless because people will simply not go there; there is no reason to do so. If pedestrians ever wander out into that space accidentally, they feel insecurity, anxiety, a feeling of being unanchored to the place, disoriented, or a feeling of being lost. This reaction is the sensation one experiences when walking in a giant parking lot. People will therefore avoid such a space, and will get out of it as fast as possible.

Poor urbanism continues to be practiced, enthusiastically supported by the universities, because both private developers and city governments refuse to accept the scientific basis of good urbanism. They continue to listen to academic experts who dismiss human-scale solutions so as to promote their own ideologically-based mechanistic fantasies. Geographical Information Systems can finally resolve some old questions and establish the knowledge basis for urban regeneration. Nowhere is this drastic revision more desperately needed than in housing for the poor.

## 17.5 Urban Complexity and Modular Decomposition

The underlying concept in creating a living city lies in accepting its hierarchical complexity (Salingaros 2005). A city must be made up of interconnected components, each of which functions in as complete a manner as possible. For example, a person's house or apartment should be as self-sufficient as possible, according to its urban geometry, whereas house clusters and apartment buildings should try to be self-sufficient and not depend for every minimal function on the rest of the city. These smaller units should then fit coherently into the larger units, thus defining a hierarchy of scales of increasing urban complexity.

This is the idea of modular decomposition, the opposite of the idea of functional segregation. Its application to cities suffers from serious misunderstandings about the complexity of each module (Salingaros 2005).

The problem is that modernist planners, in ignoring the network structure of living cities, defined simplistic geometric modules that tried to eliminate complexity. As a consequence, modularization of urban elements led to cutting the essential connections that drive a living city to function. Hardware and software modularization (and its analogies in countless biological cases), by contrast, works by containing a maximum of complexity within each module. There is a correct and a wrong way to define modules in a complex system. These results have been known for decades in the scientific literature, but are still ignored today by most architects and urbanists (Salingaros 2005).

Modernist housing blocks represent the worst sort of wrong-headed modularization of urban components. They are really industrial structures for storing people (and their lives), treating those people like machines. Housing blocks follow the typologies of early industrialization, with their visually mechanical (and inhuman) rectangular block form. The urbanist elements here consist of "standards" and "services", used in a purely quantitative mechanical vision of the city. This conception concentrates mechanical operations and then distributes them to the different urban quarters according to some technical measurements, ignoring the rules necessary for generating the network of social and connective relations of the living city. This logic can never create the street and spatial hierarchies that characterize traditional urban planning.

Buildings are then dropped into urban quarters without any possibility of ever becoming attractive nodes in the urban fabric. There is no trace of continuous street fronts created from buildings in direct relationship with the street itself, or of residences placed side by side (and obviously not isolated). A *horizontal* social relation in the neighborhood is no longer possible. The planned elimination of social urban space is then substituted with a fictitious *vertical* relation on a useless and exaggerated scale, introducing "condominium common spaces" dedicated to the lost social ties, but capable only of creating indifference, fear, alienation, and conflicts. Public buildings (and their services) no longer succeed in becoming objects of collective concern, deprived as they are of every shared symbolic or aesthetic quality, because they are based solely on "needs" established by a power base.

## 17.6 Three Different Metaphors for a City

Three useful analogies for a city are the "city-as-organism" considered by Lynch (Lynch 1981), the "city-as-machine", and finally the "city-as-computer" (Salingaros 2005). City-as-organism supposes that a living city works much like an animal (and specifically, human) body, with separate circulation systems (nerves, blood, lungs), components of body structure (skeleton, tissues, organs), central control (heart, brain), peripheries, etc. This analogy seems to fit the great cities of traditional urbanism (i.e. pre-twentieth century urban fabric). It has proved a useful metaphor, but one easily twisted into an inhuman environment that only superficially copies structural complexity. The point is that we must build urban fabric according to its own structural laws and not by copying visual complexity from another system. Architects and planners fall into this trap by thinking artistically instead of scientifically.

The second analogy, the city-as-machine, is often proposed as the opposite model for a city: the city as a working machine, not biological but entirely artificial. Contrasting the two opposites, the city can range from being more like an organism (good) or more like a machine (bad). Describing the city in these terms is useful, and gives some practical insight into urban structure and function. Nevertheless, every city is necessarily an artificial construction, hence by definition a machine, though a very sophisticated one. It would be more accurate to describe a city as a combined biological/mechanical system.

The city-as-machine is a rudimentary instrument because it lacks diffused hierarchical connections. Placing a formal simplistic plan side by side with the plan of a historical city center, one cannot help but perceive the poverty of the former and the spontaneous beauty of the latter (unless you happen to have gone through architectural training). In urban design, the true difficulty lies in reproducing, or trying to approach as closely as possible, the level of complexity of an organism. The city-as-organism is something more than a metaphor. A city constituted from overlapping, interacting networks approaches the structure of an organism more substantially (and less and less metaphorically) because it increases its degree of useful complexity.

The third analogy, the "city-as-computer", distinguishes the two parts of any computing system into hardware and software (Salingaros 2005). A computer is clearly separated into its physical components (the built pieces and circuits), and its software (which is strictly informational): each relies upon the other to work together. In a city, hardware is built into solid structures (buildings, roads, infrastructure, etc.) whereas software consists of the moving elements (people, cars, goods, energy, etc.). A city provides the solid framework upon which movement of information (the analogy to software) can take place.

A functioning information-exchange system requires a level of complexity (Salingaros 2005). Below a certain threshold of complexity, the system is either inefficient or dysfunctional for its purpose. Therefore, we can find two types of computer systems (and many others that fall in-between). At one end, the hardware contains the requisite complexity in terms of hard-wired computational capability, while the software is relatively simple. At the other end, the hardware is simplified by modularization, thus shifting the requisite complexity into the software. In the latter case, the software has to assume the burden of complexity and becomes more prone to breakdown (i.e., less stable). Either way seems to work. The point is that the total complexity (hardware plus software complexity) is an invariant according to the computational task to be accomplished.

Let us now draw the analogy with urban systems. Traditional urban fabric mimics hardware complexity in that the built structures are highly complex, showing connective components and structures on all different scales (as for example in the historic city center and in the self-built favela). These complex structures help people to interconnect and to enrich their information exchange experience; more important, they exhibit a high capacity to accompany social and economic change, again at all scales, through the continuous adaptation of the overall city fabric through time. Such changes take place by means of countless initiatives outside any central control.

By contrast, the modernist city has simplified its hardware to either of two typologies: simplistic isolated high-rise blocks, and isolated suburban houses, with a series of buildings that carry out collective functions arranged randomly. It is still possible for people to achieve complex information exchange while living in these simplistic environments, but they have to pay the price: which is an extremely complex "software" consisting of cars, public transport, electronic communication, etc. Those connections turn out to be socially exclusive, in that they are not available to all individuals. Moreover, modernist urban environments are extremely rigid. They completely exclude continuous processes of bottom-up change because of the prevalence of large-scale developers (no matter whether private or public) and spatial units of development, coupled in all developing countries with patterns of "instant growth".

The metaphor of city-as-computer reveals the restrictions of simplistic modernist urban typologies. Whereas relatively wealthy and healthy adults can engage in complex movements in order to act out their lives as complex software, the poor, the young, the old, and the infirm cannot: they have no urban life at all. They are dependent upon precarious or non-existent connective systems to lead anything resembling an emotionally fulfilling life. Often, they simply exist as urban prisoners. All the efforts of city governments to provide the complex connectivity needed by those marginalized groups of individuals can never duplicate the effortless connective capabilities of complex "hardware" already present in the complex connective urban fabric of traditional city structure.

Let us not forget that the wealthy living in dead urban environments (say, a luxurious condominium in a tower isolated in a park, or a palatial house in remote suburbia) pay a lot of money to vacation in a European capital city in order to "re-charge" their urban information deficit. Ordinary suburban residents, and those living in monofunctional residential blocks on a city's periphery, pay transportation costs to experience urban life in a private commercial center or shopping mall. City life has, in many cases around the world, been transferred to private interests by governments who became convinced that the contemporary city doesn't need to encourage it.

## 17.7 Do We Wish to Connect to Our Neighbor?

It is neither useless nor utopian to demand the restoration of relations today with our neighbors. Maybe we have nothing in common with our house neighbor, and in fact, most of our social and professional conversations could be via the internet One small store under our apartment or near our house, reached comfortably on foot, could have a poor selection of goods and limited choice, and we certainly prefer to go by car on a highway to a shopping center or "big-box" store. In the same way, we take the car to reach the hospital, the school, the concert hall, and the soccer game. Urban life today offers a variety of choices, privileging the excellent ones via the mechanism of market competition. If the best is not next door, then we have access to it wherever it is located. At least, until we have a catastrophic energy crisis.

That is not the whole point, however. As described accurately by Hillel Schocken in his essay "Intimate Anonymity" (Schocken 2003), human beings have a craving for community, and it could just as well be a community of strangers. Seeing other people up close has a biologically beneficial effect on our organism (Kellert et al. 2008). This is one aspect of biophilia: we crave intimate contact with plants, natural environments, other animals, and other human beings. We cannot satisfy this need for contact with only our close friends, thus the traditional urban environment of non-threatening strangers turns out to be a key factor in an emotionally-nourishing city (Oliva-i-Casas 2001).

Scientists point out the importance of weak links between networks that are strongly connected internally (Granovetter 1973). Those weak links tie distinct networks together into a larger network. We interpret this mechanism in the urban context as follows. People have a strongly-connected network supporting their everyday life. Strong connections do not necessarily mean nearby ones, however (a drastic reversal from village life). We could be telecommuting, working for a company in another city, or driving our children to a good school far from home. Those distant links are the strong ones. The weak links in this case could be the persons and urban nodes close to one's own residence or workplace. Opening up the possibilities for casual contact and pleasurable direct experience outside one's normal routine is what makes the city alive.

Unlike in the modernist planning philosophy, where behavior is strictly imposed on the population, we are referring to creating situations that make individual choice possible. We wish to facilitate the random exposure and contact with other human beings, which in a properly designed urban environment is not chaotic. Gathering of people as well as spatial changes at the micro-scale of the urban structure are not pre-determined by anyone else, but they are *influenced* by urban geometry and spatial morphology. Nobody should be forced to meet other people, or to merge two adjacent lots, or to change the land use of her/his property, but all must have these possibilities. We (planners) must guarantee this degree of freedom to everyone. These provide the social environment for developing weak links. The possibility of establishing such weak links has driven urbanization ever since the earliest city.

Even mono-functionality is allowed, but should never be imposed. For example, small shops of the same kind do cluster under the appropriate circumstances. The spontaneous clustering of similar small retail commercial premises to form homogeneous commercial areas is a timeless pattern: we can find it anywhere, at anytime in history. This counters the idea that shops should be located at a convenient distance from each other in order to avoid the overlapping of market catchment areas, which stands behind many artificial models of regional and urban planning. The rationale underpinning the clustering pattern is exactly that small shops simply cannot offer, as single trading units, a sufficient range of goods to potential customers, whereas they can when taken altogether. What develops is the self-organized bazaar neighborhood. This argument supports the idea that retail commerce must be conceived in clustered areas characterized by spatial proximity, in order to allow such win-win dynamics to emerge.

Generally, given two adjacent urban quarters, there is a single road that joins them and another road perpendicular to the connecting one, which divides them. This gives rise to the following double problem. Movement between one zone and the other is obligatory, and we have traffic jams on the sole connector road. To go from one urban quarter to another, a pedestrian must cross a traffic artery that has no urban qualities, and thus usually abandons the idea of doing so. The weak links (those on the pedestrian scale) necessary for including the two urban quarters into a larger encompassing network are missing.

## 17.8 Respecting and Re-creating Complex Urban Fabric

Representing the type of living city people prefer is not very difficult. We have examples in the historical city centers of thousands of cities and villages. The difficult task is not to see the city-as-organism and to represent it, but to reproduce it anew every time someone writes up a master plan, divides open territory into building lots, or even designs a single room. Why it is so difficult to reproduce something that we believe (and claim) to know so well? And why does it seem that in the past (but apparently not nowadays) it was so easy to "invent" an organic urban fabric?

The reason is that nothing was ever invented at the large scale in the ordinary urban fabric of the city. People just built such fabrics by following their "collective intelligence", which itself derived from a more closed and thus less complex society than today's. People building traditional settlements followed rules that are in harmony with nature, or better said, in harmony with the geography. Those rules satisfied a society's basic requirements first of all (i.e. defense, presence of water, the availability of long-distance connecting roads, an agricultural hinterland, etc.). After those priorities, citizens tried to satisfy higher-level needs and urban requirements (i.e. commerce, political organization, religious life, etc.). Cities have a strong conceptual identity. There was a strong sense of belonging to a city, and having a house in the city was necessary to become a citizen of that city.

The modernist city doesn't work because human beings are not a product that can be commanded to move in a great metallic machine, like bottles in a bottling plant, without losing their humanity. Human beings are made up of feelings and instincts. They require the freedom to choose. In the city-as-machine the apparent choices are many, but the actual ones very few. In order to go from one place to another you have efficient roads of fast flow, but you are forced to take the car and submit to traffic stress. In Europe, you don't dare try to find an old acquaintance's house without using a satellite navigator in your car since you risk remaining blocked in disorienting dead-end roads from which you don't know how to exit. In order to meet other persons you have no other choice than to pay in order to enter specialized premises, or to go spend your money in a shopping mall or big-box hypermarket.

You do not have the freedom to let your children play by themselves in front of the house, unless you possess a private garden (preferably enclosed); you do not have the freedom to send small children off to school by themselves; you do not have the freedom to go out in the summer evenings to eat an ice cream in the company of other people without being forced to take the car, unless you live in the city's (historical) center. Otherwise, you resign yourself to an ice-cream cone consumed while driving along a squalid road full of cars.

The city-as-machine anticipates only programmed mechanical movements, and has few degrees of freedom because there is little choice among alternative flows. This holds true not only for the paths themselves, but also for the means of transportation along those paths. A city has to guarantee the mutual coexistence of pedestrians, cars, bicycles, etc., without any particular bias or specialization, except for special needs in very limited regions. Living cities are invariably those that are highly connected by internal networks (Salingaros 2005) and which show a fractal geometry. There exists complex structure at every scale of magnification, from the size of the entire city, to the size of a neighborhood, to a cluster of buildings, to a single building, down to the urban spaces, and further down to the scale of sidewalks, trees, and street furniture. Living urban structure reproduces, in abstract form, complex natural and biological structures, which exhibit fractal scaling.

The living city is understood from an appreciation of its networks, and many solutions to problems of urban planning follow from network thinking. Urban permeability is obvious from the point of view of vehicular traffic, but also has a major impact on personal and social relations, and thus on the general functioning of the city with its innumerable activities. A city has to guarantee the largest number of degrees of freedom to each citizen. The city-as-machine is the exact opposite of this choice.

We know of urban plazas in historic cities, where the success of the urban space is entirely independent of the architectural (but not geometrical) qualities of the surrounding buildings. Those might not exhibit any particular architectural interest, sometimes the façades are degraded, badly restored after interventions or war damage, and were not necessarily elegant even before historical wear and tear. The space, however, is articulated because it was established, connected, and reinforced during successive times. Such central spaces of the city have a great advantage by being extremely lively, used by many pedestrians, often containing markets of some sort, and have remained popular, even contributing to connect the historic center to the periphery. The point is that the urban space, by being alive, has rejuvenated the surrounding buildings. By adapting to the pedestrian uses of the urban space they enclose, the buildings have adapted typologically to the rest of the living urban fabric. Curiously, they have often evolved to look more like the older buildings in the historic center than the post-war buildings in the periphery.

## 17.9 The Problem of Designing the City's Periphery

Governments struggle with problems of new housing. In Europe, the population pressure tends to result in high-rise dormitories on the city's edge; whereas in the USA the same pressures are responsible for the sprawling (and unsustainable) dormitory suburbs constituted of isolated single-family houses. Both typologies are monofunctional, hence do not and cannot give birth to living urban fabric. A living city, by definition, has to combine functions in as close an intermixing as possible without any one function harming the others (Salingaros 2005).

An immense body of planning laws and regulations behind architects and urbanists determines the basic choices of future residents. Those regulations obstinately follow a modernist city model based on a building centered on its lot: they deny the need of the street acting as both ordering and connecting element, and totally neglect to relate a new building with the rest of the city. Streets are designed for fast vehicular movement, taking no account even of the existence of pedestrians (Hall 2008). Building norms that have imposed limitations of every kind were set up on the principle of giving people a roof, but not a home that is contained organically in a city. Those norms certainly have as their project's objective a faceless population of proletarians with the most basic mechanical needs rather than human beings with feelings.

The true responsibility for the dissolution of cities lies with the modernist regulating plan. People who would otherwise walk or bicycle to a local market in some central location, crossing the smaller streets of the historic center and running into friends and acquaintances, are now exiled to a no-man's-land. People are sent to live in the periphery where streets are no longer streets because they are not bounded by building façades. Houses or apartment buildings are rigorously separated 10 meters from each other, and set back from the street line. This creates useless pieces of lawn, garden, and open space. Streets thus become asphalt ribbon conduits that serve only the car. The trip to the center, deprived of stores and the possibility of encounters, is classed simply as "urban mobility".

Choosing to erect anonymous blocks without the slightest reference to the essential properties of a house, projects are then executed with pseudoscientific accuracy. Their architect painstakingly verifies the surface area of every single space based on a "predetermined need" as set into law by an elite that has decided for everybody else. Living quarters whose inhabitants cannot recognize as familiar places with which to identify have nothing in common with the traditional city as it was known for so many generations. Such new constructions exist on the same conceptual level as the periphery destroyed by speculative building: they have been planned, exalted, advertised, and studied in all the universities. Those projects have been taught as positive examples to students, by architects who have transformed a vision they originally declared to be "ethical" into an "aesthetic" dimension, which ended up as a mix of mechanization and political ideology.

Lest readers think we are exaggerating, we recall several gross failures of social housing in our times. The Pruitt-Igoe high-rises in Saint Louis, Missouri were dynamited by the government after they became so degraded and crime-ridden that they were an embarrassment to the city. In the USA, this failed project is now used as a case study of inhuman urbanism. The situation is different in Europe. Four examples of public housing built in Italy: Monte Amiata in Milan, Corviale in Rome, Scampia in Naples, and Zen in Palermo, were condemned by European urbanists in 1991 as being total and abysmal failures. Nevertheless, 15 years later, those very projects were spotlighted in an exhibition of innovative Italian architecture, which toured the major Italian universities (Porta 2008). These examples, wherein similar cases gave rise to opposite lessons, underline that the discipline itself stubbornly sticks to a failed ideology.

There exists a fundamental principle that is indifferent to the iconic and stylistic preferences of any individual (especially an architectural or planning "expert"): architecture has a civic value because it belongs to everybody. The city is a common and collective good. Since the city is not private, citizens have the right to choose the environment in which to live, to move, to work, to enjoy their recreation, to love, to study, to travel in, etc. How can we make it possible for citizens to choose their environment directly? A good starting point is for administrators to stop following self-referential planning and building regulations, and realize that every written rule has an immediate consequence on projects. Those rules are typically not written to oppress the citizen, but instead to obtain a narrow typological and morphological result.

Today everything from the past is disowned, and this may be a good thing in part because it is necessary to look ahead; but without a truly intelligent review of the causes of the urban disasters in the periphery, we risk making the very same mistake under a different disguise. We call attention to three dangerous contemporary trends.

1. Many people call for the figure of an "appointed" architect to fight all of society's ills, discharging all the present and past political mistakes, but who necessarily draws from the ideology of the generation of architects after the events in Paris of May 1968.
2. Urban models completely unrelated to those of our historical cities continue to be implemented, with plans indistinguishable in substance from those that are now targets for demolition. Those new proposals are apparently ennobled by architecture slightly more "in fashion". Nevertheless, in the next "season" they will also become out of style. Today's urban projects are enriched by new symbols, for example the skyscraper, which every self-respecting mayor fervently desires, using the ideological cover of the new urbanistically-correct catchword "sustainability". Sustainability thus becomes the generating element of the project, in a way that everything completely instrumental becomes substantially false. For example, we see the glass skyscraper presented as a sustainable typology, whereas it is the exact contrary.
3. Even the social mix often talked about seems oriented more to a totalitarian operation of social engineering (on an ethical level, to be sure) than to an organic idea supported by city design. Only human-scale urban planning can render the goal of social mix effectively possible, driven by the mechanism that borrows from the geometry achieved in the traditional city.

This is neither a stylistic issue, nor any romantic nostalgia for the past, as those architects who pretend to be passionate about modernity would have us believe. Negative slogans referring to "obsolete forms of social life" and "historically superseded public space" continue to support dead urban typologies that prevent the emergence of living urban fabric. Their ideological followers perpetuate dogmatic images of an idealized modernity, blind to the damage those wreak on the living city. Their goal is simply to relegate supporters of a humanistic approach to the city (and the appropriate architecture for accomplishing this) to the margins of the cultural debate, and above all to exclude the humanist architects/urbanists from professional and university jobs.

It is not only the fault of architects. Politicians, mayors, and administrators who have no generative urbanist ideas mask this insufficiency by copying fashionable architecture. When city officials look through an architectural magazine that illustrates all the marvels of the latest "star" architecture, they dream of a single coup that will resolve their problems: building something that will make them famous. For precisely this reason, a mayor will hire a celebrity architect (not even an urbanist) to design a new region of the city. An irresponsible game of urban planning has given origin to entire quarters that are today recognized as being places of urban degradation, of the lowest architectonic quality and quality of social cohabitation. The political choice of assuring a house at low cost, either to rent or to purchase, for those who do not have housing, follows the choice of a particular architectural and urban planning typology. To reverse the failures of the periphery, we first need to understand the failure of social housing.

## 17.10 Spatio-Temporal Scale Jumps and Their Implications

The dramatic consequences of housing billions of people loom ahead for humanity. Such a problem can be sized up only if the magnitude of the processes of urbanization in developing countries is taken into account. It is literally billions of new urbanities, most of them poor, which are expected to be sheltered in some way in global cities in the next few decades. Nothing "strange" is happening, but the administrative answer to that massive pressure is, and will be, given in orthodox modernistic fashion, which will ultimately lead to environmental and social failures of historical relevance for the planet as a whole.

The reasons for which our current practices are increasingly negative towards the creation of human space are of course rooted in attitudes of contemporary society. To cite the most structural ones, we have witnessed a collapse of processes in both space and time, a collapse that takes the form of huge jumps in scale. Formal urban development is now led by extremely large interventions of extremely powerful public/private subjects, resulting in a jump in the spatial scale. Moreover, the urbanization process now occurs in an extremely short pace of time, which is a jump in the temporal scale. The combined effect of these two collapses makes it impossible for us to hope for a restoration of the traditional mechanisms of city evolution for the contemporary city, while at the same time setting the agenda of an entirely new role for city planning. City planning is therefore ready for a profound renovation of its conceptual/technical toolbox; this renovation might be so radical that a new discipline will emerge. Because the characteristics of such a discipline must be to recognize and put into place structural *drivers of urban self-organization* instead of structural and super-structural *programs of urban organization*, we might well call it "urban seeding" instead of "urban planning".

It should be pointed out that modernist planners, after all, simply followed the visual modernist vocabulary, which, in wishing to disown traditional architectural and urban forms, insisted precisely upon unnatural scale jumps. This is the "anti-fractal" philosophy that turned into a complete design ideology. The tight relation between political ideology and the ideology of the architect's professional world cannot be denied. Regardless of any design ideology, human beings are endowed with sensory mechanisms that recognize unnatural scale jumps (Kellert et al. 2008). Therefore, as pointed out elsewhere, the elimination of architectural ornament does indeed have major and unsuspected consequences for the dehumanization of cities on the urban scale (Salingaros and Masden 2008).

## 17.11 Spontaneous Settlements: What We Can Learn from Them

Looking at any image of obviously spontaneous settlements, as in a favela, we are able to observe that the basic urban fabric works far better than that planned by architects. Apart from its social problems, a favela creates urbanism that is understandable in a very simple way. It is not visually simple, unlike social housing or entire quarters of private buildings planned by architects and cited in books and architectural reviews. Those turn out to be too simple to contain human complexity. Certainly, the houses in a favela are often poorly made and there is a total lack of buildings for collective and public functions. Most serious is the absence of even minimal services. All of these are essential to define a city. Nevertheless, this is normal because we are not in the presence of an autonomous society, but instead a collection of marginalized citizens trying to survive outside an organized and structured society.

The collective intelligence responsible for informal settlements is no longer present in more technologically advanced and richer societies. Making a great simplification, we believe that the growth of individual freedom has triggered the increasing autonomy of individuals from their social group, with the unavoidable loss of some key values. Generally this is a positive transformation for society, but one of its negative consequences is the end of the type of city founded upon a shared collective vision. People replaced their intuitive sense of connecting to each other and to a common built environment, with an exclusively individual conception of the world. They detached themselves and manipulated the form of the built environment to enhance their isolation.

The overwhelming majority of people (architects aside) are not satisfied with the contemporary city. Does this tell us something? If historical centers have the highest real estate values, this gives us a motivation to do something. Connections and the relationship to our environment are really personal and physical relations, of which human beings have a basic need. It is not possible to substitute such connections with contacts from the web, which becomes rather an additional resource, one more way to meet and know other people in a different manner. The web is not an alternative to physical and sexual contact that human beings need and cannot do without. Today's young people chat and blog on the internet, exchanging their feelings and sentiments, but in the end, they meet each other in physical places, both old and new.

Our study on social housing (Salingaros et al. 2006) reveals an incessant battle between people trying to create their own living environment, and government agencies determined to impose their own "geometry of power". That geometry is indeed tied to the modernist ideology of industrial power, and which we now know precludes living urban fabric. And yet, governments of all political persuasions are obsessed with images of that formalistic geometry. They are insistent upon stamping out self-built human-scale urbanism, because that encourages the non-mechanical qualities of human beings who have not yet lost their humanity.

Countless examples of government-built social housing as residential towers have failed miserably. Either the people refused to leave their owner-built shanties (and the rich social networks of the favela), or they were forced to move to the new towers after their shacks were bulldozed. In the latter case, the residents eventually ended up destroying the towers as much as they physically could, expressing their hatred at this prison-like typology. The only success – but one that is denied by most governments and agencies – is the gradual upgrading of the favela, accomplished by the residents themselves with materials and infrastructure provided by the government or other agency.

It is nowadays rare to find the insertion of the appropriate urban structures on every human scale that we know to be necessary for encouraging and containing urban life. The aim and visions of modernist architects with respect to the form and substance of buildings leads to the imposition of an "originality" that is alien to human beings. Even early post-war housing settlements tried to use the high-tech materials of their day, and encouraged (instead of trying to tame) the intrinsic gigantism of such structures. Within the urban context (in contrast to, say, a petrochemical plant), the modernist claim that it is necessary to abandon traditional forms and structures falls apart. In reality, the search for novel architectural forms in most cases turns in on itself, and tends to ignore and condemn its context. It thus becomes a purely promotional exercise for the designer, where the architect becomes the end instead of the means.

It seems that what has driven social housing is less consumer demand than vested commercial interests that live on images, and above all a discipline that has lost any contact with reality. It ignores the reality of people who really wish to live in traditional cities and buildings, and that ask precisely for such environments for the past 60 years. People don't ask for a degraded and alienating industrial environment. That part of the discipline that is exquisitely technical, which is concerned with easy to determine standards, necessarily focuses on simplistic geometries for necessary city components. It neglects the basis of past experience, and avoids taking that older knowledge forward into difficult forecasts in a society in constant change. Urbanism now has to make a qualitative jump, and, like a genetic mutation, the correct technical analysis must now become the generating criterion of the project.

In the great historical cities this picture is surely more dramatic than elsewhere. Some urbanists point their finger at "the dinosaurs of popular housing blocks", which have failed because of the concentration of people and weak or non-existent services, forcing social homogenization upon the resident population. Politicians now speak of creating a true social mix: proposing the elimination of single use and single destination in new buildings, in order to make it possible to mix social functions such as work, residences, offices, light industry, handicrafts, students, etc. Despite all these nice phrases pointing out the social failure of the modernist city model while promising change, the action for remedy is caught up in politics and ideology. The reason for lack of progress is an entrenched formalist thinking that is still being taught in our universities, and legislated into current planning regulations. If politicians wish to generate living cities, then their first priority is to scrap all the modernist planning laws that prevent the construction of living urban fabric. The second priority is to learn from the spontaneous geometry of informal settlements (Salingaros et al. 2006).

We come back to the metaphor of the city-as-organism. A city behaves as an organism, or it behaves according to a plan, authoritarian in its very nature, in which someone decides for everyone else and assigns functions, tasks, and roles. The most totalitarian power an institution can wield over people is to tell them exactly where they should live, where they can or cannot walk, exactly how large (or small) a space they are allowed to inhabit, etc. Governments propose a "new great synthesis", without admitting that human-scale solutions can already be found within the traditional city. Meanwhile, the city-as-machine continues to grow more inhuman and to proliferate around the globe. The most popular urban model today, constructed in the best of cases with all of its pieces put in their proper place, as computed from the modernist conception of a city, still doesn't function.

## 17.12 Conclusion

We have inherited a rich variety of invariant patterns for urban structures. These have changed very little from what we see in traditional settlements, so we can apply those typologies to generate *living urban structure* today, i.e., a type of urban fabric that fosters the informal human exchanges which generate "life between buildings" (Gehl 1996). Living urban fabric evolves with time through countless unplanned and unpredictable grass-roots contributions by citizens and social actors at all scales. The problem is that most architects and planners, influenced by decades of anti-traditionalist practices, have forgotten the morphology of such living urban structure. In fact, the discipline of architecture and urban planning itself as we know it needs to be substantially re-framed in a new "urban seeding" approach, so as to embrace the idea of self-organization as a key feature of successful urban spaces. Lacking those insights, whenever arrogant modernist and "star" architects try to design urban fabric today, it turns out to be dead. Those who argue most fanatically against the use of traditional forms are ironically the same persons who defend the deadening sameness of the sterile modernist forms they wish to apply for every case and for every locality. The diversity and adaptivity of traditional typologies is our guarantee against homogenization. There exist common bases, biological and perceptual, for any architecture and urbanism, and every human being can verify if those are adaptable to our living environment. We change the forms, following changing cultural traditions and needs, in which the common rules of behavior are manifested. What remains invariant, however, is the biological perception common to all people.

**Acknowledgments** This essay began as a series of online conversations between one of us (PP) and Vilma Torselli, who we thank for her inspiring comments.

## References

Alexander, C. (2001–2005) *The nature of order*, Books 1–4, *Center for Environmental Structure*. (Berkeley, CA, USA). Book 1: *The phenomenon of life*, 2001; Book 2: *The process of creating life*, 2002; Book 3: *A vision of a living world*, 2005; Book 4: *The luminous ground*, 2004

Alexander, C., Ishikawa, S., Silverstein, M., Jacobson, M., Fiksdahl-King, I. & Angel S. (1977). *A pattern language*. (New York: Oxford University Press)

Alexander, C., Neis, H., Anninou, A. & King, I. (1987). *A new theory of urban design*. (New York: Oxford University Press)

Batty, M. (2002). A decade of GIS: what next? *Environment and Planning B: Planning and Design*, 29, 157–158

Batty, M. (2008) Whither network science? *Environment and Planning B: Planning and Design*, 35, 569–571

Batty, M. & Longley, P.A. (1994). *Fractal cities: a geometry of form and function*. (London: Academic Press)

Bertuglia C.S. & Staricco, L. (2000). Complessità, Autoorganizzazione, Città, Franco Angeli. Milan, Italy (in Italian)

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Wang, D. (2006). Complex networks: structure and dynamics. *Physics reports*, 424, 175–308

Campagna M. (Ed.) (2005) *GIS for sustainable development*. (London: CRC Press)

Caniggia, G. & Maffei, G.L. (2001). *Architectural composition and building typology*. (Florence, Italy: Alinea Editrice)

Carver, S. (2001). Public participation using web-based GIS. *Environment and planning B: Planning and design*, 28, 803–804

Donnay, J., Barnsley, M.J. & Longley, P.A, (Eds.) (2001). *Remote sensing and urban analysis*. (London, UK: Taylor and Francis)

Faust, N.L. (1995). The virtual reality of GIS. *Environment and Planning B: Planning and Design*, 22, 257–268

Gehl, J. (1996). *Life between buildings: using public space*. (Copenhagen, Denmark: Arkitektens Forlag)

Golicnik, B. (2007). GIS behavior mapping for provision of interactive empirical knowledge. In K. Thwaites, O. Romice, S. Porta & M. Greaves (Eds.), *Urban sustainability through environmental design: Approaches to time, people and place responsive urban spaces* (pp. 136–140). (Abingdon: Routledge, Milton Park)

Granovetter, M (1973). The strength of weak ties. *American Journal of Sociology*, 78, 1360–1380

Hall, R. (2008). Planning for walkable streets. In T. Haas (Ed.), *New urbanism & beyond: Designing cities for the future* (pp. 153–157). (New York: Rizzoli)

Harris, A. (2008). Complex systems theories for understanding. [electronic version]. Retrieved from www.iaacblog.com/complexsystems/wp-content/uploads/2008/06/final-complex syte-ms.pdf

Jacobs, A.B., Macdonald, E. & Rofè, Y. (2003). *The boulevard book: history, evolution, design of multiway boulevards*. (Cambridge, MA: The MIT Press)

Jiang, B. & Claramunt, C. (2002). Integration of space syntax into GIS: new perspectives for urban morphology. *Transactions in GIS*, 6, 295–309

Kellert, S.R., Heerwagen, J. & Mador, M. (2008). *Biophilic design: the theory, science and practice of bringing buildings to life*. (New York: John Wiley)

Krier, L. (1998). *Architecture: choice or fate*. (Windsor: Andreas Papadakis Publisher).

Lynch, K. (1981). *A theory of good city form*. (Cambridge, MA: MIT Press)

Marzot, N. (2002). The study of urban form in Italy. *Urban Morphology*, 6(2), 59–73

Oliva-i-Casas, J. (2001). *Confusion in urban design*. (Amsterdam, Holland: Techne Press)

Porta, S. (1999). The community and public spaces. *Futures*, 31, 437–456

Porta, S. (2008). The world's longest mistake? Sustainable urban design and the renovation of social housing estates. *Urban Scrawl*, 2, 18–21, Retrieved from humanspacelab.com/UploadedFiles/PA2008_05.pdf

Porta, S., Latora, V. & Crucitti, P. (2006). The network analysis of urban streets: a primal approach, *Environment and Planning B: Planning and Design*, 33, 705–725

Salingaros, N.A. (2005). *Principles of urban structure*. (Amsterdam, Holland: Techne Press)

Salingaros, N.A., Brain, D., Duany, A.M., Mehaffy, M.W. & Philibert-Petit, E. (2006). Favelas and social housing: the urbanism of self-organization. Presented at the Brazilian and Ibero-American Congress on Social Housing, Florianópolis, Brazil, 2006

Salingaros, N.A. & Masden K.G. (2008). Neuroscience, the natural environment, and building design. In S.R. Kellert, J. Heerwagen & M. Mador (Eds.), *Biophilic design: the theory, science and practice of bringing buildings to life* (pp. 59–83). (New York: John Wiley)

Schocken, H. (2003). Intimate anonymity: breaking the code of the urban genome. *INTBAU Essays*, 1(5), Retrieved from www.intbau.org/essay5.htm

Senseable City Lab (2006) Real time rome, massachusetts institute of technology, Cambridge MA, USA. http://senseable.mit.edu/realtimerome/

Steil, L., Salingaros, N.A. & Mehaffy, M.W. (2008). Growing sustainable suburbs: an incremental strategy for reconstructing sprawl. In T. Haas (Ed.), *New urbanism & beyond: designing cities for the future* (pp. 262–274). (New York: Rizzoli)

Thwaites, K. & Simkins, I. (2007). *Experiential landscape: an approach to people, place and space*. (Abingdon: Routledge, Milton Park)

Wagner, D.F. (1997). Cellular automata and geographic information systems. *Environment and planning B: Planning and design,* 24, 219–234

Whyte, W.H. (1988). *City: rediscovering the center*. (New York: Doubleday)

# Chapter 18
# Analyzing Spatial Patterns of Late-Stage Breast Cancer in Chicago Region: A Modified Scale-Space Clustering Approach

**Lan Mu, Fahui Wang, and Sara McLafferty**

**Abstract** Breast cancer has ranked highest in cancer incidence in Illinois for years. Detection at an early stage helps cancer patients live longer and maintain a better quality of life. Previous research identified two groups of potential risk factors for late-stage breast cancer diagnosis: spatial factors including access to healthcare, and nonspatial factors including socioeconomic and demographic characteristics. Literature suggests that risk factors for late-stage diagnosis behave differently between rural and urban areas, and research needs to separate study areas into various geographic settings. This chapter focuses on an urban area of six counties in Chicago region, and examines possible associations between several risk factors and late-stage breast cancer diagnosis. Based on the data at the zip code level, the study uses the modified scale-space clustering (MSSC) method to form various geographic areas. The MSSC considers both attribute similarity and spatial adjacency while minimizing the loss of information in the clustering process. Therefore, area units defined by the method are more coherent in terms of attribute and spatial closeness for research than geopolitical units. For instance, health literature often suggests the need to separate a study area into urban, suburban and rural areas or even finer-grained area classifications. The method can be used to generate more meaningful geographic divisions than traditional schemes. The analysis results are generally consistent across multiple area units, demonstrating the effectiveness of the method in mitigating the modifiable areal unit problem (MAUP).

**Keywords** Cancer clustering · Late-stage breast cancer diagnosis · GIS · Modified scale-space clustering (MSSC)

L. Mu (✉)
Department of Geography, University of Georgia, Athens, GA 30602, USA
e-mail: mu@UGA.EDU

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_18,

## 18.1 Late-Stage Breast Cancer and Risk Factors

Breast cancer is the most common cancer among women. Early diagnosis of breast cancer by mammography screening has become a critical way to reduce mortality (Sheehan and DeChello 2005). The importance of early diagnosis has led to an emphasis on understanding "late-stage" cancer risk. Late-stage breast cancer is cancer that is first detected after it has spread to distant tissues or organs. Patients whose cancer is diagnosed at a late stage have higher risks of morbidity and mortality from the disease than do other patients. How late-stage risk varies among people and places is a critically important topic for research investigation. Studies show that women without adequate accessibility to timely mammography screening are more likely to develop late-stage breast cancer (Mandelblatt et al. 1991). Racial and ethnic disparities are also significant: the chance of African American women dying from breast cancer within five years of diagnosis is twice as great as for Caucasian women, and the chance of Hispanic women dying is 1.5 times greater than that for Caucasian women (Eley et al. 1994). These differences in mortality in part reflect differences in late-stage diagnosis.

In addition to these socio-demographic disparities, late-stage breast cancer risk has been shown to vary geographically. Some studies find significant rural-urban disparities, reflecting barriers such as poor geographical access to primary care and cancer screening and treatment services, lack of insurance, and lack of knowledge about screening guidelines (Coughlin et al. 2002). At the intra-urban scale, rates of late-stage diagnosis are much higher in low-income neighborhoods than in more affluent neighborhoods (Barry and Breen 2005). A recent study (Wang et al. 2008) indicates that not only does late-stage risk vary geographically, but also the social and spatial correlates of late-stage risk vary from place to place. In Wang et al. (2008), spatial accessibility of primary care was statistically significant in affecting late-stage risk outside the city of Chicago, but not in the city. The dense availability of health care facilities in Chicago made spatial barriers less important there than in the more rural and suburban parts of the state. These studies highlight the need to understand geographic variation in late-stage breast cancer risk by focusing on data for small geographic areas. When diverse areas are pooled together, important geographic differences may be masked.

Although investigating small area variations in cancer risk is highly desirable from a research and policy perspective, small area analysis often leads to challenging statistical problems such as the small numbers problem (Diehr 1984). The small numbers problem refers to the high variance associated with data for small areas. This high variance means that disease rates for small areas are unstable and unreliable. This research describes how a new tool, the *modified scale-space clustering (MSSC) method* (Mu and Wang 2008), can be used to enhance geographic investigations of late-stage cancer risk over small areas. The focus of this paper is to construct larger geographic areas from zip code areas using the MSSC in order to stabilize the late-stage cancer rates. We then examine whether the findings on late-stage breast cancer risks are consistent over multiple geographical scales.

### 18.1.1 Data and Methods

This research focuses on a large urban area: comprising six counties (Cook, Will, DuPage, Lake, Kane and McHenry) in the Chicago metropolitan region (Chicago region) (Fig. 18.1). The study area includes 278 zip code areas. All data were compiled from a previous study (Wang et al. 2008). Breast cancer data were extracted from cancer incidences from the Illinois State Cancer Registry (ISCR) 1998–2000, Illinois Department of Public Health. The ISCR dataset has the records of all Illinois patients, including those diagnosed in neighboring states, such as Missouri and Wisconsin. Lehnerr and Havener (2002) estimated that the case ascertainment reached 98% completeness. The dataset includes individual records of cancer incidence, geocoded to zip code areas, with variables such as cancer type, age group, sex, race, diagnosis stage, and year. Privacy and confidentiality prevent access to relevant cancer data from areas smaller than zip codes. Cases of late-stage cancer are defined as cases with cancer stage in the range from 2 to 7 at time of diagnosis.

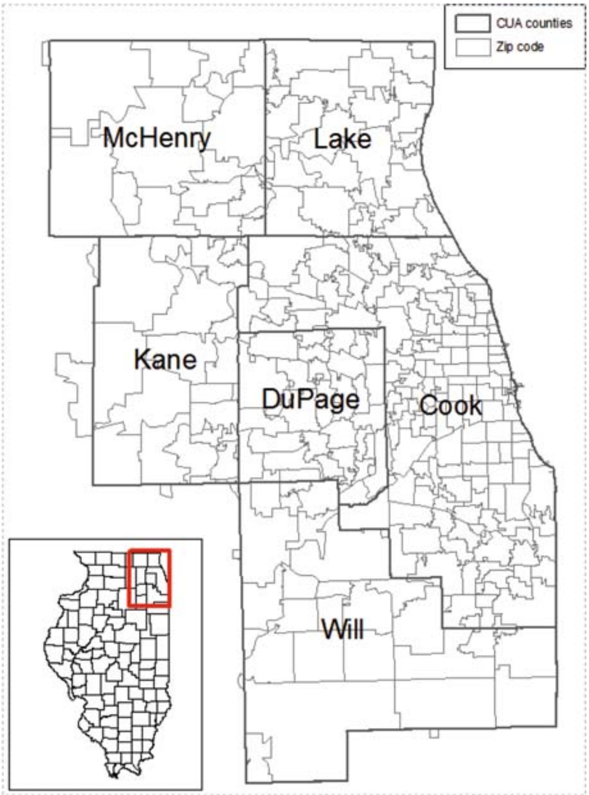



**Fig. 18.1** Six-county Chicago region

Cancer risk factors can be grouped into two broad categories: (1) spatial factors such as geographic access to primary care and travel time to the nearest mammography facility and (2) nonspatial factors such as socioeconomic and demographic characteristics of individuals and communities. As discussed earlier, both types of factors strongly affect the risk of late-stage breast cancer. To investigate the association between late stage cancer and these two types of factors, we construct spatial variables and three non-spatial variables

Geographic access to primary care was estimated using the *two-step floating catchment area method (2SFCA)* (Luo and Wang 2003; Wang 2006). A high value for this geographic access measure represents a high ratio of supply to demand – a large number of primary care physicians in the local area compared to demand. Travel times from the centroid of each zip code to the nearest mammography facility were estimated based on real-world road networks accounting for the type of road and adjusting for typically lower travel speeds in densely populated urban areas (Luo and Wang 2003).

To measure non-spatial factors, socioeconomic and demographic variables such as poverty, education and housing characteristics were obtained from the 2000 Census for zip code areas. These 11 variables were consolidated into three factors using principal components analysis (Table 18.1): socioeconomic disadvantages, sociocultural barriers, and high health care needs.

As mentioned earlier, a problem in analyzing health data for small areas like zip codes is the high variability of data due to the small numbers problem. One way to address the small numbers problem is to aggregate the data into larger areas that contain larger numbers of events. This spatial aggregation is a

**Table 18.1** Consolidating nonspatial risk factors by factor analysis

| Contents | Socio-economic disadvantages | Socio-cultural barriers | Healthcare needs |
|---|---|---|---|
| Female-headed households (%) | 0.9089 | –0.0058 | 0.0504 |
| Population in poverty (%) | 0.8662 | 0.1642 | 0.2405 |
| Non-white minorities (%) | 0.8481 | 0.2153 | –0.0153 |
| Households w/o vehicles (%) | 0.8231 | 0.1905 | 0.2699 |
| Home ownership (%) | –0.6686 | –0.3362 | –0.3922 |
| Housing units lacking basic amenities (%) | 0.4278 | 0.2703 | –0.0323 |
| Households with linguistic isolation (%) | –0.0479 | 0.9561 | 0.0164 |
| Households with >1 person per room (%) | 0.4464 | 0.7966 | –0.0631 |
| Population w/o high-school diploma (%) | 0.5800 | 0.6406 | 0.1219 |
| Population with high healthcare needs (%) | 0.0316 | –0.1050 | 0.9186 |
| Median income ($) | –0.5491 | –0.2053 | –0.5605 |
| % of variance explained by each factor | 54.08 | 28.09 | 17.83 |

common strategy; however, it is typically done in a very ad hoc way by relying on larger political units such as counties or states. We propose instead that zip codes be grouped into larger areas according to their social and spatial characteristics. A new method, MSSC, is used to perform the grouping process. The MSSC groups geographic units into clusters based on their attribute (socioeconomic measures) similarity and spatial adjacency. In the clustering process, it seeks to minimize loss of information (e.g., measured by the entropy) in aggregation of areas. By doing so, the method balances two critical criteria in regionalization: within-area homogeneity and spatial compactness. After small areas have been merged together, we can analyze spatial patterns of late-stage cancer and the associations with spatial and non-spatial factors, with less concern about the effects of small numbers.

## 18.2 Modified Scale-Space Clustering Method (MSSC)

Scale-space theory is a framework to analyze image or spatial data structures across different scales. Cartographic generalization is among many ad hoc techniques that contributed to the early development of the theory (Koenderink 1994). Key concepts of the scale-space theory were established in early 1980s with the term *scale-space* coined by Witkin (1983) and *spurious resolution* by Koenderink (1984). The theory has been developed and applied in computer vision, thermodynamics and mathematical morphology, and has been applied in image processing, spatial imagery data mining, classification of land uses, identification of seismic belts, pattern recognition, and others (Ciucu et al. 2003; Wang et al. 2005).

A major application of the scale-space theory is as a spatial clustering method. Spatial clustering methods are often challenged by difficult-to-meet data distribution criteria (Fraley and Raftery 1998), deficient cluster validity criteria (Wong 1993), random results caused by lack of a priori knowledge, and uncertainty derived from initialization, overlapping, and variability in the shape, size and density of clusters (Chakravarthy and Ghosh 1996; Ciucu et al. 2003; Gath and Geva 1989; Horn 2001; Leung et al. 2000; Miller and Rose 1996; Rose et al. 1990; Wilson and Spann 1990). Scale-space theory has been employed to address those challenges and to improve spatial clustering methods (Ciucu et al. 2003; Leung et al.; Wong 1993). Use of scale-space theory to explore socioeconomic patterns has drawn geographers' attention (Luo et al. 2002; Mu and Wang 2008). Mu and Wang (2008) developed a modified scale-space clustering (MSSC) method that accounts for both attribute homogeneity and spatial contiguity for zone-based data such as census tracts. The MSSC method is implemented in a GIS program. Additionally, a metric, level of convergence (LC), is developed to measure how soon two objects are merged into one cluster and thus how "close" they are in terms of both attribute similarity and spatial proximity. The LC indicator provides a quantitative evaluation tool for districting and boundary changes.

The MSSC method is created based on the melting algorithm (Wong 1993) and the blurring algorithm (Koenderink 1984; Leung et al. 2000; Witkin 1983; Zahn 1971). The melting algorithm simulates the melting process from solid to liquid to generate hierarchical clusters. Thermodynamical principles are applied to identify local cluster centers measured by lowest free energy. The blurring algorithm is developed from the response of human eyes when looking at a picture from close-up to far away. Because of the change of distance, the light blobs on the picture (seen by eyes) change as well, mimicking the hierarchical clustering process.

Blurring algorithm can be represented mathematically on a two-dimensional plane $R^2$.

$$P(x,\sigma) = p(x)^{*} g(x,\sigma) = \int p(x-y)\frac{1}{\sigma^2 2\pi} e^{-\frac{\|y\|^2}{2\sigma^2}} dy \tag{18.1}$$

$$\begin{cases} \frac{dx}{dt} = \nabla_x p(x) \\ x(0) = x_0 \end{cases} \tag{18.2}$$

where $P(x,\sigma)$ can characterize the scale-space image, $\sigma$ is the scale parameter, $x$ is a datum on the image, $p(x)$ is the image distribution at $x$, and $g(x,\sigma) = \frac{1}{\sigma 2\pi^2} e^{-\frac{\|x\|^2}{2\sigma^2}}$ is the Gaussian function.

The blurring process assigns cluster membership to each blob. Each blurred region (cluster) has a light blob local maximum $y$as the cluster center, and the cluster can be represented as $\{x_0 \in R^2 : \lim_{t\to\infty} x(t,x_0) = y\}$. The clustering solution can be derived from the gradient dynamic system (18.2). Blurring algorithms for nested and nonnested hierarchical clustering can be expressed through a series of iterations (Leung et al. 2000).

For polygon data, Equations (18.1) and (18.2) need to be modified. One way to capture the steepest descent criterion is to maximize correlation coefficient between adjacent polygon units across their multiple attributes (Luo et al. 2002). We propose to melt each spatial object (polygon) to its most similar neighbor, measured by their attribute distance. The *minimum-distance criterion* is designed to search for the units with the most similar attributes so that units can merge to cluster centers the fastest (Mu and Wang 2008). Equation (18.2) is modified as (18.3) and (18.4),

$$D_{ij} = \sum_t (x_{it} - x_{jt})^2 \tag{18.3}$$

$$D_{ik} = \min_{j\in \mathrm{J}} D_{ij} \tag{18.4}$$

where an object $i$ has $t$ attributes standardized as ($x_{i1}$, ..., $x_{it}$); its adjacent objects $j$ ($j \in$ J, and J $= \{1, 2, ..., m\}$) have attributes also standardized as ($x_{j1}$, ..., $x_{jt}$); and $D_{ij}$is the attribute distance between $i$ and $j$.

One may also account for relative importance of attributes in Equation (18.3) by multiplying each attribute by its corresponding weight $w_t$, such as

$$D_{ij} = \sum_t w_t (x_{it} - x_{jt})^2 \tag{18.5}$$

The MSSC is an unsupervised hierarchical classification method. A program is written and used to run the clustering process multiple rounds until all objects are grouped into one cluster.

Based on the clusters generated from MSSC, an index, *level of convergence* (LC), is developed to measure the extent to which two original objects are separated. The index is

$$LC_{ij} = 1 - \frac{r_{ij}}{R} \tag{18.6}$$

where $LC_{ij}$ is the level of convergence between objects $i$ and $j$, $r_{ij}$ is the clustering round when two objects $i$ and $j$ are first melted into one cluster, and $R$ is the total number of rounds. LC index indicates how soon two objects are grouped into one cluster and thus how "close" they are in terms of both attributive and spatial proximity.

## 18.3 Analyzing Late-Stage Breast Cancer Data in Chicago with the MSSC Method

In this section, we demonstrate how to apply the MSSC method to analysis of cancer data in the Chicago region and derive important spatial and attributive indications. Results from the MSSC method include local maxima and level of clustering convergence. Spatial regression is used to further examine the implication of the method.

### *18.3.1 Clustering Process by the MSSC Method*

The three factors in Table 18.1 capture the socioeconomic structure of zip codes, and are used as the cluster grouping variables ($x_{it}$ and $x_{jt}$ in Equation (18.5)). Their percentages of variance explained (0.54, 0.28 and 0.18) are the weights ($w_t$ in Equation (18.5)). Two other variables, access to primary care and travel time to mammography facility, are not used in identifying clusters but are carried over to each round of clustering. Population is the weight variable for aggregating values to the next round. The rate of late-stage breast cancer for the new "grouped" regions is calculated for each round of clusters. The clustering method is performed in nine rounds till all units merged to one cluster. The clusters are shown in Fig. 18.2. There are 278 zip code areas in the study area, and the numbers of clusters in the nine clustering rounds are 108, 45, 19, 11, 7, 5, 4, 2 and 1.

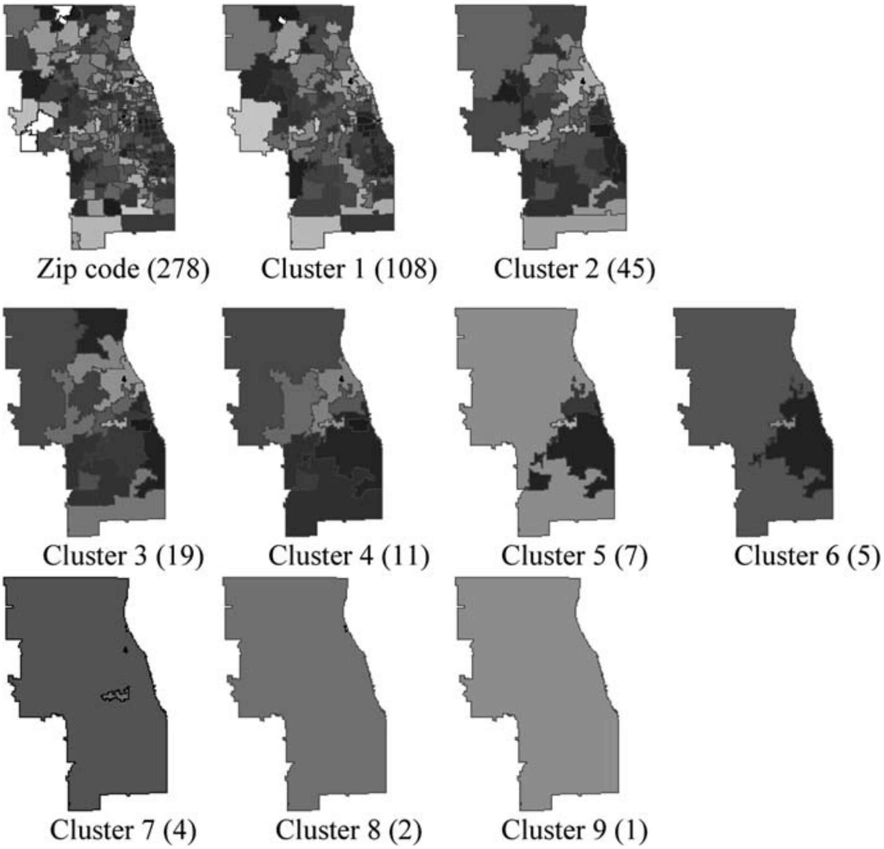



**Fig. 18.2** Clustering process based on the factors in the Chicago region

### *18.3.2 Visualizing Local Maxima*

To implement MSSC, we first establish a link between each object and its most similar adjacent object using the minimum-distance criterion (Equations (18.3) and (18.4)). The link's direction goes from the lower to the higher aggregate attribute score, and the attribute score is calculated using the same set of variables in the minimum distance calculation above. We then identify local minima and local maxima. A local minimum is an object with all directional links pointing towards other objects, and thus has the lowest aggregated attribute score among surrounding objects. A local maximum is an object with all directional links pointing towards it, and thus has the highest aggregated attribute score among surrounding objects. Clustering begins with a local maximum, follows the link direction to search and group objects along the way, and ends until a local minimum is reached.

As explained in the previous section, one important outcome from the MSSC method is local maxima at each clustering stage. Clustering usually is based on explanatory variables (i.e., socioeconomic factors affecting the late-stage diagnosis), and thus local maxima are local peaks of aggregated factor scores in this case study (map on the right in Fig. 18.3). However, it may be also useful to display local maxima that are based on the subject variable (i.e., late-stage breast cancer rate). By doing so, spatial patterns between the two can be then compared to examine how the two are related and whether the factors are valid late-stage breast cancer risks in the Chicago region. The map on the left in Fig. 18.3 shows local maxima of late-stage breast cancer rate (per 100 breast cancer cases), which is used as the only variable ($x$'s in Equation (18.3)) in MSSC. The map on the right in Fig. 18.3 shows local maxima based on all three factor scores. Note that local maxima are observations surrounded by ones of smaller values, which capture the spatial structure of data better than the plain contour lines based on the original values.

The two maps in Fig. 18.3 show similar overall spatial patterns with higher densities of local maxima in the downtown Chicago area than suburban areas. Visually, the local maxima of the late-stage breast cancer rates are less concentrated than those of the socioeconomic factor scores. There are a total of 81 local maxima of late-stage breast cancer rates, and 56% of them are in Cook County. For factor scores, there are a total of 83 local maxima, and more than 61% of them are in the central Cook County.

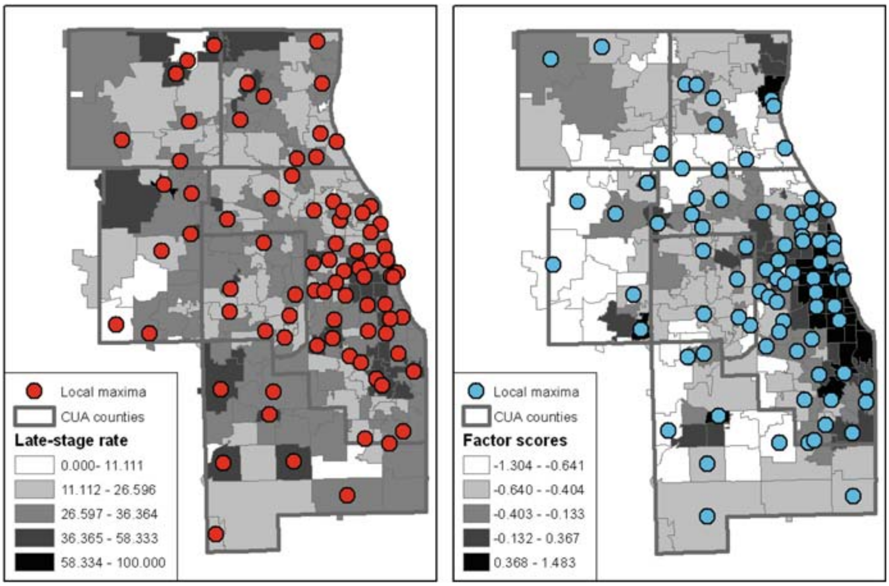



**Fig. 18.3** Local maxima of late-stage breast cancer and factor scores

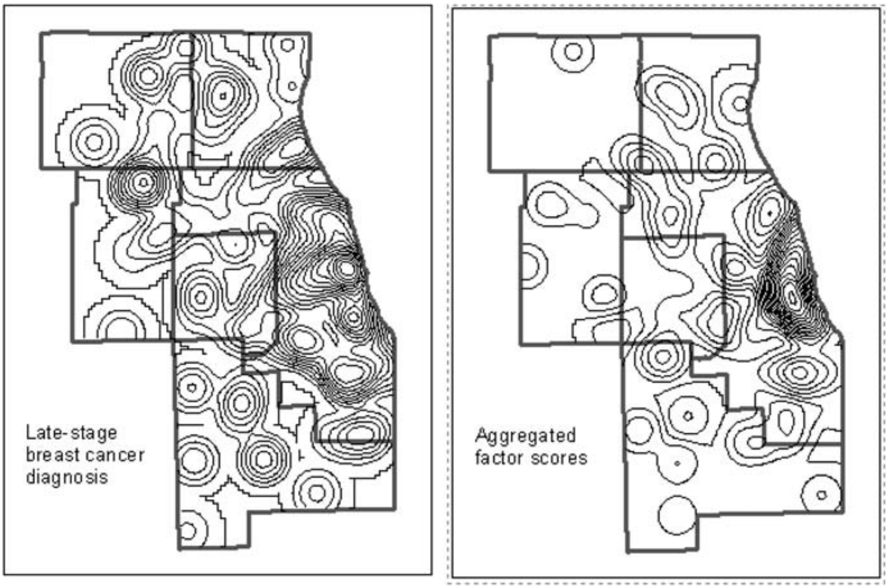



**Fig. 18.4** Contours based on local maxima of late-stage breast cancer and factor scores

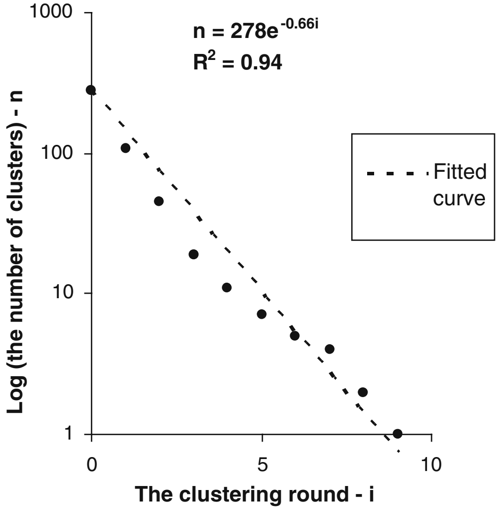
We then further explore the hot spots of late-stage breast cancer diagnosis and aggregated socioeconomic factor scores by generating contour lines from their local maxima (Fig. 18.4). The same number (20) of intervals is used in the two maps to be comparable. The contour maps confirm our previous observation that the spatial pattern of late-stage breast cancer rate is in general consistent with that of the factor scores, however, there are more local peaks of late-stage cancer than those of the factors.

### *18.3.3 Level of Convergence*

From zip code area to the ninth round of cluster, the clusters merge very quickly with a converging gradient of 0.66 (Fig. 18.5) according to Equation (18.6). A level of convergence (LC) map for each zip code area can be generated to examine how soon two zip code areas are grouped into one cluster, and thus how "close" they are. As the MSSC method considers both attribute similarity and spatial adjacency in the process, the LC value represents the closeness of area objects in terms of socioeconomic and spatial proximity.

Take zip code 60624 as an example (Fig. 18.6). The zip code area is in Garfield Park Community, and its "close" regions in descending order are its immediate neighbors (e.g., Austin, LC = 0.89), downtown area excluding Loop (e.g., Lincoln Park, LC = 0.78), Loop and the southwest (e.g. West Lawn, LC = 0.67), the southeast (e.g., Chatham, LC = 0.56), the south (e.g., South

**Fig. 18.5** Converging effect



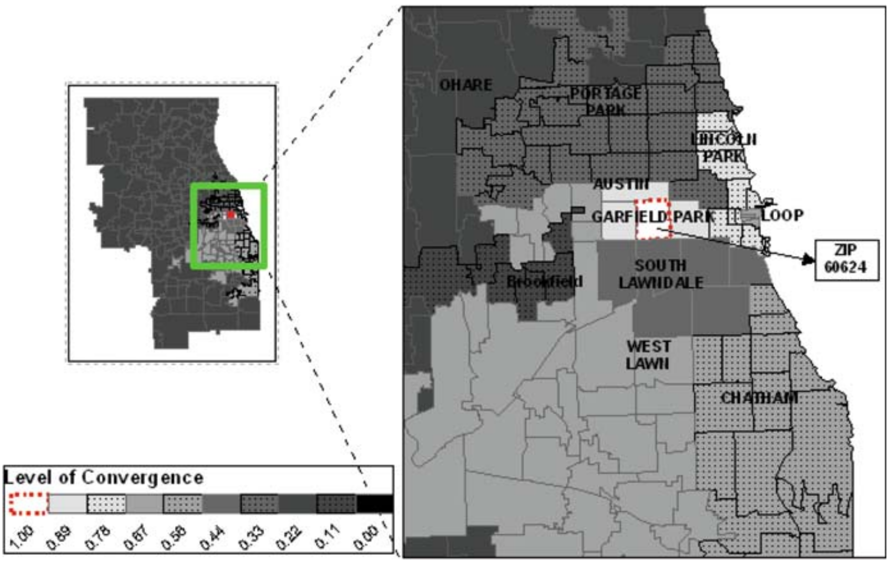




**Fig. 18.6** Level of convergence (LC) for an exemplary zip code 60624

Lawndale, LC = 0.44), the north (e.g., Portage Park, LC = 0.33), suburbs including O'Hare (LC = 0.22), a strip just to the west of Chicago (e.g., Brookfield, LC = 0.11), and a single zip code area in Glenview. The zip code area in Glenview has 0 breast cancer incidence and does not bear any resemblance to its surrounding areas in terms of socioeconomic structure, and thus is not grouped to any clusters until the last round (not shown on Fig. 18.6, LC = 0.00).

Overall, zip code 60624 shares more similarity with areas to its east and south than areas to its west and north in terms of socioeconomic structure. The clustering process balances spatial adjacency and attribute homogeneity simultaneously: it merges with areas immediately adjacent to the north, west, and east, joins with areas to the southwest, and snaps up areas to the east, and goes back to merge with areas in the Loop, South Lawndale and Brookfield, and finally towards the west and north.

## 18.3.4 Spatial Regression

Another value of MSSC method lies in its assist in choosing spatial regression models. Visually, in the case of late-stage breast cancer in the Chicago region, there is no practical need to continue the MSSC clustering after round three or four, when the numbers of clusters are down to 19 and 11 respectively. At this point, spatial variation is reduced greatly because of the formation of mega clusters (Fig. 18.2). Statistically, we use ordinary least squares (OLS) regression and spatial lag (SL) model to test the cut-off point of clustering. In both models, the dependent variable is the late-stage breast cancer rate, and the five independent variables are the three factors, access to primary care and travel time to the nearest mammography facility.

The *spatial lag model* (Anselin 1988) controls for spatial autocorrelation such as:

$$y = \rho y_{-1} + \sum_{k} b_k \, x_k + e \tag{18.7}$$

Where $y$ is dependent variable, $y$B$_{-1}$B is its spatial lag, $x_k$'s represent independent (explanatory) variables, $e$ is random error, and $\rho$ and $b_k$ are coefficients. The outcome from each round of the MSSC method is fed into OLS and SL using statistical software (Anselin 2005).

Regression results at four geographic scales ranging from zip code areas to third round clusters are presented in Table 18.2.

Three key observations can be made from the table:

1. In Chicago region, geographic access to primary care and travel time to mammography facilities have no significant association with late-stage breast cancer diagnosis. This conclusion holds across all cluster levels. The result is consistent with previous findings where geographic access only matters outside of the city of Chicago (Wang et al. 2008).
2. Among the three factors, factor 1 (socioeconomic disadvantages) and factor 2 (sociocultural barriers) are positively and significantly associated with late-stage breast cancer rate in most cases, but factor 1 is no longer significant after the third round of cluster. Factor 3 (high health care needs) is negatively related to the late-stage rate, however, such a relationship is not significant until higher aggregation level of clusters (round three) are formed.

**Table 18.2** Ordinary least squares (OLS) regression and spatial lag (SL) models

| Independent variables | Zip codes OLS ($n = 278$) | Zip codes SL ($n = 278$) | Cluster 1 SL ($n = 108$ ) | Cluster 2 SL ($n = 45$) | Cluster 3 SL ($n = 19$) |
|---|---|---|---|---|---|
| Intercept | 28.91 (11.83)*** | 29.39 (7.84) *** | 25.49 (5.19) *** | 22.54 (3.25)** | 21.53 (2.29)* |
| Factor1 | 3.80 (3.68)*** | 3.85 (3.66) *** | 3.68 (3.37)*** | 2.82 (2.20)* | 2.20 (1.04) |
| Factor2 | 2.03 (2.17)* | 2.04 (2.20) * | 2.22 (2.40)* | 2.85 (2.51)* | 3.34 (2.56)* |
| Factor3 | –0.02 (–0.02) | –0.01 (–0.01) | –0.37 (–0.35) | –2.04 (–1.72) | –6.95 (–4.39)*** |
| Access to primary care | –0.44 (–0.64) | –0.43 (–0.64) | –0.65 (–0.92) | –0.97 (–1.00) | 0.07 (0.05) |
| Travel to mammography | 0.28 (1.58) | 0.28 (1.61) | 0.22 (1.00) | –0.20 (–0.46) | –1.25 (–1.95) |
| Spatial lag (ρ) | | –0.02 (–0.18) | 0.13 (0.96) | 0.31 (1.71) | 0.38 (1.64) |

Note: The numbers are correlation coefficients with t-values in parentheses; *** significant at 0.001, ** significant at 0.01, * significant at 0.05.

3. The significance level of spatial lag ($y\mathrm{B}_{\text{-}1\mathrm{B}}$) is not significant from zip code areas to higher rounds of clusters, indicating little spatial autocorrelation. Therefore, the advantage of spatial lag model over OLS regression is minimal for this study.

## 18.4 Discussion

The MSSC method groups adjacent areal units so that within-cluster homogeneity is maximized. Shapes of histograms of the original data and clusters can demonstrate the changes in data distribution and therefore reflect the overall homogeneity of clusters (Wang 2005). To compare the within-cluster homogeneity, we calculate the average within-cluster variances of late-stage breast cancer rate from original zip-level to all levels of clusters (Fig. 18.7). Measured by the within-cluster variance, the MSSC method improves the within-cluster homogeneity for cluster levels of 1 to 5 with cluster level 1 as the best grouping solution, and it does not help anymore after level 6. Results from this case study indicate that geographic areas constructed by the method yield a generally consistent relationship between late-stage breast cancer rate and risk factors. This demonstrates the effectiveness of the method in mitigating the modifiable areal unit problem (MAUP) and small numbers problem.

The method also helps answer the question to what extent areas (zip codes) could be grouped with a minimal loss of information. For example, from the clustering process, it is not until the sixth round of cluster that all zip codes in the city of Chicago (red outline in Fig. 18.8) are merged as one cluster. If the

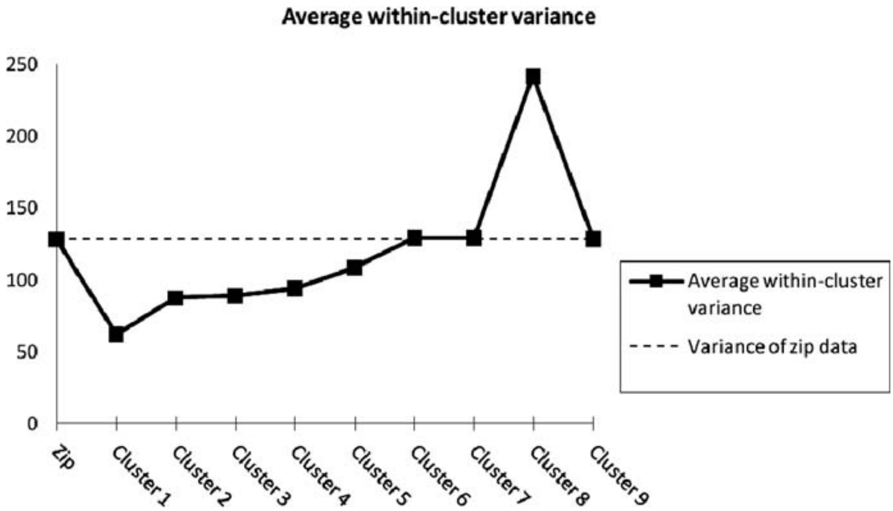



**Fig. 18.7** Average within-cluster variances of different levels of clusters

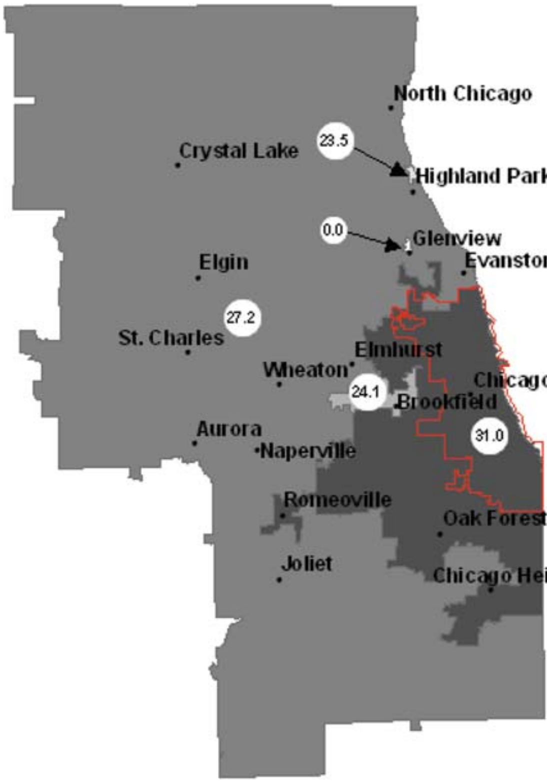



**Fig. 18.8** Five clusters from the sixth round

interest is to compare central city versus suburbia in Chicago region, it is perhaps more meaningful to use Cluster 6 instead of a geopolitical boundary (City of Chicago versus others) as the geographic dividing schemes. The sixth round yields five clusters: the cluster covering all areas in the City of Chicago and beyond ("central city") and four other clusters. This indicates more variations outside of the central city. Average late-stage breast cancer rates vary across these five clusters such as

- City of Chicago and its southwest areas including Oak Forest, Chicago Heights, and Romeoville (31.0%),
- the majority of Chicago suburbs including Evanston, North Chicago, Crystal Lake, Elgin, St. Charles, Wheaton, Elmhurst, Aurora, Naperville, and Joliet (27.2%),
- a small area just to the west of Chicago including Brookfield (24.1%),
- a small area in north suburb just to the north of Highland Park (23.5%), and
- a single zip code area in Glenview that has 0 breast cancer incidence and share no similar socioeconomic/cultural/access natures to its surrounding areas (so it never grouped to any clusters before this round) (0%).

## 18.5 Summary

Stage of cancer at diagnosis – plays a critical role in determining the prognosis of patients. Detection at an early stage helps cancer patients live longer and maintain a better quality of life. Breast cancer has ranked highest in cancer incidence in Illinois for years. Literature suggests that risk factors for late-stage diagnosis behave differently between rural and urban areas, and research needs to separate study areas into various geographic settings. This paper focuses on an urban area of six counties in Chicago region, and examines possible associations between several risk factors and late-stage breast cancer diagnosis.

Based on the data at the zip code level, the study uses the modified scale-space clustering (MSSC) method to form various geographic areas. The analysis results are generally consistent across multiple area units, demonstrating the effectiveness of the method in mitigating the modifiable areal unit problem (MAUP). In the Chicago region, geographic access to primary care and travel time to mammography facilities do not appear to be significant risk factors in late-stage breast cancer diagnosis. Among the three socioeconomic factors, factor 1 (socioeconomic disadvantages) and factor 2 (sociocultural barriers) are positively and significantly associated with late-stage breast cancer rate in most cases. However, factor 1 is no longer significant in higher rounds of clusters.

The MSSC considers both attribute similarity and spatial adjacency while minimizing the loss of information in the clustering process. Therefore, area units defined by the method are more coherent in terms of attribute and spatial closeness for research than geopolitical units. For instance, health literature often suggests the need to separate a study area into urban, suburban and rural areas or even finer-grained area classifications. The method can be used to generate more meaningful geographic divisions than traditional schemes (e.g., the rural-urban commuting areas or RUCA classification; (Hart et al. 2005)).

**Acknowledgments** The second and third authors would like to acknowledge the support of the National Cancer Institute through grant 1-R21-CA114501-01.

## References

Anselin, L. (1988). *Spatial econometrics: methods and models*. (Dordrecht; Boston: Kluwer Academic Publishers)

Anselin, L. (1988). *GeoDa (software)*. Spatial Analysis Laboratory, University of Illinois 2005 [electronic version], retrieved from http://sal.agecon.uiuc.edu/geoda_main.php.

Barry, J. & Breen, N. (2005). The importance of place of residence in predicting late-stage diagnosis of breast or cervical cancer. *Health and Place,* 11(1), 15–29.

Chakravarthy, S. V. & Ghosh, J. (1996). Scale-based clustering using the radial basis function network. *IEEE Transactions on Neural Networks,* 7(5), 1250–1261

Ciucu, M., Heas, P., Datcu, M. & Tilton, J.C. (2003). Scale space exploration for mining image information content. In O. R. Zaiane, S. Simoff & C. Djeraba (Eds.), *Mining multimedia and complex data* (pp. 118–133). (Berlin : Springer)

Coughlin, S. S., Thompson, T.D., Hall, H.I., Logan, P. & Uhler, R.J. (2002). Breast and cervical carcinoma screening practices among women in rural and nonrural areas of the United States, 1998–1999. *Cancer,* 94(11), 2801–2812

Diehr, P. (1984). Small area statistics – large statistical problems. *American Journal of Public Health,* 74(4), 313–314

Eley, J.W., Hill, H.A., Chen, V.W., Austin, D.F., Wesley, M.N., Muss, H.B., Greenberg, R.S. , Coates, R.J., Correa, P., Redmond, C.K., Hunter, C.P., Herman, A.A, Kurman, R., Blacklow, R., Shapiro, S. & Edwards, B.K. (1994). Racial-differences in survival from breast cancer-results of the national-cancer-institute black/white cancer survival study. *Jama-Journal of the American Medical Association,* 272(12), 947–954

Fraley, C. & Raftery, A.E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *Computer Journal,* 41(8), 578–588

Gath, I. & Geva, A.B. (1989). Unsupervised optimal fuzzy clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 11(7), 773–781

Hart, L.G., Larson, E.H. & Lishner, D.M. (2005). Rural definitions for health policy and research. *American Journal of Public Health,* 95(7), 1149–1155

Horn, D. (2001). Clustering via hilbert space. *Physica A-Statistical Mechanics and Its Applications,* 302(1–4), 70–79

Koenderink, J.J. (1984). The structure of images. *Biological Cybernetics,* 50(5), 363–370

Koenderink, J.J. (1994). Foreword. In T. Lindeberg (Ed.), *Scale-space theory in computer vision* (iii–iv). (Boston: Kluwer Academic)

Lehnerr, M., & Havener, L. (2002). Assessment of interstate exchange of cancer data: Illinois, 1986–1998: Illinois State Department of Public Health

Leung, Y., Zhang, J.S., & Xu, Z.B. (2000). Clustering by scale-space filtering. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 22(12), 1396–1410

Luo, W. & Wang, F.H. (2003). Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region. *Environment and Planning B-Planning & Design,* 30(6), 865–884

Luo, J., Zhou, C., Leung, Y., Zhang, Y. & Huang, Y. (2002). Scale-space theory based regionalization for spatial cells. *ACTA Geographica Sinica,* 57(2), 167–173

Mandelblatt, J., Andrews, J.H., Kerner, J., Zauber, A. & Burnett, W. (1991). Determinants of late stage diagnosis of breast and cervical-cancer – The impact of age, race, social-class, and hospital type. *American Journal of Public Health,* 81(5), 646–649

Miller, D. & Rose, K. (1996). Hierarchical, unsupervised learning with growing via phase transitions. *Neural Computation,* 8(2), 425–450

Mu, L., & Wang, F.H. (2008). A scale-space clustering method: Mitigating the effect of scale in the analysis of zone-based data. *Annals of the Association of American Geographers,* 98(1), 85–101

Rose, K., Gurewitz, E. & Fox, G. (1990). A deterministic annealing approach to clustering. *Pattern Recognition Letters,* 11(9), 589–594.

Sheehan, T.J. & DeChello, L. (2005). A space-time analysis of the proportion of late stage breast cancer in Massachusetts, 1988 to 1997. *International Journal of Health Geographics,* 4(1), 15

Wang, F. (2005). Job access and homicide patterns in Chicago: an analysis at multiple geographic levels based on scale-space theory. *Journal of Quantitative Criminology,* 21(2), 195–217

Wang, F. (2006). *Quantitative methods and applications in GIS*. (Boca Raton, FL: CRC/Taylor & Francis)

Wang, M., Luo J.C. & Zhou, C.H. (2005). Linear belts mining from spatial database with mathematical morphological operators. In X. Li, S. Wang & Z. Y. Dong (Eds.), *Advanced data mining and applications* (pp. 769-776). (Berlin: Springer)

Wang, F., McLafferty, S., Escamilla, V. & Luo, L. (2008). Late-stage breast cancer diagnosis and health care access in Illinois. *The Professional Geographer,* 60(1), 54–69

Wilson, R. & Spann, M. (1990). A new approach to clustering. *Pattern Recognition* 23(12), 1413–1425

Witkin, A.P. (1983). Scale-space filtering. In A. Bundy (Ed.), *The 8th International Joint Conference of Artificial Intelligence* (pp. 1019–1022). (Karlsruhe, Germany: William Kaufmann)

Wong, Y. F. (1993). Clustering data by melting. *Neural Computation,* 5(1), 89–104

Zahn, C. T. (1971). Graphic theoretic methods for detecting and describing Gestalt clusters. *IEEE Trans. Computers,* 20, 68–86

# Chapter 19
# Influence of Job Accessibility on Housing Market Processes: Study of Spatial Stationarity in the Buffalo and Seattle Metropolitan Areas

**Sungsoon Hwang and Jean-Claude Thill**

**Abstract** The impact of job accessibility on housing prices is examined in the Buffalo and Seattle metropolitan areas using a hedonic regression modeling framework. Global hedonic regression results show that job accessibility is positively associated with housing prices in the two study areas. Local hedonic regression modeling is also conducted to test whether the response of the housing market to job accessibility is spatially stationary. The statistical analysis reveals that the role of job accessibility in the house price-setting process varies locally in each metropolitan area. Empirical challenges with unraveling relationship between transportation and land use, and the policy implications of our findings, are discussed.

**Keywords** Job accessibility · Housing market · Geographically weighted regression · Land use/transportation interaction

## 19.1 Introduction

It is widely held that land use (activities) and transportation (linkage and movement) are intimately related (Pickrell 1999; Giuliano 1989; Cervero and Landis 1995; Newman and Kenworthy 1996). Many urban and regional policy measures are grounded in the land use/transportation linkage. For example, large-scale transportation investments such as interstate highways in United States (Rephann and Isserman 1994; Chandra and Thompson 2000) and high-speed rail in Japan (Nakamura and Ueda 1989) usually generate multiplier effects in economic activities through space-time convergence brought about by enhanced accessibility (Rietveld and Bruinsma 1998). Similarly, real estate development such as business parks, retail and recreational complexes generates new focal points for enhanced traffic flows and contributes to changes

S. Hwang (✉)
Department of Geography, DePaul University, Chicago, IL 60614, USA
e-mail: hwang.sungsoon@gmail.com

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_19,

in regional characteristics in the long term (Forkenbrock and Foster 1996; Banister and Berechman 2001).

Not all of these policy measures achieve their intended outcomes. The manner in which land use interacts with transportation is indeed heavily context-dependent. Multiple-path-dependent factors such as economic conditions and local land use/transportation policy influence the complex inner-workings of the land use/transportation relationship. Therefore, it is necessary to further examine how the land use/transportation relationship is uniquely manifested within the confines of local geographic contexts to ensure that pertinent policy measures are effective.

Gradual shifts in public opinion in developed countries support the need for research on the land use/transportation interaction and for integrating land use policy with transportation policy. In recent decades, the legislative landscape (e.g., Clean Air Act, ISTEA, TEA-21) and public sentiment have coalesced in support of an agenda centered on environmental concerns and inequity issues. Open space and agricultural land has been developed to the point that the future water consumption is at risk, while increasing automobile dependence is seen as steering society away from more environmentally sound and equitable practices (Ewing 1997; Kenworthy et al. 1999).

Land use and transportation policies evidently have a role to play in changing the course toward sustainability. Within this vein, policy measures such as transit oriented development (TOD) and smart growth have been implemented as a means to reduce automobile dependence and prevent further environmental degradation (Pickrell 1998; Porter 1998). It is important to note that these policy measures are based on the premise that urban form (e.g., density, land use mix, design) influences travel behavior. Not surprisingly, there was a surge of interest in the effects of urban form on travel behavior as documented in Boarnet and Crane (2001). Despite abundance of research, a conundrum surrounding the validity of this kind of research persists; measuring urban form pertinent to problems at hand admittedly remains a challenging task due to its complex dimension and scale factor (Krizek 2003).

In contrast to abundant research on the impact of land use on transportation, there is a relative lack of research on the effects of transportation on land use. Transportation improvement increases mobility (i.e., overall travel time is reduced), and as a result affected sites become more accessible. Accessibility influences location decision by firm and household in the long term (i.e., firms want to reduce transportation cost, households want to save on commuting, firms want to be located near a qualified labor pool). That is, a change in accessibility leads to land use change, and land use change reshapes accessibility reversely (Giuliano 2004). The concept of accessibility provides the key to understanding the link between transportation and land use.

This study examines how job accessibility influences residential land use at the neighborhood scale in the two metropolitan areas of Buffalo and Seattle. That is, in what manner and to what extent do job accessibility considerations matter in residential location choice? The hypothesis is that the demand for job accessibility in a metropolitan housing market is not spatially invariant within and across the metropolitan area. Geographically weighted regression (Fotheringham et al. 2002) is of a practical value in exploring how demand for job accessibility varies within and across a metropolitan area. This statistical modeling approach differs from global ordinary least squares (OLS) regression where coefficients are not allowed to vary spatially.

The rest of this chapter will be organized as follows. Section 19.2 discusses the literature on the effect of job accessibility on housing prices in terms of research strategy, and reviews empirical challenges with a particular emphasis on geographic aspects. Section 19.3 describes the methodology, including the formulation of job accessibility measures, the hedonic model, and the geographically weighted regression of housing prices. Section 19.4 reports on the results of the statistical analysis. Section 19.5 discusses the implications and limitations of the study.

## 19.2 Literature Review

### *19.2.1 Research Strategies*

Studies that have examined the role of job accessibility in shaping urban housing markets have followed one of three research strategies. One strand of research looks at how increase in accessibility arising from new transportation investment is capitalized into housing price over time. It is usually conducted by regressing the change in housing prices on the change in accessibility resulting from transportation improvements, while controlling for other factors. It allows the analyst to examine how housing market adjusts to change in accessibility. Empirical findings are mixed (Huang 1996; Ryan 1999; Gibbons and Machin 2008). Scale and timing of transportation investment, local economic conditions, and land use policy are found to influence how land and housing markets respond to increase in accessibility.

In a second group of studies, the relationship between housing prices and components of housing services (including job accessibility) is examined using so-called hedonic price models. The theoretical underpinnings of hedonic modeling are that the value of consumer goods (housing service in this case) is composed of a bundle of composite attributes (Lancaster 1971). Hedonic modeling attempts to estimate how the bundling of housing-related attributes is reflected in housing prices. Structural characteristics of housing units and locational characteristics of the neighborhood of each housing unit usually constitute the dimensionality of the housing service. The modeling is operationalized by a multiple regression model where housing price is the dependent variable and multiple housing-related attributes constitute independent variables. The regression coefficients of job accessibility measures serve to estimate the implicit price of job accessibility or housing price (dis)premium attached to job accessibility. Hedonic modeling establishes how demand for job accessibility comes into play in housing market processes. The present study contributes to this line of research.

A third body of research has looked at the relative importance of accessibility in residential location decisions. Decision makers choose alternatives (residential locations in this case) by maximizing utilities derived from multiple attributes that characterize the alternative in their choice set; varying influence of these attributes (including job accessibility) on housing choice is estimated by multinomial logit models and other discrete choice models (McFadden 1978). A number of empirical studies have found that accessibility is of less significance to residential location decision than other factors such as housing and neighborhood characteristics (Molin and Timmermans 2003), but many others rank job accessibility among the most significant residential choice factors (Quigley 1985; Thill and Van de Vyvere 1989). In addition, household structure is an important determinant of the magnitude of job accessibility considerations in residential location decisions (Waddell 1996).

### *19.2.2 Empirical Challenges*

This section reviews empirical challenges and evidences provided by studies that employ cross-sectional hedonic modeling (the second research strategy mentioned above). We focus on challenges of dealing with locational variables, and make recommendations for handling them.

First, it is notable that statistical results are often inconsistent depending on the specification of the job accessibility measure. The formulation of aggregate job accessibility measures ranges from a simple count of job opportunities within a certain distance of a reference location, to gravity-based forms. Gravity-based measures of accessibility are considered more accurate due to the continuous distance decay function, which is consistent with principles of spatial interaction (Roy and Thill 2004). Yet, gravity-based measures of accessibility can be specified in a number of ways, which may exhibit close collinearity (Kwan 1998; Thill and Kim 2005). Gravity-based cumulative opportunities access measures were found to be most helpful in predicting residential location according to Issam et al. (2002). Results are also inconsistent depending on whether measures are based on travel time or travel distance to job locations. In her survey paper, Ryan (1999) indicates that house price in US metropolitan areas is negatively associated with accessibility when measured on travel time, but is found to be otherwise when measured on travel distance. It is, however, unclear whether multicollinearity problems (i.e., accessibility is highly correlated with other explanatory variables) are explicitly treated in the studies reviewed. Also, the relationship may not be generalizable given the limited number of study areas. In principle, travel time-based accessibility should more accurately reflect what accessibility implies than travel distance-based measures given that commuters and other travelers are first and foremost responsive to travel time rather than to travel distance as travel time is tallied up against their time budget (Golledge and Stimson 1997). Accessibility to different types of activities (e.g., shopping, educational, recreational, and employment) is shown to have different impacts on property values, and employment accessibility positively contributes to property values in Seattle according to Franklin and Waddell (2003).

Second, research shows that including precisely measured locational variables in the hedonic model of house prices improves the performance of the model. The measurement of locational variables has become easier as geographic information systems (GIS) are increasingly used in real estate analysis (Anselin 1998). To some extent, this can mitigate biases related to omitted variables, a common occurrence in hedonic modeling. It is, however, not devoid of problems either; spatial autocorrelation can bias the model estimates as the assumption of independence among observations is violated. Moreover, more often than not, locational variables are highly correlated with other variables. Diagnostics for spatial autocorrelation (Anselin et al. 2006) and factor analysis may be useful in this regard, respectively. Once multicollinearity and spatial autocorrelation are treated statistically, it is found that house price premiums are attached to highly accessible sites in the Quebec urban community, Canada (des Rosiers et al. 2000). In Charlotte, NC, Munroe (2007) found housing value to decrease significantly with distance to the central business district (CBD) and to major employers, and to decrease in proximity of brownfield sites, whereas parks and greenways had no conclusive effect on real estate prices. Both travel time to the CBD (as a proxy of urban attraction or centrality) and gravity-based job accessibility contribute significantly to explaining spatial variation of housing prices in the southern Norwegian region (Osland and Thorsen 2008).

Third, more and more studies find that the impact of job accessibility on housing prices is not constant over the study area. Adair et al. (2000) show that job accessibility has a minimal impact on housing prices in the whole study area, but it exerts varying influence across sub-regions in the Belfast urban area (UK). It is likely that preference for employment accessibility is outweighed by the preference for large space (Alonso 1964), public service (Tiebout 1956), or amenities (Rosen 1974) in areas where job accessibility is negatively associated with housing prices. This suggests that the demand function of different attributes that compose housing quality is spatially disaggregated and non-stationary within a metropolitan area (Straszheim 1975; Maclennan et al. 1987).

In summary, the literature review suggests that research findings may be sensitive to the formulation of job accessibility measures, to the treatment of spatial autocorrelation and multicollinearity, and to the geographic scale of analysis (the so-called modifiable areal unit problem). In this study, factor analysis is employed to correct for multicollinearity among housing price factors, and a geographically weighted regression is used to deal with spatial autocorrelation and spatial non-stationarity of the relationship between job accessibility and housing prices. We now describe the methodology designed to address the empirical challenges identified here.

## 19.3 Methodology

The research methodology consists of three steps. First, we compute travel time-based job accessibility measures at the level of census tracts in two U.S. metropolitan areas of Buffalo and Seattle. Second, explanatory variables expected to influence housing value are transformed to underlying dimensions that are independent of each other by factor analysis. Third, factor scores including the job accessibility dimension are entered into the multiple regression model of hedonic house prices. This allows us to determine the magnitude of the effect of job accessibility on housing prices, which is assumed to be constant over the study area. Most importantly, we estimate local coefficients of job accessibility fitted to a geographically weighted regression model. The mapping of local coefficients permits us to examine the spatial variation of the demand for job accessibility at the neighborhood scale in the Buffalo and Seattle metropolitan areas.

### *19.3.1 Job Accessibility Measures*

The job accessibility measure is calculated according to Hansen (1959)'s formulation as follows:

$$A_i = \sum_j O_j \exp(-\beta C_{ij}) \tag{19.1}$$

where $A_i$ is job accessibility of residential origin $i$, $O_j$ is an attractiveness measure of potential commute destination $j$, $C_{ij}$ is a measure of spatial separation, and $f(C_{ij})$ is the function of spatial friction. In this study, the total number of workers employed in destination $j$ is used as a proxy for $O_j$. Mean travel time in minutes between $i$ and $j$ is used as a measure of spatial separation between tracts. Data source of $O_j$ and $C_{ij}$ is the 2000 Census Transportation Planning Package (CTPP) Part 3 data at the level of the census tract. The transformation exp(- $\beta C_{ij}$) is chosen as a functional form of spatial friction. Spatial deterrence parameter $\beta$ is calibrated by maximum likelihood estimation for each metropolitan area using a version of SIMODEL (Fotheringham and O'Kelly 1989) modified for processing CTPP Part 3 data.

## 19.3.2 Factor Analysis

Table 19.1 shows the list of explanatory variables compiled for hedonic regression modeling. Variables are chosen in accordance with the relevant literature (Follain and Jimenez 1985). The dependent variable is the neighborhood's median owner-occupied home value. Explanatory variables are related to neighborhood characteristics (school quality, crime, and job accessibility), resident attributes (income, educational attainment, occupation, life cycle, ethnicity, length of residence), and structural characteristics of the properties (house size, house type, house age). The unit of analysis is the census tract. Most variables are directly available from the 2000 U.S. Population and Housing Census, except for school quality and crime. Data are disaggregated to the census tract level through spatial overlay in GIS when data are available at a coarser geographic resolution.

As many explanatory variables are correlated with each other, factor analysis is used to extract underlying dimensions that are independent of each other from the variables listed in Table 19.1. The method of factor analysis is based on the extraction of principle components with a varimax rotation. The number of factors is determined so that the job accessibility measure indexes one of the

**Table 19.1** Candidate explanatory variables for hedonic price model estimation

| Name | Description | Data | Year | Geographic unit |
|---|---|---|---|---|
| Residents-related attributes | | | | |
| pcincome | Per capita income | Census | 2000 | Census tract |
| College | % College degree holders | Census | 2000 | Census tract |
| Managep | % Management workers | Census | 2000 | Census tract |
| Prodp | % Production workers | Census | 2000 | Census tract |
| Famcpchl | % Family with children | Census | 2000 | Census tract |
| Nfmalone | % Nonfamily living alone | Census | 2000 | Census tract |
| black_p | % Black | Census | 2000 | Census tract |
| nhwht_p | % Non-hispanic white | Census | 2000 | Census tract |
| Nativebr | % Native born | Census | 2000 | Census tract |
| Housing-specific attributes | | | | |
| Medroom | Median number of rooms | Census | 2000 | Census tract |
| Hudetp | % Detached housing units | Census | 2000 | Census tract |
| Yrhublt | Median year structure built | Census | 2000 | Census tract |
| Locational attributes | | | | |
| Ptratio | Pupil to teacher ratio | NCES[a] | 2002 | School district |
| Schexp | School expenditure per student | NCES | 2002 | School district |
| Vrlcrime | Violent crime rate | FBI[b] | 2003 | Designated place |
| Prpcrime | Property crime rate | FBI | 2003 | Designated place |
| Jobacm | Job accessibility | CTPP[c] | 2000 | Census tract |

components and that all components explain the variation in the original data adequately.

### 19.3.3 Hedonic Modeling

The relationship between housing prices and job accessibility is examined using two regression models: the first is a global regression and the other is a local regression. A global regression model can be written as:

$$y_i = \beta_0 + \Sigma_k \beta_k x_{ik} + \varepsilon_i \tag{19.2}$$

where $y_i$ is the median owner-occupied home value in tract $i$ , $x_{ik}$ is a vector of predictors, and $\beta_k$ and $\varepsilon_i$ are vectors of parameters and errors, respectively. In a global model, it is assumed that the values of parameters are constant across the study area. In this model, geographic variation in the relationship is confined to the error term. Spatial variation can be accommodated such that the relationship is not treated in the error term, by taking account of the location of the study area. As a form of local regression, Geographically Weighted Regression can be rewritten as follows:

$$y_i = \beta_0(u_i, v_i) + \Sigma_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \tag{19.3}$$

where $(u_i,v_i)$ denotes the coordinates of the $i$ th point in space and $\beta_k(u_i,v_i)$ is a realization of the continuous parametric function $\beta_k(u,v)$ at location $i$ . In essence, Equation (19.3) measures the relationship inherent in the model around each location $i$. The estimation of $\beta_k$ $(u_i,v_i)$ is a function of the geographic weighting of each of the $n$ observed data for regression point $i$ . The spatial weighting function chosen for this analysis is a bi-square function where the size of the spatial kernel (i.e., the bandwidth) is allowed to vary spatially through minimization of the Akaike Information Criterion to best fit the spatially-varying distribution of observations (Fotheringham et al. 2002, pp. 57–62). It can be noted that Equation (19.2) is a special case of Equation (19.3) in which the parameters are spatially invariant.

Factor scores are entered into both global and local multiple regression models as independent variables. Residuals derived from global hedonic prediction are mapped to determine whether spatial autocorrelation is present. Testing for spatial autocorrelation such as Moran's I will indicate whether geographically weighted regression should be considered. The software GWR 3.0 (Fotheringham et al. 2002) attaches spatially varying coefficients to each observation (census tract in this case). We determine whether the GWR coefficient of job accessibility is consistent throughout the study area using a Monte Carlo test of spatial stationarity.

## 19.4 Results

### *19.4.1 Study Area*

The Buffalo-Niagara Falls MSA (Metropolitan Statistical Area) is located in the western part of New York State, and is the second largest in the state after New York City. Like much of the Rust Belt, this region is faced with deep structural economic problems and a declining and aging population. The metropolitan area had a population of 1,161,832 according to 2000 Census. Seattle-Tacoma-Bremerton CMSA (Consolidated Metropolitan Statistical Area) is the largest metropolitan area in the Pacific Northwest. The metropolitan area had a population of 3,553,244 according to the 2000 Census of Population. In contrast to Buffalo, Seattle has enjoyed fast economic and population growth over the past three decades. Hence, the selected metropolitan areas represent highly contrasted regional economies and metropolitan housing markets. For simplicity, we will refer to the metropolitan areas as the Buffalo metropolitan area and Seattle metropolitan area, respectively, from this point forward.

Different economic conditions are well reflected in the range of housing prices as shown in Fig. 19.1. The map shows the spatial distribution of

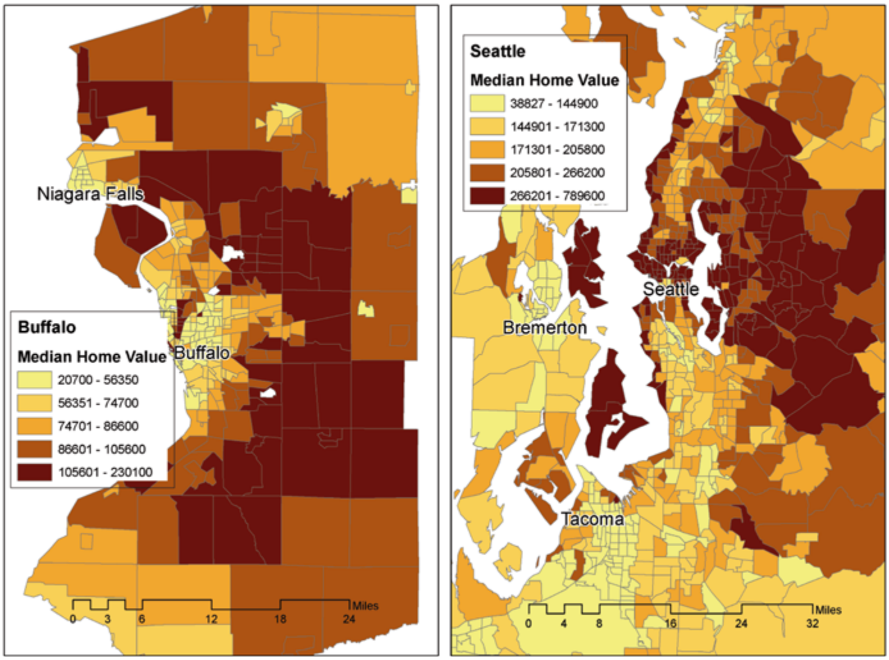



**Fig. 19.1** Housing prices in the Buffalo and Seattle metropolitan areas

median price of owner-occupied housing units by census tract. Data is classified by a quantile method for comparison purposes. Each of the five classes represents 20% of data range in order. In general, areas of high housing prices in the Buffalo metropolitan area are concentrated in the outer suburbs (e.g., Amherst, Clarence, Orchard Park, Grand Island) and Elmwood Village (the swath extending northward from the downtown Buffalo). Areas of high housing prices in the Seattle metropolitan area are concentrated in the suburbs (Bellevue, Mercer Island, Bainbridge Island) and the shore sides of Seattle.

Similarly, the spatial distribution of jobs is depicted in Fig. 19.2 using a quantile method. Figure 19.2 maps the total number of workers per acre by census tract in the Buffalo and Seattle metropolitan areas. Worker-to-population ratio of the Buffalo metropolitan area as a whole is 0.44, and that of the Seattle metropolitan area is 0.5. Thus it can be said that there are more job opportunities in the Seattle metropolitan area than the Buffalo metropolitan area. Figure 19.2 shows that in general jobs are more decentralized in the Buffalo metropolitan area compared to the Seattle metropolitan area, after absolute job density is considered. It can be seen that high employment density is associated with the proximity to major highways.

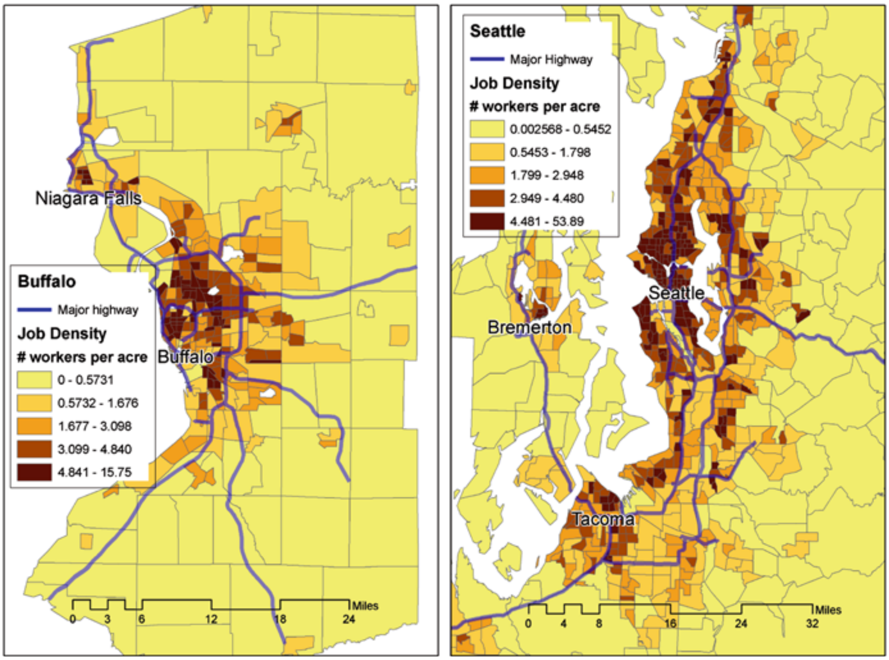



**Fig. 19.2** Employment density in the Buffalo and Seattle metropolitan areas

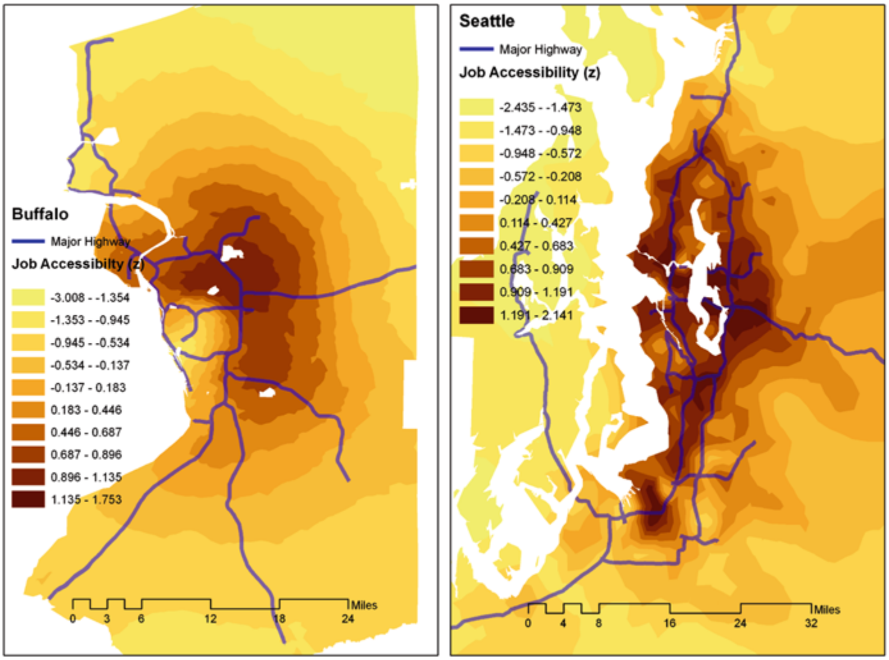



**Fig. 19.3** Kriging maps of job accessibility measures in the Buffalo and Seattle metropolitan areas

## *19.4.2 Job Accessibility*

Job accessibility measures are depicted in Fig. 19.3 as a contour map. Areas in darker shades are more accessible to job opportunities than others. Job accessibility measures are standardized (i.e., converted to *z*-scores) because job accessibility measures increase with the size of the study area (more specifically, the number of workers in this study). Job accessibility is computed on the basis of travel time, thus the proximity to major highways positively influences job accessibility measures. The spatial deterrence parameter $\beta$ calibrated for the Buffalo metropolitan area is 0.03158, and the Seattle counterpart is 0.02443.

The combination of the long-term trend toward decentralization of employment from the city center and the building of limited-access highways in the inner suburbs of the Buffalo metropolitan area have led to high job accessibility in those areas (Amherst in particular). Downtown Buffalo and Niagara Falls no longer enjoy good accessibility to jobs. It can be seen from Fig. 19.3 that job accessibility is ubiquitously high in the City of Seattle, Bellevue, and Tacoma in the Seattle metropolitan area. It should be noted, however, that the Seattle metropolitan area is unique in many ways; for example, land is confined by the ocean and other bodies of water; thus highways (esp. I-5) cannot induce employment decentralization. Seattle provides an interesting case from which we can examine how nature-induced land use control plays a role in shaping job accessibility while economic conditions need to be controlled for.

### 19.4.3 Dimensions of Housing Markets

Factor analysis extracts principal components of the housing markets in the Buffalo and Seattle metropolitan areas. Six components constitute the Buffalo housing market, accounting for 86.15% of the total variation of all 17 variables listed in Table 19.1. They are interpreted as crime, human capital, life cycle, ethnicity, job accessibility, and length of residence. The human capital factor can be seen as an aggregate of income, educational attainment, and occupation. The variables of housing type and size are highly correlated with life cycle, while housing age is correlated with crime in the Buffalo metropolitan area.

The Seattle housing market is best represented by eight components. They are human capital, life cycle, crime, length of residence, school quality, ethnicity, housing age, and job accessibility. Collectively, those components explain 92.47% of the total variance. The percentage of non-Hispanic whites is correlated with the length of residence, and the ethnicity factor is negatively associated with the percentage of Blacks. There is more correlation between income and education/occupation in the Seattle metropolitan area than in the Buffalo metropolitan area.

### 19.4.4 Global Regression

Table 19.2 shows OLS regression results. In the case of the Buffalo metropolitan area, all of six components are found to be statistically significant in

**Table 19.2** Global regression results

| (a) Buffalo-Niagara Falls MSA | | | (b) Seattle-Tacoma-Bremerton CMSA | | |
|---|---|---|---|---|---|
| Global hedonic model | B | t | Global hedonic model | B | t |
| (Constant) | 82572.66 | 84.56 | (Constant) | 211754.7 | 107.946 |
| Crime[a] | –8867.63 | –9.066 | Human capital[a] | 73266.47 | 37.325 |
| Human capital[a] | 23418.08 | 23.941 | Life cycle[a] | 8510.768 | 4.336 |
| Life cycle[a] | 8583.926 | 8.776 | Crime[b] | –3973.47 | −2.024 |
| Ethnicity[a] | 6403.728 | 6.547 | Length of residence | 1116.597 | 0.569 |
| Job accessibility[a] | 5883.911 | 6.015 | School quality | −1080.38 | −0.55 |
| Length of residence[b] | −2114.44 | −2.162 | Ethnicity[a] | 9792.13 | 4.989 |
| | | | Housing age | 318.506 | 0.162 |
| | | | Job accessibility[a] | 26195.68 | 13.345 |
| Adjusted R-square = 0.734 | | | Adjusted R-square = 0.677 | | |
| Standard error = 16,743 | | | Standard error = 54,398 | | |

[a] Significant at the 0.01.level
[b] Significant at the 0.05 level.

explaining the variation of median housing value. Five components (human capital, life cycle, crime, ethnicity, and job accessibility) out of eight components turn out to be significant dimensions of the housing market in the Seattle metropolitan area. Higher standard error and relatively lower adjusted *R*-square in the Seattle metropolitan area compared to the Buffalo metropolitan area indicates that the hedonic model does not perform quite well compared to the former. It may arise from omitted variables (e.g., site characteristics such as ocean view) and the quality of school data.

Most importantly, job accessibility turns out to be a significant determinant of housing price in both metropolitan areas. This component is statistically significant at the 0.01 level. The positive sign of the coefficient of job accessibility component implies that sites accessible to job opportunities are considered more desirable, and good access to job offers house price premium in both Buffalo and Seattle metropolitan area. Interestingly, the elasticity of housing price to job accessibility is over four times larger in Seattle than in Buffalo. This could be the result of Buffalo's greater overall job accessibility (partly due to its smaller size), which provides little incentive for local residents to pay a real estate premium for housing that is more accessible to job opportunities.

We calculate Moran's I statistic to determine whether residuals from the global regression exhibits spatial autocorrelation. Buffalo's Moran's I index is 0.05 ($z$-score is 10.14), and that of the Seattle metropolitan area is 0.04 ($z$-score is 15.04). This indicates that regression residuals are positively spatially autocorrelated in both metropolitan areas. Hence we turn our attention to the estimation of geographically weighted regression models.

### 19.4.5 Local Regression

A summary of GWR estimation results is provided in Table 19.3. The rightmost column shows the statistical significance of the Monte Carlo test of spatial variability of local regression coefficients. It tests whether the null hypothesis that local coefficients of the respective component are constant over the study area (i.e., spatial stationarity of coefficients) can be rejected. It is notable that all components except for ethnicity and length of residence reject the null hypothesis at the 0.05 level in the Buffalo metropolitan area. That is, the impact of those components on housing prices varies spatially. Job accessibility is not an exception. While the median of the job accessibility coefficients is positive (which is consistent with the global regression results), their values range from positive to negative. In some cases, job accessibility depresses housing value as indicated by the negative coefficients estimated in the Buffalo metropolitan area.

Overall results in the Seattle metropolitan area are rather similar to those of Buffalo with regard to the spatial variability of the relationship between factors and housing price; all factors except for life cycle, crime, and housing age exhibit statistically significant variability at the 0.05 level. Combined with global regression

**Table 19.3** Local regression results

(a) Buffalo-Niagara Falls MSA

| Label | Minimum | Lwr quartile | Median | Upr quartile | Maximum | Sig. |
|---|---|---|---|---|---|---|
| Intercept | 70618.20 | 75859.86 | 78340.75 | 82873.55 | 92678.50 | 0.00 |
| Crime | −24158.39 | −12158.78 | −8722.78 | −7116.00 | −1070.97 | 0.00 |
| Human capital | 13400.33 | 17533.58 | 20502.66 | 25632.61 | 29127.53 | 0.04 |
| Life cycle | −809.24 | 3237.95 | 7803.32 | 12602.80 | 19399.02 | 0.00 |
| Ethnicity | −2393.25 | 5269.02 | 6493.57 | 7658.22 | 18719.68 | 0.29 |
| Job accessibility | −5852.97 | −559.46 | 5200.96 | 7752.86 | 12191.91 | 0.00 |
| Length of residence | −9558.35 | −5352.72 | −3572.81 | −1937.71 | 1385.32 | 0.07 |
| Adjusted R-square = 0.861 | | | | | | |
| Standard Error = 13110 | | | | | | |

(b) Seattle-Tacoma-Bremerton CMSA

| Label | Minimum | Lwr Quartile | Median | Upr Quartile | Maximum | Sig. |
|---|---|---|---|---|---|---|
| Intercept | 125495.11 | 177423.88 | 191496.49 | 206157.82 | 224078.04 | 0.00 |
| Human capital | 30531.06 | 44466.25 | 71933.14 | 97384.99 | 122000.87 | 0.00 |
| Life cycle | −6447.06 | 129.18 | 4346.28 | 8624.06 | 17008.99 | 0.81 |
| Crime | −18481.04 | −6580.01 | −3622.20 | 328.81 | 27509.61 | 0.08 |
| Length of residence | −32583.44 | −12513.29 | −948.12 | 5804.33 | 18776.00 | 0.00 |
| School quality | −21208.30 | −6388.85 | −1256.65 | 11191.19 | 47893.10 | 0.00 |
| Ethnicity | −22438.75 | 6765.75 | 12552.73 | 14965.08 | 32997.48 | 0.00 |
| Housing age | −17552.95 | −5642.52 | 1776.13 | 4084.50 | 10070.33 | 0.17 |
| Job accessibility | 2782.17 | 13093.78 | 18189.55 | 43108.87 | 75501.66 | 0.00 |
| Adjusted R-square = 0.79 | | | | | | |
| Standard Error = 44695 | | | | | | |

results, it can be inferred that school quality and length of residence play important roles in some neighborhoods although they do not significantly contribute to explaining the variation of housing prices at a metropolitan scale, given high variability of corresponding local coefficients. In comparison to Buffalo, it is notable that job accessibility offers house price premium all across the Seattle metropolitan area. This premium is highest in the CBD and in Bellevue.

Geographically weighted regression improves the performance of hedonic modeling significantly as indicated by the higher share of variation explained and the lower prediction errors. More pertinent to real estate analysis, local regression results suggest that using a global regression model for housing valuation or estimating implicit price of attributes comprising housing quality is not quite valid due to the spatial non-stationarity of the relationship between predictors and housing prices. This study signifies that the manner in which some attributes exert influence on housing value is contingent upon local contexts. Consequently, spatial processes are more at play in the operation of housing markets than it has been recognized thus far. It warrants further study to better understand the spatial nature of housing market processes.

To examine how the implicit price of job accessibility varies at a fine geographic granularity, we map $t$-values of local job accessibility coefficients. Figure 19.4 shows how the relationship between job accessibility and housing

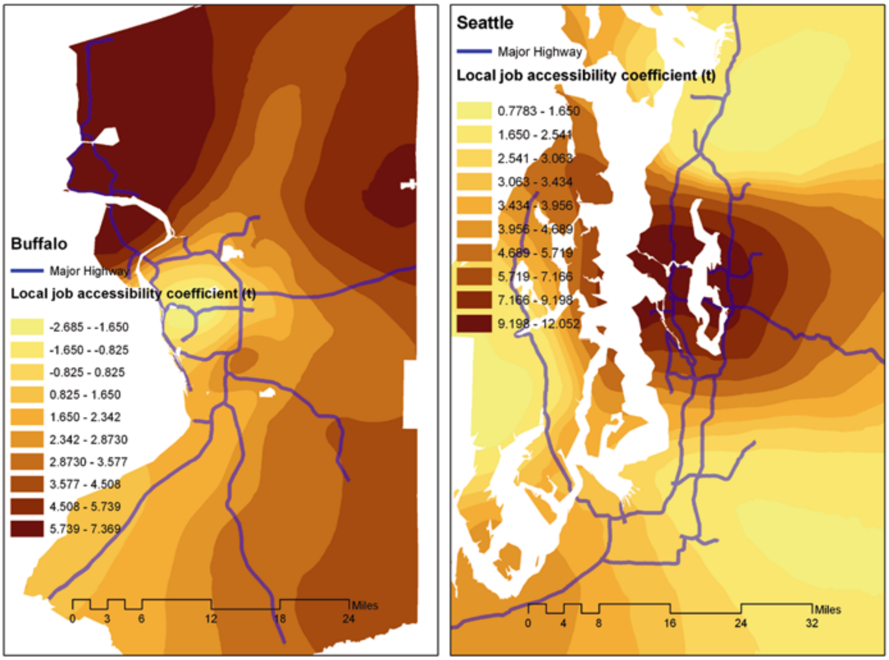



**Fig. 19.4** The impact of job accessibility on median housing value as given by estimated $t$ values in the Buffalo and Seattle metropolitan areas

prices varies spatially. Job accessibility positively influences home value in areas with $t$-values greater than 1.65 (significant at the 0.05 level). Job accessibility is negatively associated with housing prices in areas with $t$-values less than -1.65. Areas with $t$-values between -1.65 and 1.65 can be interpreted as areas where job accessibility does not influence the local housing market.

Striking difference between the Buffalo and Seattle metropolitan area can be observed with regard to the effect of job accessibility on housing prices. Areas with high demand for job accessibility are concentrated in the outer suburbs of the Buffalo metropolitan area. Strong positive demand for job accessibility is present throughout Seattle and Bellevue. Negative implicit price of job accessibility in the City of Buffalo is quite noteworthy. Although the hedonic model using aggregate data does not directly capture behavioral aspects, results indirectly suggest that nearly all city (Buffalo) residents and inner suburbia in the Buffalo metropolitan area are either indifferent to increases in job accessibility or do not consider job accessibility to be important in housing choice. The same interpretation can be made about the outer suburbs of the Seattle metropolitan area. Further research on this issue using disaggregate analysis is needed to validate this point.

## 19.5 Conclusions

Despite the importance of and the need for better understanding the land use/transportation relationship, the task of unraveling their intimate relationship remains challenging. This study attempted to overcome some of these empirical challenges. More specifically, we employed factor analysis to resolve multicollinearity of locational variables in hedonic modeling, and geographically weighted regression to examine the spatially variant role of job accessibility in articulating metropolitan housing markets.

A local hedonic model based on geographically weighted regression in the metropolitan areas of Buffalo and Seattle allows us to infer the implicit price of job accessibility with regard to housing choice. Compelling empirical evidence from two very dissimilar metropolitan areas indicates that the market response to job accessibility is not spatially stationary. Instead, a clear geographic pattern exists with regard to how job accessibility may influence housing value. Our analysis suggests that suburbanites are more willing to pay for additional increases in job accessibility in housing consumption than urban residents in the Buffalo-Niagara Falls MSA. In contrast, residents living near the urban core of Seattle are more likely to accept to pay a real estate premium for high job accessibility than those who live further away.

It is necessary to look at neighborhood-scale operations to fully grasp the role of job accessibility in housing markets. Inconsistent findings in this body of literature (Priemus et al. 2001) are not free from the scale of analysis. This study demonstrates the utility of geographically weighted regression in examining how the relationship between job accessibility and land use plays out locally. Results shed light on the localized nature of housing market processes.

Mapping local impacts of job accessibility on housing prices can inform policy analysts on how residents may respond to gains in job accessibility stemming from potential transportation investments in the short term. Marginal transportation improvement project in areas of low demand or negative demand for job accessibility (e.g., city of Buffalo, and outer suburbs of the Seattle metropolitan area) is not likely to harvest anticipated policy outcomes such as urban revitalization or economic development. Transportation policy alone would not be sufficient in reaping potential benefits in areas characterized as such, and should be supplemented with other land and human capital development policies. Conversely, sections of metropolitan areas with high demand will quickly respond to even marginal increases in job accessibility, and thus fully benefit from the regional multiplier effects.

Despite the rather complex nature of the concept of accessibility, we have shown that it deserves greater attention in studies of land use/transportation interaction because it provides a crucial medium to understanding this complex linkage. As demonstrated above, accessibility is in touch with equity and efficiency arguments (e.g., urban revitalization) of land use/transportation policies (Cervero et al. 1999).

A limitation of this study can be traced to quality of the data. Using more disaggregate data such as actual real estate transactions in addition to census data would improve the performance of hedonic models to understand housing market processes. In addition, neighborhood characteristics of crime, school quality, and accessibility to a range of urban amenities are best measured at a fine spatial resolution. Our analysis has demonstrated that housing markets have a strong local component and that market processes are fundamentally non-stationary over space. Spatial processes are known to often operate at multiple scales. Likewise, rejoining Pace and Gilley (1997), it can be conjectured that different factors shape metropolitan housing markets at different spatial scales. Therefore, an integrated multi-scale analysis may provide a well-suited research framework for the study of housing markets.

## References

Adair, A., McGreal, S., Smyth, A., Cooper, J. & Ryley, T. (2000). House prices and accessibility: the testing of relationships within the Belfast urban area. *Housing Studies,* 15(5), 699–716

Alonso, W. (1964). *Location and land use*. (Cambridge: Harvard University Press)

Anselin, L. (1998). GIS research infrastructure for spatial analysis of real estate markets. *Journal of Housing Research,* 9(1), 113–133

Anselin, L., Syabri, I. & Kho, Y. (2006). GeoDa: an introduction to spatial data analysis. *Geographical Analysis,* 38, 5–22

Banister, D. & Berechman, Y. (2001). Transport investment and the promotion of economic growth. *Journal of Transport Geography,* 9, 209–218

Boarnet, M.G. & Crane, R. (2001). *Travel by design: the influence of urban form on travel.* (New York: Oxford University Press)

Cervero, R. & Landis, J. (1995). The transportation-land use connection still matters. *Access,* 7, 2–10

Cervero, R., Rood, T. & Appleyard, B. (1999). Tracking accessibility: employment and housing opportunities in the San Francisco Bay Area. *Environment and Planning A,* 31, 1259–1278

Chandra, A. & Thompson, E. (2000). Does public infrastructure affect economic activity?: evidence from the rural interstate highway system. *Regional Science and Urban Economics,* 30(4), 457–490

des Rosiers, F., Theriault, M. & Villeneuve, P.-Y. (2000). Sorting out access and neighborhood factors in hedonic price modeling. *Journal of Property Investment & Finance,* 18(3), 291–315

Ewing, R. (1997). Is Los Angeles-style sprawl desirable? *Journal of the American Planning Association,* 63, 107–126

Follain, J. & Jimenez, E. (1985). Estimating the demand for housing characteristics: a survey and critique. *Regional Science and Urban Economics,* 15, 77–107

Forkenbrock, D.J. & Foster, N.S.J. (1996). Highways and business location decisions. *Economic Development Quarterly,* 10(3), 239–248

Fotheringham, A.S. & O'Kelly, M.E. (1989). *Spatial interaction models: formulations and applications*. (Dordrecht; Boston: Kluwer Academic Publishers)

Fotheringham, A.S., Brundsdon, C. & Charlton, M. (2002). *Geographically weighted regression: the analysis of spatially varying relationships*. (New Jersey:John Wiley & Sons)

Franklin, J.P. & Waddell, P. (2003). A hedonic regression of home prices in King County, Washington, using activity-specific accessibility measures. *Paper presented at Transportation Research Board Meeting,* **Washington**, D.C

Gibbons, S. & Machin, S. (2008). Valuing school quality, better transport and lower crime: evidence from house prices. *Oxford Review of Economic Policy,* 24(1), 99–119

Giuliano, G. (1989). New directions for understanding transportation and land use. *Environment and Planning A,* 21(2), 141–159

Giuliano, G. (2004). Land use impacts of transportation investments. In S. Hanson & G. Giuliano (Eds.), *The geography of urban transportation* (pp. 237–273). (New York: Guilford)

Golledge, R.G. & Stimson, R.J. (1997). *Spatial behavior. A geographic perspective*. (New York: Guilford Press)

Hansen, W. (1959). How accessibility shapes land use. *Journal of the American Institute of Planners,* 25, 73–76

Huang, H. (1996). The land-use impacts of urban rail transit systems. *Journal of Planning Literature,* 11(1), 17–30

Issam, M.S., Kockelman, K.M. & Dunn, T.P. (2002). Accessibility indices: a connection to residential land prices and location choices. *Transportation Research Record,* 1805, 25–34

Kenworthy, J.R., Laube, F.B. & Newman, P. (1999). *An international sourcebook of automobile dependence in cities 1960–1990*. (Colorado: University Press of Colorado)

Krizek, K.J. (2003). Operationalizing neighborhood accessibility for land use-travel behavior research and regional modeling. *Journal of Planning Education and Research,* 22(3), 270–287

Kwan, M.P. (1998). Space-time and integral measures of individual accessibility: a comparative analysis using a point-based framework. *Geographical Analysis,* 30, 191–216

Lancaster, K.J. (1971). *Consumer demand: a new approach*. (New York: Columbia University Press)

Maclennan, D., Munro, M. & Wood, G. (1987). Housing choices and the structure of housing markets. In B. Turner, J. Kemeny, & L.J. Lundqvist (Eds.), *Between state and market:*

*Housing in the post-industrial era* (pp. 26–52). (Stockholm: Almqvist & Wiksell International)

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*, 672, 72–77

Molin, E. & Timmermans H.J.P. (2003). Accessibility considerations in residential location decisions: accumulated evidence from the Benelux. *Paper presented at the Annual Transportation Research Board Meeting,* Washington, D.C

Munroe, D.K. (2007). Exploring the determinants of spatial pattern in residential land markets: Amenities and disamenities in Charlotte, NC, USA. *Environment and Planning A,* 34, 336–354

Nakamura, H. & Ueda, T. (1989). The impact of the Shinkansen on regional development. *Paper presented to the Fifth World Conference on Transportation Research,* Antwerp, Belgium

Newman, P.W.G. & Kenworthy, J.R. (1996). The land use-transport connection. *Land Use Policy,* 13(1), 1–22

Osland, L. & Thorsen, I. (2008). Effects on housing prices of urban attraction and labor-market accessibility. *Environment and Planning A,* 40(10), 2490–2509

Pace, R.K. & Gilley, O.W. (1997). Using the spatial configuration of the data to improve estimation. *Journal of Real Estate Finance and Economics,* 72, 34–55

Pickrell, D. (1998). Smart transportation for smart growth. *Smart growth: economy, community, environment* (pp. 12–19). (Washington, D.C.: Urban Land Institute)

Pickrell, D. (1999). Transportation and land use. In J.A. Gomez-Ibanez, W.B. Tye, & C. Winston (Eds.), *Essays in transportation economics and policy: a handbook in honor of John R. Meyer* (pp. 403–435). (Washington, D.C.: Brookings Institution Press)

Porter, D. (1998). Transit-focused development and light rail systems: the lite connection. *Transportation Research Record,* 1623, 154–169

Priemus, H., Nijkamp, P. & Banister, D. (2001). Mobility and spatial dynamics: an uneasy relationship. *Journal of Transport Geography,* 9, 167–171

Quigley, J. (1985). Consumer choice of dwelling, neighborhood and public services. *Regional Science and Urban Economics,* 15, 41–63

Rephann, T.J. & Isserman, A.M. (1994). New highways as economic development tools: an evaluation using quasi-experimental matching methods. *Regional Science and Urban Economics,* 24(6), 723-751

Rietveld, P. & Bruinsma, F. (1998). *Is transportation infrastructure effective? transportation infrastructure and accessibility: impacts on the space economy*. (Berlin: Springer-Verlag)

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy,* 82(1), 34–55

Roy, J.R. & Thill J.C. (2004). Spatial interaction modeling. *Papers in Regional Science,* 83(1), 339–361

Ryan, S. (1999). Property values and transportation facilities: finding the transportation-land use connection. *Journal of Planning Literature,* 13(4), 412–427

Straszheim, M.R. (1975). *An econometric analysis of the urban housing market*. (New York: National Bureau of Economic Research)

Thill, J.C., & Van de Vyvere, Y. (1989). Workplace and locational choice of residence: a hierarchical approach. *Sistemi Urbani,* 3, 339–365

Thill, J.C. & Kim, M. (2005). Trip making, induced travel demand, and accessibility. *Journal of Geographical Systems,* 7, 229–248

Tiebout, C. (1956). A pure theory of local expenditures. *Journal of Political Economy,* 64, 415–424

Waddell, P. (1996). *Accessibility and residential location: the interaction of workplace, residential mobility, tenure, and location choices*. Paper presented at the Lincoln Land Institute TRED Conference

# Chapter 20
# How do Socioeconomic Characteristics Interact with Equity and Efficiency Considerations? An Analysis of Hurricane Disaster Relief Goods Provision

**Mark W. Horner and Michael J. Widener**

**Abstract** Spatial analytic research has explored the issue of where to best site hurricane relief distribution facilities, but it has largely concentrated on the efficient provision of these services. However, equity considerations may also impact decisions on where to locate facilities. Questions of efficiency vs. equity become all the more acute when more detailed assessments of peoples' socio-economic characteristics are made as a part of these decisions. This paper examines the issue of siting hurricane disaster relief facilities based on equity vs. efficiency objectives, in light of populations' socioeconomic differences. Population differences are measured in terms of a household income variable. *p*-median and vertex *p*-center problems are applied to find relief center locations in a Southeastern U.S. city. Results show that income differences interact with the location strategies employed to produce variation in people's accessibility to relief goods.

**Keywords** Hurricanes · Distaster relief · Spatial model · GIS · Equity

## 20.1 Introduction

Researchers have given increased attention to issues of hurricane disaster relief and preparedness following extremely active hurricane seasons in 2004 and 2005. This is due in part to New Orleans' disastrous experience with Katrina in late summer 2005 and the widespread damage it and other storms caused throughout the southeastern U.S. (Cutter et al. 2006; Fu and Wilmot 2006; Horner and Downs 2007).

Transportation is a key concern for government agencies that plan for hurricane emergencies. While it is important that evacuation strategies be put

M.W. Horner (✉)
Department of Geography, The Florida State University, Tallahassee, FL 32306-2190, USA
e-mail: mhorner@fsu.edu

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_20,

into place so that people are compelled to leave affected regions in a safe and timely manner (Fu and Wilmot 2006; Lindell and Prater 2007), some people will choose to remain behind during a storm and not evacuate (Horner and Downs 2007), for reasons ranging from personal unfamiliarity with hurricane disasters and their potential threat (e.g., new residents), to people's desire to protect personal property (Baker 1991). In some cases, it may be that remaining behind is not a "choice," particularly for those who are elderly or lack personal transportation. With respect to New Orleans and Katrina, specifically, there were at least 100,000 people who remained behind, (Wolshon 2006), many of these presumably unable to evacuate. For those who do not leave the hurricane-affected area, governmental and humanitarian agencies must plan for providing relief services, including making basic provisions available such as food, water, and ice.

Planning for hurricane disaster relief goods transportation using spatial technologies and geographic information systems (GIS)-based modeling has great potential (Altay and Green 2006; Wright et al. 2006; Horner and Downs 2007). Research conducted to date has focused on siting relief distribution centers so that their services are made accessible to populations (Horner and Downs 2007, 2008a). The distribution strategies suggested to date have tended to emphasize the *efficiency* associated with siting relief facilities. Such modeling seeks to minimize the total transportation effort expended by all people in neighborhoods to reach relief facilities (Horner and Downs 2007, 2008a). From a system-wide perspective, this approach is logical as it ensures that people expend the smallest possible transport effort as overall accessibility is maximized.

However, from an *equity* standpoint (ReVelle and Eiselt 2005; Daskin 2008), such planning and modeling approaches may not be wholly satisfactory. From a system perspective, it may be impressive to design a facility location plan where for example "the average neighborhood is about 3.5 min from their nearest facility." Though if individual neighborhood's service levels are analyzed under such a plan, and some neighborhoods are over 15 min from their nearest facility, that plan would appear less impressive. Therefore, one way to avoid such wide disparities in service is to formally incorporate equity objectives when designing hurricane disaster relief distribution systems.

Besides the potential for different configurations of facilities to arise as a result of considering equity vs. efficiency strategies, there is also the question of how socioeconomic differences are best accounted for in the context of system planning. We know that certain populations are more vulnerable to disasters and may benefit greatly from increased access to goods and services (Fothergill et al. 1999). In fact, prior research has argued the importance of examining socioeconomic characteristics in the context of placing distaster relief facilities (Horner and Downs 2008a). Thus, discussions of efficiency and equity in hurricane relief goods distribution should not be separated from broader socio-economic issues.

In this chapter, we examine the implications of equity vs. efficiency based modeling strategies for siting hurricane disaster relief facilities. Moreover, we analyze how pursuing these contrasting objectives impacts socioeconomic groups' accessibility to relief distribution centers. We first begin by providing background that more formally establishes "equity" vs. "efficiency" as planning goals. From there, we move to discuss related transport literature on accessibility and transport modeling. Next, we provide an overview of a distribution system architecture, and follow that with a description of the spatial models we employ in our analysis. We then describe our study area and the details of our modeling scenarios, and present results. We conclude with a brief discussion and suggestions for future research.

## 20.2 Background and Literature

We review major strands of literature related to the issues of siting accessible facilities to provide hurricane disaster relief services. An important dichotomy can be drawn at the outset of this discussion between approaches aimed at making *private* as opposed to *public* facilities available. As the former are often associated with profit maximizing objectives (ReVelle et al. 1970), we shall work with models of the latter type, which are concerned with maximizing various forms of social welfare (Drezner 2004). This distinction notwithstanding, choosing the appropriate objective(s) for public facility location problems is not always clear and frequently involves many interrelated considerations (ReVelle and Eiselt 2005).

### *20.2.1 Efficiency Versus Equity?*

There are longstanding debates as to the appropriate objectives for addressing certain societal decision problems (Harsanyi 1975; Truelove 1993). Some public service provision questions are not always neatly dichotomized as simply efficiency vs. equity (DeVerteuil 2000; Drezner 2004) because other, interrelated objectives might also have bearing (ReVelle and Eiselt 2005; Daskin 2008). The issue of siting waste facilities, for example, certainly hinges on the efficient routing of waste trucks and pickup schedules. Commensurately, equitable levels of service may be a priority in the facility location plan, as there could be concern for ensuring the public is not unduly burdened by their siting in terms of air pollution, proximity to foul smells, etc. (ReVelle and Eiselt 2005). Clearly, public service location problems can take on a number of considerations. However, for purposes of analyzing hurricane service provision, limiting our discussion to contrasting efficiency vs. equity is an appropriate starting point.

The goal of hurricane disaster relief goods service provision is to make supplies available to people at a pre-determined number of distribution points located throughout the planning area. The questions involved then are (1) where to place distribution facilities, and (2) what objective(s) should be followed in making that decision. Some public facility location models may seek to make facilities accessible based on *efficiency* criteria that stress finding an "optimal" solution. Many of these models seek to minimize system transport costs, and as described by Morrill and Symons (Morrill and Symons 1977) this definition of efficiency is "independent of the distribution of benefits among individuals or locations." In the context of hurricane disaster relief, this means efficiency-maximizing approaches could result in some neighborhoods being disproportionately far from their nearest distribution center. If the socioeconomic characteristics of neighborhoods are also accounted for relative to facility location placements, efficiency based approaches may further exacerbate the equity situation (Marsh and Schilling 1994).

Equity stresses fairness, and according to Marsh and Schilling (1994) for public sector facility locations we must ask "are those who are affected by the decision treated fairly or equitably?" With respect to the hurricane distaster relief problem, the idea of several neighborhoods being 15 min or more from their nearest center is not palatable if many others are less than 7 min away from comparable facilities. Thus, the approach inherent to many equity-oriented facility location models is to try reducing the variance in the level of service demand units might receive (Drezner 2004; ReVelle and Eiselt 2005). In the present case, this would entail making decisions with an eye toward improving the plight of those neighborhoods that are farthest from relief centers (Marsh and Schilling 1994). Of course, such an equity-based facility strategy would not be as "optimal" from a systems perspective, as many neighborhoods with very good service under an efficiency scenario would likely have their accessibility degraded at the expense of improving service for more distant neighborhoods. Moreover, if socioeconomic concerns are accounted for in the modeling effort, the ability to achieve equity in facility location becomes all the more challenging (Marsh and Schilling 1994; ReVelle and Eiselt 2005).

In sum, it is important to point out that emphasizing either efficiency or equity will place relief facilities such that they may be accessed by neighborhood populations. However, an interest of this paper is the potentially differential levels of *accessibility* socioeconomic groups in neighborhoods may have to these centers after they are sited. To address this need we will turn to briefly reviewing some of the literature dealing with transport planning for accessibility.

### 20.2.2 Accessibility, Hurricane Relief Service Provision, and Transportation

There has been increased interest during the last several years in measuring accessibility, which is often defined as the ease with which people are able reach particular goods or services (Liu and Zhu 2004). However, accessibility issues have long been a part of broader transport debates on social exclusion (Preston and Rajé 2007), the spatial mismatch of workers and jobs (Schwanen 2003), and concerns about sprawl and land use imbalances (Cervero and Duncan 2006). Accessibility includes a land use dimension that accounts for the variability in the spatial distribution of people and activities (Scott and Horner 2008). In turn, it is the transportation system that provides the necessary linkages connecting origins and destinations (Liu and Zhu 2004).

Hurricane disaster relief is a question of accessibility when viewed as people utilizing the transport system to reach distribution centers. The approach adopted in this study is to take the transportation system as a given, and look at how changes in the distribution and arrangement of service centers affect people's accessibility to goods. We also wish to examine how this accessibility varies along socioeconomic dimensions. In this way, our study shares basic commonalities with efforts examining people's accessibility to activities such as shopping (Smoyer-Tomic et al. 2006), recreation (Talen and Anselin 1998), and medical and other services (Scott and Horner 2008), particularly given a context of recognizing accessibility disparities across socioeconomic groups (Horner and Mefford 2005).

## 20.3 Relief Distribution System Architectures

Comprehensive plans such as those written by the State of Florida (FDEM 2004, 2006) detail guidelines for conducting relief operations in the hurricane context. Improvements in hurricane warning and tracking systems in recent years (Elsner and Jagger 2006) increasingly provide better meteorological data that can be used to shape preparation efforts. As it has been detailed in other research (Horner and Downs 2007, 2008a, b), hurricane disaster relief goods distribution involves making available key supplies (e.g. food, water, ice, etc.) at designated distribution points in the planning area. Relief effort preparation makes use of advanced warnings and sets up distribution points before the onset of a storm, so that they are prepared to distribute supplies to people after the system passes. Placing distribution centers in proximity to populations is a necessity.

There are many potential designs for a hurricane disaster relief goods distribution system. Generally, goods flow from larger warehouses, of which there are relatively few in a region, to distribution points, which would be more numerous and act as the access points where people obtain goods (Horner and Downs 2007). From the resident's perspective, it is the linkage between their neighborhood and the nearest relief point that is most pressing. Although modeling efforts can account for the movement of goods along the full supply chain, some research has argued that the accessibility of neighborhoods to distribution points be given the most weight (Horner and Downs 2008b).

In this research, we focus specifically on the neighborhood-distribution facility linkage, and as such, pursue the requisite modeling approaches. We do not include the warehouse to distribution linkage as a formal component of our modeling. Doing so would detract from our goal of exploring the possible differential socioeconomic impacts of pursuing efficiency vs. equity scenarios. The tradeoffs involved with emphasizing the importance of one part of the supply chain over another have been explored elsewhere (Horner and Downs 2007). There are also site design issues in terms of the characteristics of the relief facilities themselves, which could be considered (FDEM 2004, 2006). In this research, we assume a single type of uncapacitated distribution facility. We do not explore the impacts of siting various types of relief facilities, which may vary in terms of their capacity, range of goods/services available, etc. (FDEM 2006). Readers interested in the implications of variable facility types for providing disaster relief are referred elsewhere (Horner and Downs 2008b).

Given this scope, operationalizing "efficiency" vs. "equity" becomes a question of placing relief facilities relative to neighborhoods, given their socioeconomic makeup. In the next section, we describe two modeling approaches we use to implement facility siting based on efficiency and equity criteria.

### *20.3.1 Model Overview*

The $p$-median model is a classical approach to efficiently allocating services. It seeks to minimize the total transportation costs of providing services to populations by siting accessible facilities on networks. Applications of this model range from locating medical centers to refueling facilities (Oppong 1996; Nicholas et al. 2004). The $p$-median formulation presented below is adapted from several sources (ReVelle and Swain 1970; Daskin 1995; Miller and Shaw 2001) and is given by

Minimize

$$\sum_{i \in I} \sum_{j \in J} a_i c_{ij} x_{ij} \tag{20.1}$$

Subject to

$$\sum_{j \in J} x_{ij} = 1 \quad \forall i \in I \tag{20.2}$$

$$\sum_{j \in J} x_j = p \tag{20.3}$$

$$x_{ij} - x_j \geq 0 \quad \forall i \in I, \forall j \in J \tag{20.4}$$

$$x_{ij}, x_j \in (0, 1) \quad \forall i \in I, \forall j \in J \tag{20.5}$$

where

$i$ = index of all neighborhood locations in $I$
$j$ = index of all distribution center candidate locations in $J$
$a_i$ = demand for relief services at neighborhood $i$
$c_{ij}$ = transportation costs between neighborhood $i$ and distribution candidate site $j$
$p$ = user defined number of distribution facilities to be sited
$x_{ij}$ = 1 if neighborhood $i$ is served by distribution facility $j$, 0 otherwise
$x_j$ = 1 if a distribution facility is sited at candidate site $j$, 0 otherwise

The objective function to be minimized in Equation (20.1) captures total interaction costs between neighborhoods and distribution centers. Optimal total transportation costs output from the $p$-median problem may be divided by the total number of interactions $\Sigma_i \Sigma_j x_{ij}$ to obtain *average* transport costs. Constraints in Equation (20.2) ensure that each neighborhood is assigned to exactly one sited distribution facility. Equation (20.3) stipulates that exactly $p$ distribution centers are to be sited. Per Equation (20.4) neighborhoods are only able to be serviced by sited distribution centers. Binary integer bounds on the decision variables are given by Equation (20.5).

The other modeling approach used in this research is the $p$-center problem. It is considered an equity-oriented model because it seeks to minimize the maximum transport cost between sited distribution facilities and neighborhoods. That is, it places facilities so that the final configuration of centers is the best in terms of the most inaccessible neighborhood overall. It is often used in similar modeling situations of the $p$-median problem (Daskin 1995; Cheng et al. 2007). The model formulation is

Minimize

$$W \tag{20.6}$$

Subject to

$$\sum_{j \in J} x_{ij} \quad \forall i \in I \tag{20.7}$$

$$\sum_{j \in J} x_j = p \tag{20.8}$$

$$x_{ij} - x_j \geq 0 \quad \forall i \in I, \forall j \in J \tag{20.9}$$

$$W \geq \sum_{j \in J} c_{ij} x_{ij} \quad \forall i \in I \tag{20.10}$$

$$x_{ij}, x_j \in (0,1) \quad \forall i \in I, \forall j \in J \tag{20.11}$$

where, $W =$ maximum cost (e.g. miles, time) of accessing a relief center from a neighborhood, with all other notation as defined previously.

The objective to be minimized (20.6) is the maximum transport cost between a neighborhood and its closest distribution facility. Constraints in (20.7–20.9) function identically as they do in the previously described $p$-median formulation. The one additional constraint in (20.10) ensures that the maximum cost $W$, is greater than the interaction costs between any neighborhood and the relief center that services it.

Scanning the models' descriptions and structures, it is apparent that they may be used in related modeling scenarios, with the key factor in choosing one over the other being whether pursuing efficiency ($p$-median) or equity ($p$-center) is of greatest concern. Of interest in the present research is exploring how utilizing these modeling approaches might impact service provision to various socioeconomic groups.

## 20.4 Analysis and Results

We now turn to describing our analysis and results. The data sources we use are discussed briefly below, as well as in previous work by Horner and Downs (2008a), who also looked at socioeconomic issues in the same study area. Horner and Downs (2008a) focused on how assorted parameterizations of a commodity flow model affected socioeconomic groups' accessibility to relief centers. In contrast, this research explores whether modeling approaches favoring equity vs. efficiency translate into divergent service outcomes for various population groups.

### *20.4.1 Study Area and Data*

Leon County, Florida, which contains the city of Tallahassee, was selected as our study area. As the neighborhood is considered as the basis for providing service, we used Leon County's 594 traffic analysis zones (TAZs) from the Census Transportation Planning Package (2000) as representative units for our analysis. Similar to Horner and Downs (2008a), we use household counts by TAZ as the basis for demand, and disaggregate these households by their total incomes. We do not control for household size membership. The source of the data is "Table 66" from Part 1 of the CTPP. Table 66 tabulates households by TAZ according to several income calculations. The income ranges for each class are narrow enough such that it is easy to separate households into low, medium, and high income classes. According to these data, there were 95,637 households present in Leon County (2000). Horner and Downs (2008a) reported some detailed tabulations of these data. For purposes of this study, it is noteworthy that approximately 20% of households (20,333 – 21.26%) have incomes of $14,999 or less. At the other end of the spectrum, roughly 20% of households (19,141 – 20.01%) have incomes of $75,000 or more. Our use of income to stratify demand is based on an assumption that it directly proxies a household's socioeconomic characteristics. Both sets of income counts along with the Leon County study area are shown in Fig. 20.1. Lastly, road

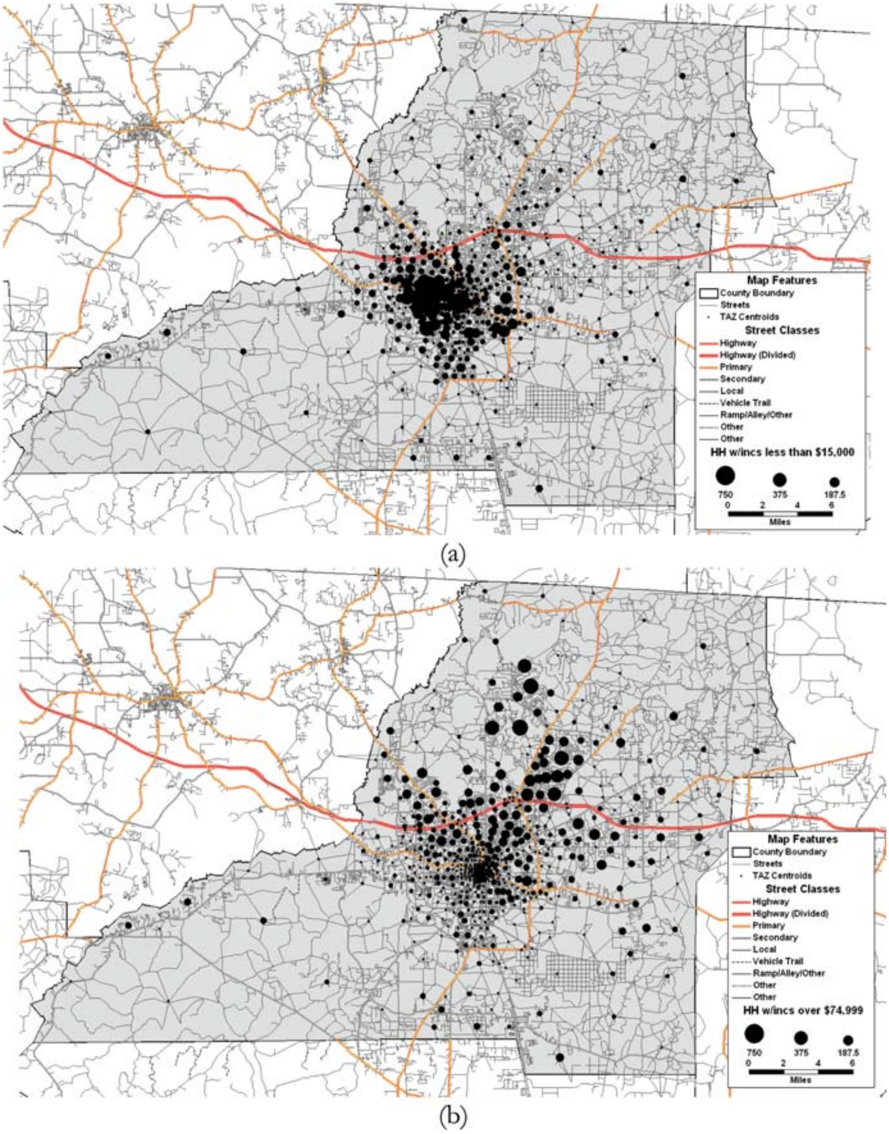



**Fig. 20.1** Leon county study area and lower/upper income counts of households by TAZ: (**a**) Counts of households by TAZ with incomes less than or equal to $14,999; (**b**) Counts of households by TAZ with incomes greater than $74,99

infrastructure data were taken from the US Census TIGER files from the year 2002, which contain the largest roads such as Interstate Highways, as well as smaller-scale features like local neighborhood streets.

## 20.4.2 Computational Environment and Analytical Setting

The analysis was conducted on a dual core 3.0 GHZ personal computer with 3.25 GB of memory running a standard commercial operating system. Procedures available within the GIS (TransCAD vs. 4.7) were used to find solutions for all facility location problems. The GIS uses implementations of an interchange heuristic (Teitz and Bart 1968; Mirchandani and Francis 1990) to find near-optimal solutions to the location models. The advantage of using the GIS in this capacity is the relative ease of having the spatial data and solution procedures tightly coupled, which allows users to solve a wide range of problems quickly, with minimal user overhead.

### 20.4.2.1 Measuring Transport Costs from Neighborhoods to Distribution Centers

Using the street network from the TIGER line file, network travel times were estimated between each neighborhood and all distribution candidate locations. The streets file contains functional class information, which is used to assign speeds to each of the road segments. With speeds present on all roadway linkages, we calculated free-flow travel times for each segment, then added a delay factor of 30% to account for delays due to stoplights, turns, etc. Similar to Horner and Downs (2008a) additional congestion impacts were not included in the modeling, based on the assumption that roads would have adequate capacity following the passing of the weather system. We also assumed that all roadways would be open and did not subject any of our model runs to cases where particular streets were shut down.

## 20.4.3 Efficiency and Equity Analysis

In this section, we describe the results of our modeling efforts with the two representative spatial models for equity ($p$-center) and efficiency ($p$-median). Each of the models was solved using the same specifications of $p$, ranging from $p = 1$ relief distribution facility, to $p = 35$ facilities, in various increments. Given the size of the Leon County study area, a maximum $p$ of 35 represents an estimated ceiling of approximately how many facilities could be used in practice, though it is difficult to be exact given that they are uncapacitated. At the other end of the spectrum, clearly relief agencies would activate more than one relief facility ($p = 1$). Nonetheless, the model runs with lower values of $p$ are included for sake of completeness and to assist in interpreting trends in the numerical results. TAZ centroids are used as the candidate locations for relief distribution facilities.

Each of the modeling runs makes use of the same household definitions and income breaks as taken from the CTPP TAZ. However, these data are used differently in the models. As the classic *p*-center problem is an "unweighted" model, the size of the household counts in a particular income class does not impact facility placement. In the *p*-median model, zones with larger counts of households in a particular class, ceteris paribus, will compel solution algorithms to place facilities closer to them, thereby maximizing access. It is important to note that we do not reduce the number of households under consideration due to likely evacuations. As evacuation estimates are often performed based on assuming a constant rate of departure across all neighborhood units (e.g. 60%) (FDEM 2004; Horner and Downs 2008b), bypassing this step does not impact the analytical results.

#### 20.4.3.1 Equity analysis with the *p*-center problem

Table 20.1 reports our modeling results with the *p*-center problem. Problems are solved for 1–35 sited facilities (columns), and for 13 income definitions (rows). Statistics are reported on the average time zones are from their nearest facilities (weighted by the number of households of a particular group in the zone), the maximum time any zone is from its nearest facility, and the standard deviation of the travel time from the zones to their nearest facility.

Operationalizing the household income definitions for use in the *p*-center problem was necessary as it is not naturally demand-sensitive. In our analysis, if a zone contained at least one household in a particular income category, then that zone was considered in need of service and it was entered into the model. For example, the first income class contains those households with incomes up to and including $9,999. If a zone contained at least one household with an income at or below $9,999, it was deemed in need of service.

As the *p*-center problem is primarily concerned with minimizing the maximum time any one zone is away from their nearest sited facility, the primary statistic of interest is the maximum time. In the case of $p = 1$ the maximum time is about 43–46 min. When the study area is more saturated with facilities at $p = 35$, the maximum time any one zone is from their nearest facility is slightly more than 9 min. Consistent with other facility location studies, there are diminishing returns in terms of adding new facilities (Daskin 1995). That is, adding additional facilities to systems of relatively few facilities has a greater impact on decreasing service times than the addition of new facilities to systems with an already large number of facilities. To illustrate, increasing *p* from 1 to 5 decreases the maximum minutes traveled on the order of over 20 min for all 13 income definitions, while increasing *p* from 30 to 35 reduces times from the most distant facility from around 10.1 to about 8.8 min.

Holding *p* constant and exploring the effects of income definition changes on service provision, it is noted that there are not substantial changes in the corresponding maximum time statistics. Looking down any given column, the

**Table 20.1** Facility location solutions based on *p*-center (equity) objective

| | | *p* | 1 | 5 | 9 | 15 | 20 | 25 | 30 | 35 |
|---|---|---|---|---|---|---|---|---|---|---|
| Counts of HH with incomes up to, and including | $ 9,999 | Avg. min | 14.900 | 8.847 | 8.327 | 7.483 | 5.223 | 5.463 | 5.148 | 3.922 |
| | | Max. min | 43.287 | 21.922 | 17.492 | 13.971 | 11.292 | 9.099 | 8.923 | 8.050 |
| | | Std. dev. | 7.535 | 4.980 | 3.546 | 2.988 | 2.781 | 2.242 | 2.310 | 2.102 |
| | $ 14,999 | Avg. min | 15.105 | 12.663 | 12.041 | 7.379 | 8.009 | 6.879 | 5.461 | 4.737 |
| | | Max. min | 43.287 | 22.751 | 16.517 | 13.834 | 12.597 | 10.609 | 8.923 | 8.203 |
| | | Std. dev. | 7.506 | 4.625 | 3.342 | 2.903 | 2.892 | 2.45 | 2.285 | 2.147 |
| | $ 19,999 | Avg. min | 15.219 | 8.666 | 7.307 | 8.036 | 8.249 | 6.679 | 6.420 | 5.386 |
| | | Max. min | 43.673 | 23.510 | 17.028 | 13.753 | 11.841 | 10.998 | 9.832 | 8.775 |
| | | Std. dev. | 7.636 | 7.636 | 3.499 | 2.941 | 2.765 | 2.504 | 2.397 | 2.289 |
| | $ 24,999 | Avg. min | 15.507 | 9.033 | 7.389 | 8.545 | 5.971 | 6.641 | 6.259 | 5.236 |
| | | Max. min | 43.673 | 23.510 | 17.028 | 13.753 | 13.011 | 10.998 | 9.832 | 8.081 |
| | | Std. dev. | 7.604 | 5.160 | 3.526 | 2.982 | 3.253 | 2.497 | 2.431 | 2.130 |
| | $ 29,999 | Avg. min | 15.717 | 9.250 | 7.573 | 8.529 | 6.063 | 6.630 | 5.496 | 5.382 |
| | | Max. min | 43.673 | 23.802 | 17.089 | 13.753 | 13.011 | 10.998 | 9.787 | 8.775 |
| | | Std. dev. | 7.620 | 5.238 | 3.605 | 3.027 | 3.260 | 2.512 | 2.319 | 2.269 |
| | $ 34,999 | Avg. min | 12.363 | 8.875 | 7.853 | 7.346 | 6.846 | 7.402 | 5.845 | 4.406 |
| | | Max. min | 45.932 | 24.215 | 17.089 | 14.064 | 12.984 | 10.998 | 10.132 | 8.775 |
| | | Std. dev. | 8.070 | 4.152 | 3.458 | 3.318 | 2.886 | 2.714 | 2.363 | 2.256 |
| | $ 39,999 | Avg. min | 12.537 | 8.921 | 7.878 | 7.350 | 6.889 | 7.345 | 5.820 | 4.423 |
| | | Max. min | 45.932 | 24.215 | 17.089 | 14.064 | 12.984 | 10.998 | 10.132 | 8.775 |
| | | Std. dev. | 8.036 | 4.153 | 3.442 | 3.307 | 2.877 | 2.719 | 2.366 | 2.256 |
| | $ 44,999 | Avg. min | 12.789 | 8.976 | 7.939 | 7.378 | 6.944 | 7.265 | 5.783 | 5.472 |
| | | Max. min | 45.932 | 24.215 | 17.089 | 14.064 | 12.984 | 10.998 | 10.132 | 8.775 |
| | | Std. dev. | 8.021 | 4.159 | 3.436 | 3.325 | 2.890 | 2.715 | 2.372 | 2.261 |
| | $ 49,999 | Avg. min | 12.996 | 9.013 | 7.979 | 7.410 | 6.963 | 7.215 | 5.766 | 5.447 |
| | | Max. min | 45.932 | 24.215 | 17.089 | 14.064 | 12.984 | 10.998 | 10.132 | 8.775 |
| | | Std. dev. | 8.023 | 4.148 | 3.430 | 3.324 | 2.881 | 2.707 | 2.375 | 2.259 |

**Table 20.1** (continued)

| | *p* | 1 | 5 | 9 | 15 | 20 | 25 | 30 | 35 |
|---|---|---|---|---|---|---|---|---|---|
| $ 59,999 | Avg. min | 13.293 | 9.066 | 8.056 | 6.894 | 7.022 | 6.502 | 5.746 | 5.166 |
| | Max. min | 45.932 | 24.215 | 17.089 | 14.064 | 12.984 | 11.174 | 10.132 | 8.835 |
| | Std. dev. | 8.081 | 4.171 | 3.437 | 3.245 | 2.891 | 2.539 | 2.394 | 2.258 |
| $ 74,999 | Avg. min | 13.677 | 9.136 | 9.537 | 6.974 | 7.069 | 6.484 | 6.155 | 5.166 |
| | Max. min | 45.932 | 24.215 | 18.046 | 14.064 | 12.984 | 11.174 | 10.132 | 8.941 |
| | Std. dev. | 8.068 | 4.153 | 3.638 | 3.261 | 2.910 | 2.541 | 2.496 | 2.300 |
| $ 99,999 | Avg. min | 13.958 | 9.182 | 9.625 | 7.055 | 7.111 | 6.497 | 6.122 | 5.155 |
| | Max. min | 45.932 | 24.215 | 18.046 | 14.064 | 12.984 | 11.174 | 10.132 | 8.941 |
| | Std. dev. | 8.095 | 4.144 | 3.649 | 3.263 | 2.900 | 2.532 | 2.488 | 2.297 |
| $ 100,000 | Avg. min | 14.323 | 12.605 | 9.739 | 7.956 | 7.177 | 6.201 | 6.097 | 5.172 |
| | Max. min | 45.932 | 25.915 | 18.046 | 13.834 | 11.909 | 10.998 | 10.132 | 9.281 |
| | Std. dev. | 8.185 | 5.501 | 3.649 | 3.015 | 2.747 | 2.824 | 2.497 | 2.348 |

maximum time statistics typically vary in a range of 1–2 min. This stability in the solutions is due to the way in which the income definitions were operationalized. Inclusion of a zone in the model occurs as long as the zone has at least one household belonging to the particular income under consideration. Although zones may be constructed based on neighborhood and economic characteristics, they are still aggregate spatial units, and thus likely to contain many different types of households (Horner and Murray 2002). In other words, the picture of households to be served, spatially, at income $44,999 for example, is not that dissimilar from the more inclusive upper income categories.

#### 20.4.3.2 Efficiency Analysis with the *p*-median Problem

The results of siting disaster relief facilities with the *p*-median location model are shown in Table 20.2. As the *p*-median model is concerned with minimizing system wide costs of neighborhoods accessing relief centers, a key statistic of interest is the average service time. Adding relief facilities decreases (at a decreasing rate) the average time a neighborhood is from its nearest facility. When $p = 1$, average times range from 6.7 to about 10.9 min depending on the demand specification. At the other end of the spectrum, for $p = 35$, average times range from 1.5 to 2.7 min.

In contrast with our *p*-center problem, which is not demand weighted, the *p*-median model is demand weighted and alternative demand specifications clearly have an effect on the average travel time to relief facilities. Expanding the definition of who will need service can greatly influence average time. The case of $p = 15$, for example, demonstrates that broadening the service population definition from only the households with incomes of $9,999 or less to all households increases average time (2.7 vs. 4.1 min). It also results in the maximum time any one neighborhood is from a facility increasing from 23.7 min to almost 29 min. What this shows is the potentially significant effect of demand specification on the efficient provision of relief services. It is suggestive of the spatial variation in the distribution of Leon County's high and low income households.

#### 20.4.3.3 Comparisons of *p*-median and *p*-center Results

Though the *p*-center and *p*-median pursue different objectives during facility location, it is possible to compare how they provide service. If we focus on the case of $p = 20$ for both models and the situations where the demand is all households with incomes up to $24,999, and all households regardless of income, there are some interesting contrasts that may be drawn between "efficiency" and "equity". In both income cases, the average travel time under the *p*-center scenario is substantially greater than the corresponding *p*-median average. For households making $24,999 or less, under the *p*-center solution, the average neighborhood is more than double the travel time from its nearest relief center as opposed to the *p*-median's configuration of sites (5.97 vs. 2.74).

**Table 20.2** Facility location solutions based on *p*-median (efficiency) objective

| | | *p* | 1 | 5 | 9 | 15 | 20 | 25 | 30 | 35 |
|---|---|---|---|---|---|---|---|---|---|---|
| Counts of HH with incomes up to, and including | $ 9,999 | Avg. min | 6.755 | 4.352 | 3.444 | 2.675 | 2.231 | 1.893 | 1.662 | 1.485 |
| | | Max. min | 49.798 | 48.372 | 31.261 | 23.662 | 24.185 | 23.662 | 18.562 | 18.562 |
| | | Std. dev. | 8.608 | 6.588 | 5.457 | 4.498 | 3.968 | 3.867 | 3.537 | 3.174 |
| | $ 14,999 | Avg. min | 7.247 | 4.660 | 3.650 | 2.859 | 2.474 | 2.129 | 1.863 | 1.683 |
| | | Max. min | 50.074 | 46.417 | 31.261 | 24.388 | 24.388 | 23.514 | 18.562 | 18.562 |
| | | Std. dev. | 8.583 | 6.847 | 5.472 | 4.653 | 4.030 | 3.597 | 3.436 | 3.478 |
| | $ 19,999 | Avg. min | 7.581 | 4.876 | 3.854 | 3.032 | 2.578 | 2.259 | 2.014 | 1.806 |
| | | Max. min | 51.650 | 47.664 | 31.000 | 24.388 | 24.388 | 19.809 | 19.809 | 18.516 |
| | | Std. dev. | 8.776 | 7.087 | 5.250 | 4.609 | 4.084 | 3.582 | 3.638 | 3.420 |
| | $ 24,999 | Avg. min | 8.039 | 5.150 | 4.146 | 3.217 | 2.744 | 2.419 | 2.178 | 1.971 |
| | | Max. min | 51.879 | 47.664 | 31.000 | 24.388 | 24.388 | 19.809 | 19.809 | 18.516 |
| | | Std. dev. | 8.703 | 6.787 | 5.051 | 4.126 | 3.824 | 3.490 | 3.505 | 3.332 |
| | $ 29,999 | Avg. min | 8.372 | 5.377 | 4.270 | 3.349 | 2.828 | 2.513 | 2.257 | 2.032 |
| | | Max. min | 52.339 | 47.644 | 31.000 | 24.388 | 24.388 | 19.809 | 18.516 | 18.516 |
| | | Std. dev. | 8.795 | 7.047 | 5.281 | 4.093 | 3.831 | 3.610 | 3.471 | 3.374 |
| | $ 34,999 | Avg. min | 8.601 | 5.497 | 4.370 | 3.443 | 2.917 | 2.588 | 2.320 | 2.100 |
| | | Max. min | 52.339 | 40.083 | 31.000 | 24.388 | 24.388 | 18.516 | 18.516 | 18.516 |
| | | Std. dev. | 8.954 | 6.592 | 5.283 | 4.077 | 3.848 | 3.521 | 3.543 | 3.393 |
| | $ 39,999 | Avg. min | 8.793 | 5.589 | 4.476 | 3.526 | 2.993 | 2.636 | 2.390 | 2.175 |
| | | Max. min | 52.466 | 47.664 | 31.000 | 24.388 | 24.388 | 23.165 | 18.516 | 18.516 |
| | | Std. dev. | 8.904 | 6.954 | 5.253 | 4.077 | 3.848 | 3.672 | 3.543 | 2.985 |
| | $ 44,999 | Avg. min | 9.345 | 5.771 | 4.618 | 3.642 | 3.125 | 2.740 | 2.461 | 2.266 |
| | | Max. min | 52.466 | 47.664 | 31.459 | 24.388 | 24.388 | 23.165 | 23.165 | 18.516 |
| | | Std. dev. | 8.866 | 6.932 | 5.138 | 3.915 | 3.835 | 3.668 | 3.395 | 2.979 |
| | $ 49,999 | Avg. min | 9.345 | 5.916 | 4.716 | 3.718 | 3.164 | 2.797 | 2.511 | 2.301 |
| | | Max. min | 54.466 | 47.664 | 31.459 | 24.388 | 23.165 | 23.165 | 23.165 | 19.739 |
| | | Std. dev. | 8.835 | 6.893 | 4.942 | 3.904 | 3.735 | 3.378 | 3.352 | 3.116 |

**Table 20.2** (continued)

| | *p* | 1 | 5 | 9 | 15 | 20 | 25 | 30 | 35 |
|---|---|---|---|---|---|---|---|---|---|
| $ 59,999 | Avg. min | 9.674 | 6.109 | 4.856 | 3.847 | 3.251 | 2.884 | 2.611 | 2.391 |
| | Max. min | 52.466 | 47.664 | 31.459 | 24.388 | 23.165 | 23.165 | 23.165 | 23.165 |
| | Std. dev. | 8.908 | 6.947 | 4.996 | 3.981 | 3.794 | 3.418 | 3.413 | 3.425 |
| $ 74,999 | Avg. min | 10.100 | 6.362 | 5.027 | 3.898 | 3.370 | 3.006 | 2.724 | 2.484 |
| | Max. min | 52.466 | 47.664 | 31.459 | 24.673 | 23.165 | 23.165 | 23.165 | 23.165 |
| | Std. dev. | 8.881 | 6.920 | 4.982 | 3.983 | 3.475 | 3.404 | 3.419 | 3.364 |
| $ 99,999 | Avg. min | 10.467 | 6.555 | 5.161 | 4.052 | 3.489 | 3.111 | 2.830 | 2.581 |
| | Max. min | 52.488 | 47.664 | 31.858 | 28.906 | 23.165 | 23.165 | 23.165 | 21.935 |
| | Std. dev. | 8.854 | 6.921 | 5.130 | 4.141 | 3.476 | 3.415 | 3.387 | 3.206 |
| $ 100,000 | Avg. min | 10.876 | 6.796 | 5.307 | 4.152 | 3.624 | 3.242 | 2.933 | 2.676 |
| | Max. min | 52.488 | 47.664 | 31.858 | 28.906 | 23.165 | 23.165 | 20.631 | 20.631 |
| | Std. dev. | 8.906 | 6.977 | 5.218 | 4.111 | 3.742 | 3.480 | 3.212 | 3.202 |

With respect to the maximum time statistic, those are again almost mirror opposites from one model to the other. Looking at $p = 20$, where the demand scenario considers all households, the maximum time for the $p$-median solution is 23.17 min – almost double the corresponding $p$-center problem time of 11.09.

Clearly the models are performing as they should, but the wide range of objective function values raises interesting questions about how "equitable" or "efficient" these facility location configurations are for particular population groups. If we map the solutions for $p = 15$ (Fig. 20.2), focusing on households with the lowest 20% of incomes (those making less than $15,000) it becomes apparent that there are substantial facility location changes under the two different objectives. The $p$-center solution seems to favor a more dispersed pattern, while the $p$-median facilities are more centralized. This is interesting and consistent with the prior figure (Fig. 20.1) that showed higher concentrations of lower income households in the central portions of the study areas. The $p$-median solutions are clearly affected by the concentration of low income households. This issue can be explored in greater detail by examining the geography of the facility locations versus the various parameters that were used to find them, and the people who are assumed to be in need of service.

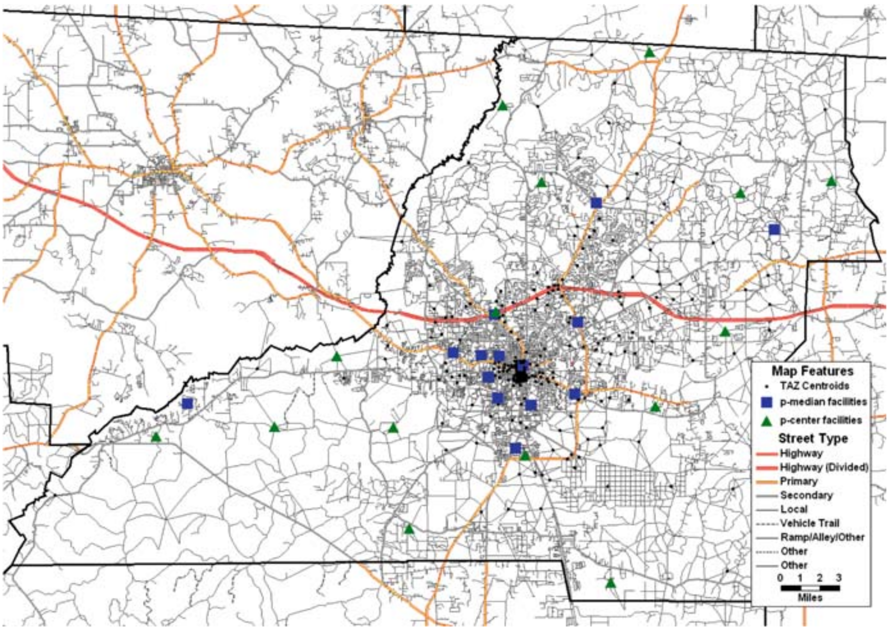



**Fig. 20.2** Comparison of facility locations, $p = 15$, median vs. center, households with incomes less than or equal to $14,999 used as demand basis

#### 20.4.3.4 Exploring Service Differences Across Income Groups

We wanted to explore how particular facility configurations and demand specifications affected the "efficiency" and "equity" associated with service provision to different socioeconomic groups. Given discussion suggesting that relief need is tied to socioeconomic and income characteristics (Cutter et al. 2006), we chose to examine the two ends of the household income spectrum; those households with incomes of $14,999 or less, and those households with incomes of $75,000 or more. As previously mentioned when introducing the dataset, these class breaks roughly correspond to the upper and lower 20% of household incomes. In each of the queries, we considered the location statistics based on $p = 15$, 25, and 35. Results appear in Table 20.3.

**Table 20.3** Service difference comparisons for high and low income groups

*p*-center problem – all nonzero TAZs served

| | Lower 20th income percentile | | | Upper 20th income percentile | | |
|---|---|---|---|---|---|---|
| | $p = 15$ | $p = 25$ | $p = 35$ | $p = 15$ | $p = 25$ | $p = 35$ |
| Avg. min | 8.056 | 5.713 | 5.555 | 8.033 | 6.535 | 5.025 |
| Max. min | 13.834 | 10.780 | 9.201 | 13.834 | 10.911 | 9.281 |
| Min. min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

*p*-median problem – demand is HH with incomes < = $14,999

| | Lower 20th income Percentile | | | Upper 20th income percentile | | |
|---|---|---|---|---|---|---|
| | $p = 15$ | $p = 25$ | $p = 35$ | $p = 15$ | $p = 25$ | $p = 35$ |
| Avg. min | 2.859 | 2.129 | 1.683 | 6.774 | 5.624 | 5.392 |
| Max. min | 24.390 | 23.510 | 18.560 | 25.780 | 23.510 | 18.520 |
| Min. min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

*p*-median problem – demand is all HH

| | Lower 20th income percentile | | | Upper 20th income percentile | | |
|---|---|---|---|---|---|---|
| | $p = 15$ | $p = 25$ | $p = 35$ | $p = 15$ | $p = 25$ | $p = 35$ |
| Avg. min | 3.494 | 2.467 | 2.185 | 4.935 | 4.100 | 3.162 |
| Max. min | 27.250 | 23.160 | 20.630 | 27.250 | 23.160 | 20.630 |
| Min. min | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

First, we examined solutions to the *p*-center problem that were solved based on serving all TAZs with at least one household (these solutions would correspond to the last row in the bottom of Table 20.1). Based on the query results, both the upper and lower income households have equitable levels of service under this approach. The equity measure, the maximum number of minutes, is virtually the same across both groups for each of the three facility location scenarios. For example, when $p = 25$, the worst-case for both the lower and upper income zones is slightly less than 11 min away from their nearest relief facility. Similar trends emerge if the average time to the nearest facility is examined; the average times are comparable across groups for the three facility scenarios.

We also looked at how well the low and high income percentiles fared under two *p*-median demand scenarios. The first is when demand is assumed to be households making $14,999 or less (this corresponds to the second set of trials (rows) in Table 20.2). The results of the query naturally give the statistics of interest for the reference group (HH < $14,999). With further work, we may also determine how well this configuration of facilities serves those in the upper income percentile. Perhaps unsurprisingly, the average time statistics are much lower for the reference group in all facility location cases. In the case of $p = 35$, for example, the average service time for the lower income group is less than a third of the high income group (1.68 vs. 5.39). From an equity standpoint, both groups have similar service as measured by their maximum statistics.

When the *p*-median problem is solved using all households as the demand, the nature of the efficiency changes substantially for the two groups. Lower income households are still more efficiently served than their higher income counterparts, but not as well served under the prior *p*-median scenario when they are the sole weight for demand. On the one hand, when all incomes are considered, the average times for lower incomes increase by some fraction of a minute in all three facility location specifications. On the other hand, because their spatial configuration factor into facility locations, the higher income households are better served under these scenarios; actually more than 2 min better in the case of $p = 35$ (5.39 vs. 3.16). Maximum time statistics are very close between the two groups.

These results suggest several interesting interactions between efficiency and equity goals. One is that a consistent level of equity can be achieved across the two income percentiles while solving the model based on efficiency objectives. The first set of queried *p*-median solutions favor efficiency for the lower income percentile, but not at the expense of equity for both groups. In the second set of queried *p*-median solutions, the level of equity is comparable between groups, though the equity quality is degraded from the prior solution set in terms of both groups experiencing higher maximum times. With respect to comparing the *p*-center solutions to those found using the *p*-median problem, it is noteworthy how much efficiency suffers (in both *p*-median scenarios) for the lower income group in order to achieve greater equity. Average time to nearest facility ranges from about 5.5 to a little over 8 min in the three *p*-center runs. When that is contrasted with the substantially lower average times found in the *p*-median runs, it is apparent that the *p*-center is making service worse for a lot of neighborhoods in order to improve the situation for those that are most inaccessible.

## 20.5 Discussion and Conclusions

Recognizing the need for new strategies that provide accessible hurricane disaster relief, this paper has explored issues involving the potential conflicts of efficiency vs. equity objectives in goods distribution center location decisions.

As we have discussed, populations' socioeconomic characteristics have bearing on how services should best be provided. Our analytical results show how changes in the socioeconomic representation of demand impacts overall goods accessibility. Results point out the role of population heterogeneity (as represented by data on household incomes) in influencing appropriate facilities placement.

There are many opportunities for future research. We will mention three broad possibilities. First, additional efforts may be spent exploring other socio-economic variables to stratify relief service demand. By considering other variables in location decisions (e.g. auto ownership, poverty status, etc.) it is possible that new geographies of service accessibility would emerge. Second, future work might consider the tradeoffs between efficiency and equity in a more subtle fashion, perhaps by treating the problem as multiobjective in nature, rather than limiting location decisions to being a function of one or the other. Such multiobjective modeling is commonplace in transport and location studies (ReVelle and Eiselt 2005). Lastly, new work could look to develop alternative pictures of the demand for relief goods in terms of the complex interactions between socioeconomic status and the propensity to evacuate. As our research used a constant rate of evacuation that was applied to non-emergency derived census population counts, future efforts could impart greater realism by incorporating actual evacuation data sources. Pursuing these and other opportunities will better enhance our understanding of the best facility location strategies for designing effective hurricane relief distribution systems.

**Acknowledgments** This chapter is based upon work supported by the U.S. National Science Foundation under Grant No. (BCS-0550330) awarded to the first author (Horner). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

## References

Altay, N. & Green, W.G. (2006). OR/MS research in disaster operations management. *European Journal of Operational Research,* 175(1), 475–493

Baker, E.J. (1991). Hurricane evacuation behavior. *International Journal of Mass Emergencies and Disasters,* 9(2), 287–310

Cervero, R. & Duncan, M. (2006). Which reduces vehicle travel more: jobs-housing balance or retail-housing mixing? *Journal of the American Planning Association,* 72(4), 475–490

Cheng, T.C.E., Kang, L. & Ng, C.T. (2007). An improved algorithm for the *p*-center problem on interval graphs with unit lengths. *Computers & Operations Research,* 34(8), 2215

Cutter, S.L., Emrich, C.T., Mitchell, J.T., Boruff, B.J., Gall, M., Schmidtlein, M.C., Burton, C.G. & Melton, G. (2006). The long road home: Race, class, and recovery from Hurricane Katrina. *Environment,* 48(2), 8–20

Daskin, M.S. (1995). *Network and discrete location: models, algorithms, and applications.* (New York: John Wiley and Sons)

Daskin, M.S. (2008). What you should know about location modeling. *Naval Research Logistics,* 55(4), 283–294

DeVerteuil, G. (2000). Reconsidering the legacy of urban public facility location theory in human geography. *Progress in Human Geography,* 24(1), 47–69

Drezner, T. (2004). Location of casualty collection points. *Environment and Planning C-Government and Policy,* 22(6), 899–912

Elsner, J.B. & Jagger, T.H. (2006). Comparison of hindcasts anticipating the 2004 Florida Hurricane season. *Weather and Forecasting,* 21(2), 182–192

Florida Division of Emergency Management (FDEM). (2004). *Florida Comprehensive Emergency Management Plan*. Available on http://floridadisaster.org/documents/CEMP/floridaCEMP.htm.

Florida Division of Emergency Management (FDEM). (2006). *Florida Comprehensive Emergency Management Plan County Logistics Planning Standard - Operating Guideline (CEMP 2355)*. Available on http://floridadisaster.org/bpr/Response/CIandLOG.htm

Fothergill, A., Maestas, E.G.M. & Darlington, J.D. (1999). Race, ethnicity and disasters in the United States: a review of the literature. *Disasters,* 23(2), 156–173

Fu, H. Q. & Wilmot, C.G. (2006). Survival analysis-based dynamic travel demand models for hurricane evacuation. *Transportation Research Record,* 1964, 211–218

Harsanyi, J.C. (1975). Can maximin principle serve as a basis for morality – critique of rawls. J Theory. *American Political Science Review,* 69(2), 594–606

Horner, M.W. & Downs, J.A. (2007). Testing a flexible GIS-based network flow model for routing hurricane disaster relief goods. *Transportation Research Record,* 2022, 47–54

Horner, M.W. & Downs, J.A. (2008a). An analysis of the effects of socio-economic status on hurricane disaster relief plans. *Transportation Research Record,* 2067, 1–10

Horner, M.W. & Downs, J.A. (2008b). Optimizing Hurricane Disaster Relief Goods Distribution and Transportation: Florida Department of Emergency Management Comprehensive Plan. Paper read at Annual Meeting of the Transportation Research Board, at Washington, DC.

Horner, M. W. & Mefford, J.N. (2005). Examining the spatial and social variation in employment accessibility: a case study of bus transit in austin, texas. In D. Levinson & K. Krizek (Eds.), *Access to Destinations* (pp. 193–214). (Oxford: Elsevier)

Horner, M. W. & Murray, A.T. (2002). Excess commuting and the modifiable areal unit problem. *Urban Studies,* 39(1):131–139

Lindell, M.K. & Prater, C.S. (2007). Critical behavioral assumptions in evacuation time estimate analysis for private vehicles: Examples from hurricane research and planning. *Journal of Urban Planning and Development-Asce,* 133(1), 18–29

Liu, S.X. & X. Zhu, A. (2004). Accessibility analyst: an integrated GIS tool for accessibility analysis in urban transportation planning. *Environment and Planning B-Planning and Design,* 31(1), 105–124

Marsh, M.T. & Schilling, D.A. (1994). equity measurement in facility location analysis – a review and framework. *European Journal of Operational Research,* 74(1), 1–17

Miller, H. J. & Shaw, S.-L. (2001). *Geographic information systems for transportation*. (New York: Oxford University Press)

Mirchandani, P. B. & Francis, R.L. (1990). *Discrete location theory*. (New York: John Wiley and Sons)

Morrill, R. L. & Symons, J. (1977). Efficiency and equity aspects of optimum location. *Geographical Analysis,* 9(3), 215–225

Nicholas, M. A., Handy, S. L & Sperling, D. (2004). Using geographic information systems to evaluate siting and networks of hydrogen stations. *Transportation Research Record, 1880,* 126–134

Oppong, J.R. (1996). Accommodating the rainy season in third world location-allocation applications. *Socio-Economic Planning Sciences,* 30(2), 121–137

Preston, J. & Rajé, F. (2007). Accessibility, mobility and transport-related social exclusion. *Journal of Transport Geography,* 15(3), 151

ReVelle, C., Marks, D. & Liebman, J.C. (1970). Analysis of private and public sector location models. *Management Science Series a-Theory,* 16(11), 692–707
ReVelle, C.S. & Eiselt, H.A. (2005). Location analysis: a synthesis and survey. *European Journal of Operational Research,* 165(1), 1
ReVelle, C.S. & Swain, R.W. (1970). Central facilities location. *Geographical Analysis,* 2(1), 30–42
Schwanen, T. (2003). Transport planning, logistics, and spatial mismatch: a regional science perspective. *Journal of Regional Science,* 43(1), 206–209
Scott, D.M. & Horner, M.W. (2008). Examining the role of urban form in shaping people's accessibility to opportunities: an exploratory analysis. *Forthcoming in the Journal of Transportation and Land Use*.
Smoyer-Tomic, K.E., Spence, J.C. & Amrhein, C. (2006). Food deserts in the prairies? Supermarket accessibility and neighborhood need in Edmonton, Canada. *Professional Geographer,* 58(3), 307–326
Talen, E. & Anselin, L. (1998). Assessing spatial equity: an evaluation of measures of accessibility to public playgrounds. *Environment and Planning A,* 30(4), 595–613
Teitz, M. B. & Bart, P. (1968). Heuristic methods for estimating generalized vertex median of a weighted graph. *Operations Research* 16(5), 955–961
Truelove, M. (1993). Measurement of spatial equity. *Environment and Planning C-Government And Policy,* 11(1), 19–34
Wolshon, B. (2006). Evacuation planning and engineering for hurricane katrina. *The Bridge*, 36(1), 27–34
Wright, P.D., Liberatore, M.J. & Nydick, R.L. (2006). A survey of operations research models and applications in homeland security. *Interfaces,* 36(6), 514–529

# Chapter 21
# Visualizing and Diagnosing Coefficients from Geographically Weighted Regression Models

**David C. Wheeler**

**Abstract** Visualizing and interpreting regression coefficients from spatially varying coefficient models, such as geographically weighted regression (GWR), can be challenging, given the amount of information the models provide the spatial analyst. Adding to the visualization dilemma are various diagnostic tools for checking the coherence of the model. One such diagnostic tool is based on decomposing the regression coefficient variance matrix to test for collinearity, which is relevant because previous research has shown that collinearity in GWR can lead to estimated regression coefficients for multiple regression terms that are strongly correlated with each other. In this paper, visualization tools, such as linked scatter plots, parallel coordinate plots, and maps, are presented for diagnosing correlation in estimated regression coefficients that can be problematic for statistical inference of relationships between variables. These tools help explain patterns of dependence between regression terms apparent in scatter plots of estimated coefficients and link the structure in the scatter plots to locational information in a map. Visualization of this information should aid in the typical spatially varying coefficient model estimation process. A case study of census undercount in Franklin County, Ohio is presented as an illustrative example of applying the visual diagnostic approach in a GWR analysis.

**Keywords** GWR · Regression diagnostics · Collinearity · Visualization · ESDA · Census undercount

## 21.1 Introduction

The declaration from John Chambers, "There is no single statistical tool that is as powerful as a well-chosen graph." (Chambers et al. 1983, p. 1), is a fundamental starting point for this paper. The simple, but compelling, idea expressed

D.C. Wheeler (✉)
Harvard School of Public Health, Harvard University, Boston, MA, USA
e-mail: wheelerdc@mail.nih.gov

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6_21,

by Chambers is that by visualizing information, one can see what was not previously apparent and gain some insight into what is being studied. A graph can convey information about a pattern that a general statistical measure alone cannot. This idea is especially relevant for analyzing data with spatially varying coefficient models. One such type of model that has become popular in geography is geographically weighted regression (GWR), which repeatedly applies a spatial kernel function to spatial data to produce regression coefficient estimates that may vary over space. Visualizing and interpreting regression coefficients from spatially varying coefficient models, such as GWR, can be challenging given the amount of information the models provide the spatial analyst. Adding to the visualization dilemma are various diagnostic tools for checking the coherence of the GWR model. One such diagnostic tool, the variance-decomposition proportion and associated condition index (Belsley 1991; Wheeler 2007), is based on decomposing the estimated regression coefficient variance matrix from GWR to test for collinearity. Testing for collinearity is important and especially relevant for inference because previous research (Wheeler and Tiefelsdorf 2005; Wheeler 2007; Wheeler and Calder 2007; Waller et al. 2007; Griffith 2008) has shown that it can lead to estimated regression coefficients for different regression terms that are strongly correlated with each other in GWR, and this dependence among coefficients can interfere with the interpretation of coefficients for one term marginally. Related to the problem of strong dependence among estimated regression coefficients in GWR is the problem of inflated variance and over-smoothing of estimated coefficients, as previous work (Wheeler 2007) has suggested that strong correlation in the overlapping kernel weighted explanatory variables that GWR uses increases the estimated variance of regression coefficients. This in turn increases the magnitude of estimated coefficients and induces over-smoothing in the coefficients.

The approach described in this paper builds on the powerful ideas of visualizing information and diagnosing regression model output by using well-chosen graphs of model diagnostics for geographically weighted regression, linked to estimated GWR coefficients, to find areas that suggest problems with the GWR coefficients. The visualization tools used for diagnosing correlation in estimated regression coefficients include linked scatter plots, parallel coordinate plots, and maps. The primary goal of the visualization is to highlight areas where coefficients can be problematic for statistical inference on relationships between variables. The selected visualization tools help explain patterns of dependence evident in scatter plots of estimated coefficients for different regression terms and link structure in the scatter plots to locational information on a map. Synergistic visualization of this information should aid users in the typical spatially varying coefficient model fitting process. To demonstrate the application of these tools, a case study of analyzing census undercount in Franklin County, Ohio using GWR is presented.

## 21.2 Methods

### *21.2.1 Geographically Weighted Regression*

The technical details underlying GWR have been described elsewhere (Fotheringham et al. 2002), but the basics are presented here for completeness. In GWR, a regression model can be fitted at each observation location in the dataset, although the model calibration locations are not restricted to observation locations. The spatial coordinates of the data points are used to calculate inter-point distances, which are input into a kernel function to calculate weights that represent spatial dependence between observations. For each location, $s = 1, \ldots, n$, where the model is calibrated, the GWR model is

$$y(\mathbf{s}) = \mathbf{X}(\mathbf{s})\beta(\mathbf{s}) + \varepsilon(\mathbf{s}), \tag{21.1}$$

where $y(s)$ is the dependent variable at location $s$, $\boldsymbol{\beta}(s)$ is a column vector of regression coefficients at location $s$, $\mathbf{X}(s)$ is a row vector including the explanatory variables at location $s$, and $\varepsilon(s)$ is the random error at location $s$. The vector of $p$ estimated regression coefficients (including the intercept) at location $s$ is

$$\hat{\beta}(s) = [\mathbf{X}^T \cdot \mathbf{W}(s) \cdot \mathbf{X}]^{-1}\mathbf{X}^T \cdot \mathbf{W}(s) \cdot \mathbf{y}, \tag{21.2}$$

where y is the $n \times 1$ vector of dependent variables; $\mathbf{X} = [\mathbf{X}(1);\ \mathbf{X}(2);\ \ldots;\ \mathbf{X}(n)]^T$ is the design matrix of explanatory variables, including a column of 1's for the intercept; and $\mathbf{W}(s) = diag[w_1(s), \ldots, w_n(s)]$ is the diagonal weights matrix calculated for each calibration location $s$.

The weights matrix, W($s$), is calculated from a kernel function that places more weight on observations that are closer to the calibration location $s$ than those that are more distant. In this paper, the exponential function is used as the kernel function. The weights from the exponential kernel function are calculated as

$$w_j(s) = \exp(-d_{sj}/\gamma), \tag{21.3}$$

where $d_{sj}$ is the distance between the calibration location $s$ and another location $j$, and $\gamma$ is the kernel bandwidth parameter.

To fit the GWR model, one must first estimate the kernel bandwidth using cross-validation, which is an iterative process that finds the kernel bandwidth that minimizes the prediction error of all the $y(s)$. Next, one must calculate the weights at each calibration location using the kernel function. Finally, one estimates the regression coefficients at each calibration location along with the response variable estimates by the expression $\hat{y}(s) = \mathrm{X}(s)\hat{\beta}(s)$. In typical applications of GWR to a spatial dataset, analysts have mapped the estimated regression coefficients and then attempted to interpret the spatial pattern in the coefficients in the context of the problem under study and the regression model variable relationships.

## *21.2.2 Diagnostic Tools*

Due to the attention in the literature (Wheeler and Tiefelsdorf 2005; Wheeler and Calder 2007; Waller et al. 2007; Griffith 2008; Wheeler 2009) on potentially strong correlation in sets of estimated GWR coefficients, there is a need to diagnose model collinearity and correlation of estimated coefficients for pairs of different regression terms in GWR. A commonly used diagnostic tool for collinearity in ordinary least squares (OLS) regression is the variance inflation factor (Neter et al. 1996), which measures how much estimated variances of regression coefficients are increased by collinearity in the model. A drawback of the variance inflation factor (VIF) as a collinearity diagnostic is that it is a single summary measure for each coefficient and does not illuminate the nature of the collinearity. For example, if there are ten regression coefficients in an OLS model and there is collinearity in several regression terms, the VIF for any coefficient can indicate the presence of collinearity for that term but cannot indicate the terms with which it is collinear (see Belsley 1991 for more details). In addition, the global VIF diagnostic does not easily transfer to the local model setting of GWR. To assess collinearity in the GWR model requires a local-based diagnostic tool that provides information regarding both the strength of and nature of model collinearity. One diagnostic tool that provides this information is the combination of variance-decomposition proportions and the associated condition index introduced by Belsley (1991) and modified for GWR by Wheeler (2007). The variance decomposition of Belsley (1991) is done through singular value decomposition (SVD) of the design matrix, where SVD factors a matrix into distinct pieces which can then be used to conveniently calculate quantities of the matrix, such as the variance-covariance (see below). The variance-decomposition proportion is the proportion of the variance of a regression coefficient that is affiliated with one component of its decomposition. It has an associated condition index, which is the ratio of the largest singular value to the smallest singular value.

The diagnostic tool as modified by Wheeler (2007) uses singular value decomposition of the kernel weighted design matrix to form condition indexes of this matrix and variance-decomposition proportions of the coefficient covariance matrix. The singular value decomposition of the design matrix in the GWR framework is

$$\mathbf{W}^{1/2}(i)\mathbf{X} = \mathbf{UDV}^T, \tag{21.4}$$

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal $n \times (p+1)$ and $(p+1) \times (p+1)$ matrices respectively, $\mathbf{D}$ is a $(p+1) \times (p+1)$ diagonal matrix of singular values of decreasing value down the diagonal starting at position (1,1), $\mathbf{X}$ is the column scaled matrix (by its norm) of explanatory variables including the constant, and $\mathbf{W}^{1/2}(i)$ is the square root of the diagonal weight matrix for calibration location $i$ calculated from the kernel function. Using the SVD, the variance-covariance matrix of the regression coefficients is

$$Var(\hat{\beta}) = \sigma^2 \mathbf{V}\mathbf{D}^{-2}\mathbf{V}^T, \tag{21.5}$$

and the variance of the $k^{th}$ regression coefficient is

$$Var(\hat{\beta}_k) = \sigma^2 \sum_{j=1}^{p} \frac{v_{kj}^2}{d_j^2}, \tag{21.6}$$

where the $d_j$'s are the singular values and the $v_{kj}$'s are elements of the $\mathbf{V}$ matrix. The variance-decomposition proportion is the proportion of the variance of the $k^{th}$ regression coefficient affiliated with the $j^{th}$ component of its decomposition. Following from Belsley (1991), the variance-decomposition proportions are

$$\pi_{jk} = \frac{\varphi_{kj}}{\varphi_k}, \tag{21.7}$$

where

$$\varphi_{kj} = \frac{v_{kj}^2}{d_j^2} \tag{21.8}$$

and

$$\varphi_k = \sum_{j=1}^{p} \varphi_{kj}. \tag{21.9}$$

The condition index for column $j = 1,\ldots., p + 1$ of $\mathbf{W}^{1/2}(i)\mathbf{X}$ is

$$\eta_j = \frac{d_{\max}}{d_j}, \tag{29.10}$$

where $d_j$ is the $j^{th}$ singular value of $\mathbf{W}^{1/2}(i)\mathbf{X}$.

In diagnosing collinearity, the larger the condition index, the stronger is the collinearity among the columns of the GWR weighted design matrix. In a general ordinary least squares regression setting, Belsley (1991) recommends a conservative value of 30 for a condition index that indicates collinearity, but suggests the threshold value could be as low as 10 when there are large variance proportions for the same component. The presence of two or more variance-decomposition proportions greater than 0.5 in one component of the variance decomposition indicates that collinearity exists between at least two regression terms, one of which may be the intercept. These guidelines are applied in diagnosing collinearity in a GWR model used later in this paper. An advantage of the variance-decomposition proportion and condition index diagnostic tool is that it indicates collinearity locally, at the same spatial scale of dependence and same locations where the GWR models are calibrated. This enables one to construct maps or graphs of the local diagnostic values and link them explicitly to the GWR coefficients. An approach for visualizing the diagnostic tools in such a useful way is described in the following sections.

### *21.2.3 Visualization Tools*

"The scatter plot may well be the single most powerful statistical tool for analyzing the relationship between two variables, $x$ and $y$." (Chambers et al. 1983, p. 75)

The above quote by Chambers and coauthors about the power of the scatter plot for investigating the relationship between two variables is certainly true in general, and scatter plots have been used to assess problems in estimated GWR coefficients, such as strong dependence in coefficients for pairs or regression terms, counterintuitive signs, and inflated variance (Wheeler and Tiefelsdorf 2005; Wheeler 2007; Waller et al. 2007; Griffith 2008). However, when assessing and diagnosing GWR models, there is a need to move beyond simple scatter plots and correlation coefficients of GWR coefficients from pairs of regression terms. The problem is that with spatially varying coefficient models there are multivariate spatial data and one must evaluate the relationships between several variables at a time and look at the spatial nature of the relationships, as well as investigate whether the relationships are being represented properly or are being distorted by statistical artifacts. Scatter plots alone are not sufficient in an analysis of the relationships in multivariate spatial data as represented by spatially varying regression coefficients.

Instead, the proposal put forth here is to use a combination of linked graphics and maps to visualize GWR coefficients and diagnostics together to assess model output and gain insight into relationships between variables over space. The approach is novel in the context of GWR analysis in that it combines exploratory spatial data analysis tools with statistical diagnostic tools specifically for analyzing GWR model results. The role of linked plots in statistical analysis is described by Cleveland and McGill (1988) and Tierney (1990) and these plots are often called dynamic graphs. Linked plots allow one to use the principle of data brushing, which is the action of interactively selecting data observations in one graph and having these observations highlighted simultaneously in all linked graphs, to better visualize relationships in elements of different types of information. This technique is also present in geographic information system (GIS) software, manifested as subset selections of data. GIS software packages such as ArcGIS (ESRI 2005) and GeoDa (Anselin 2003) have linked graphs for subsets of observations on maps and graphs, such as scatter plots. Selection sets can be created by selecting observations through attribute query, selecting within a graphical window, or brushing across a plot.

A graph, in addition to the scatter plot, that is useful for visualizing relationships between variables is the parallel coordinate plot, which in contrast to the scatter plot, is designed for multivariate data. Parallel coordinate plots (see Inselberg 1985; Wegman 1990; Moustafa and Wegman 2006) visualize the values of multiple variables simultaneously for all observations in a dataset by joining the values for each variable for an observation by a line, where each variable is an axis, either as a row in a horizontal parallel coordinate plot or as a column in a vertical plot (plots are effective in either orientation). By visually following a line through the layout of the parallel coordinate plot, one can see how the values for different variables are related for an observation. Parallel coordinate plots, therefore, allow one to distinguish where values tend to be high for one variable and low for another variable across a dataset. Another graph that is valuable in assessing model output is the histogram, which is designed for showing the distribution of one variable alone.

Together, these graphics, scatter plots, parallel coordinate plots, histograms, and maps, along with the techniques of linking graphs and brushing observations, can be combined effectively to visualize and diagnose problems with GWR model output. Moving forward, it is recommended that spatial data analysts using GWR employ, at a minimum, scatter plots of estimated GWR coefficients for pairs of regression terms linked to parallel coordinate plots containing variance-decomposition proportions and condition indexes for observations and also to maps of estimated coefficients. This allows the analyst to select observations with problematic diagnostic values and locate them on a map, in addition to selecting observations with coefficients that are correlated for different terms in a scatter plot and viewing their diagnostic values and location on coefficient maps. One may also link parallel coordinate plots of estimated GWR coefficients and parallel coordinate plots of diagnostic values to see trends in coefficients for observations that have problems with collinearity and shared variance, as indicated by the diagnostics. Also, one could produce scatter plots of pairs of variance-decomposition proportions and select observations with large proportions for both regression terms to investigate any spatial pattern or clustering in estimated coefficient maps. In addition, histograms of condition indexes are beneficial as a graphical indication to the extent of collinearity in the GWR model. The goal of the visualization of diagnostics and coefficients described here is congruous with the philosophy of data analysis expressed by Tukey (1977) who states, “The greatest value of a picture is when it forces us to notice what we never expected to see”. Examples of the linked plots described above will be presented in the case study that follows.

## 21.3 Census Undercount in Franklin County, Ohio

The following example exploring census undercount in Franklin County, Ohio illustrates how the visual and diagnostic tools described above function effectively together. The objective in this example is to build a model to explain percent census undercount in Franklin County, which contains the city of Columbus, OH. The US Census Bureau attempts to take a complete enumeration of the population every 10 years, but invariably the census does not include every person. To account and adjust for this, the bureau takes a post-enumeration survey some time after the census and attempts to match people from the census and the post-enumeration survey to determine how many people were not counted in the census. This estimate is the census undercount. To model the log percent census undercount, we will use the following explanatory variables: MNRT, the percent minority population; RENT, the percent renter population; PVRT, the percent of population under the poverty line; and MALE, the percent male population. Nationally, undercount is higher for minorities, such as Hispanics and blacks, compared with whites, and is also higher for males (Steffey 1997; Darga 1998). The other variables may indicate areas where it is difficult to reach people in a census, for example, in higher crime areas or areas with more homeless people. The log percent census undercount is mapped by block group in Franklin County in Fig. 21.1. Figure 21.2

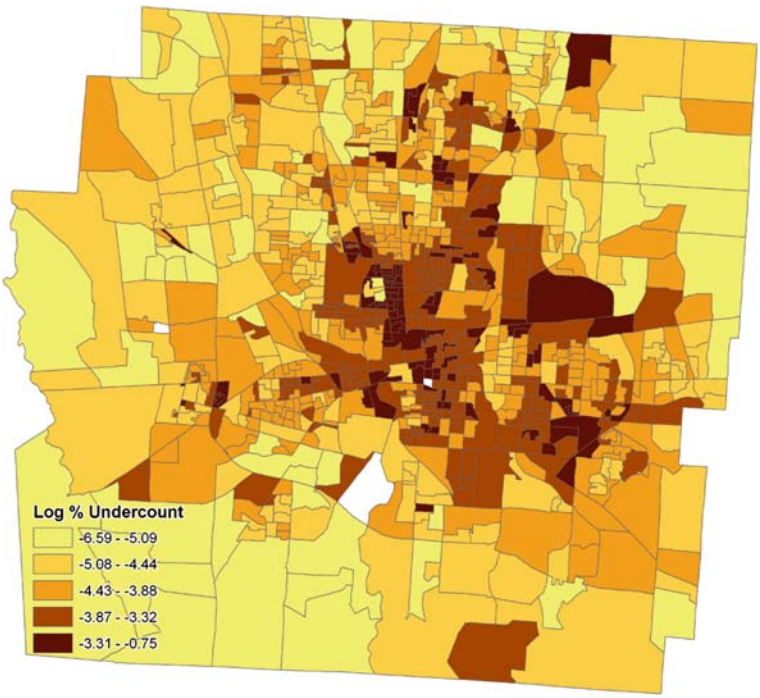



**Fig. 21.1** Log percent of census undercount by block group for Franklin County, Ohio

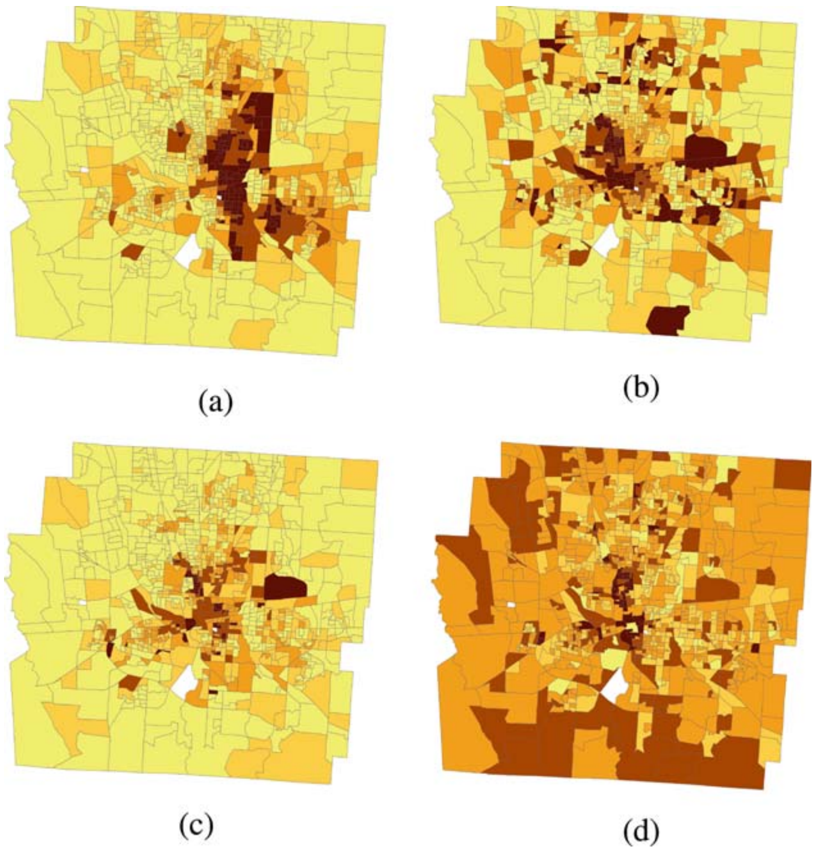



**Fig. 21.2** Explanatory variables used to model variation in census undercount: percent minority (**a**), percent renter (**b**), percent poverty (**c**), and percent male (**d**)

shows the spatial pattern of the four explanatory variables by block group. These maps of the dependent and explanatory variables are made using natural breaks classification. There appears to be some spatial pattern in the log percent of undercount by block group. It also seems that some of the spatial variation in undercount can be explained by several of the covariates, such as percent minority and percent renter.

To get a base for the analysis, it is prudent to start with an ordinary least squares, or global, linear regression model with the four explanatory variables of interest and an intercept:

$$\begin{aligned} y(s) = \beta_1 + \beta_2 \cdot \mathrm{MNRT}(s) + \beta_3 \cdot \mathrm{RENT}(s) + \beta_4 \cdot \mathrm{PVRT}(s) \\ + \beta_5 \cdot \mathrm{MALE}(s) + \varepsilon(s). \end{aligned} \tag{21.11}$$

The coefficient of determination for this model is $R^2 = 0.57$, hence, the OLS model explains a substantial amount of the variation in percent census undercount. The estimated OLS regression coefficients are $\hat{\beta}_1 = -6.123$, $\hat{\beta}_2 = 0.874$, $\hat{\beta}_3 = 0.776$, $\hat{\beta}_4 = 1.100$, $\hat{\beta}_5 = 2.702$ and all are significant at the 0.001 level. The positive coefficients agree with relationships in national trends and seem logical and intuitive. Therefore, there are no obvious problems with this four-variable model in the signs of the regression coefficients.

There is a suggestion of collinearity in the OLS model, as the regression coefficient correlation coefficient is –0.99 for the intercept and the percent male coefficient. In practice, analysts sometimes ignore strong correlation with the intercept and other regression coefficients, arguing that some correlation, perhaps strong, is expected given the nature of the relationship between the intercept and a slope in fitting a line to a set of data points. Belsley (1991), however, argues that collinearity with the intercept is important and should not be ignored, and it can impact the standard errors of the regression coefficients. The single summary measure of the collinearity and the merely global, as opposed to the local, estimated regression coefficients in the OLS model make it fairly easy to overlook the collinearity. When moving to the GWR model, however, it becomes impossible to ignore collinearity and its effects when looking at the estimated local regression coefficients and local variance-decomposition proportions and condition indexes. While generally ignoring collinearity in OLS regression models is not recommended, focus now turns to the GWR model to make the point that collinearity effects in OLS are amplified or exacerbated in GWR for collinearity that includes the intercept. It has been demonstrated elsewhere that collinearity among predictor variables (not the intercept) can produce substantially more harmful effects in the GWR model compared to the OLS model (Wheeler and Tiefelsdorf 2005; Wheeler 2007).

The comparable GWR model to the OLS model is

$$y(s) = \beta_1(s) + \beta_2(s) \cdot \text{MNRT}(s) + \beta_3(s) \cdot \text{RENT}(s) + \beta_4(s) \cdot \text{PVRT}(s) + \beta_5(s) \cdot \text{MALE}(s) + \varepsilon(s), \tag{21.12}$$

which has the same regression terms as the OLS model, but now the intercept and covariate effects are allowed to vary over space. To fit the GWR model, the GWR kernel bandwidth is first estimated as $\hat{\gamma} = 1.58$ through cross-validation and then the regression coefficients at each block group center are estimated using R software (Venables and Smith 2002). Since analytical interest is often on inference of potentially varying relationships over space, the GWR coefficient estimates are mapped at every data location. The estimated GWR coefficients for the intercept and the percent male term are mapped in Fig. 21.3. There is a clear complimentary pattern in the mapped coefficients. In general, the intercept is low where the percent male coefficient is high, and vice versa. The strong negative correlation in the estimated GWR coefficients across block groups for these two terms is evident in the scatter plot of the coefficients in Fig. 21.4.

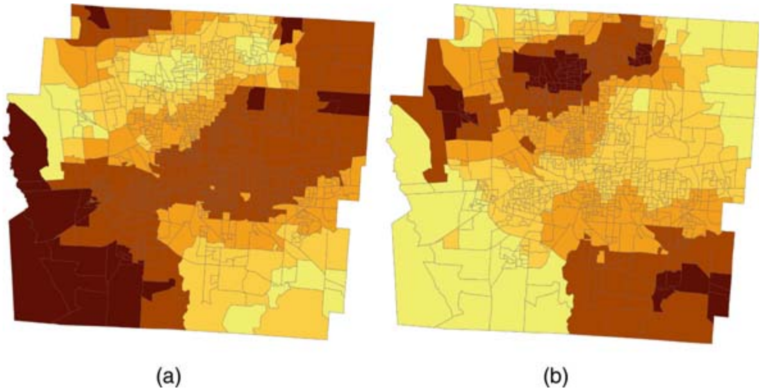



**Fig. 21.3** The estimated GWR coefficients for the intercept (**a**) and percent male (**b**)

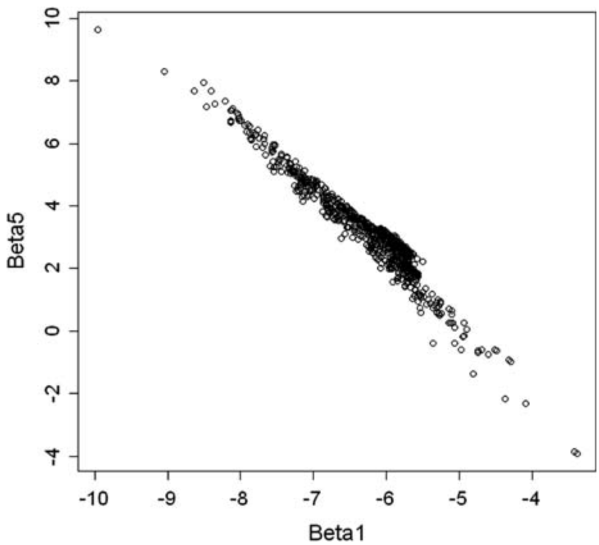



**Fig. 21.4** Scatter plot of the estimated GWR coefficients for percent male versus the intercept

The problem with such strong dependence between terms is in attempting to interpret the spatial pattern in the coefficients for individual terms, i.e. marginal inference. The spatial pattern in the percent male coefficients is a near direct, and negative, result of the spatial pattern in the intercept. Any inference on the relationship between percent males and percent census undercount will be biased by the strong association with the intercept. While a strong regression coefficient correlation for the intercept and another term may go overlooked in the OLS regression model, the strong negative correlation in local coefficients is a more obvious and observable problem in the GWR model. Another clear indication of a collinearity problem in the GWR model is that the estimated coefficients for percent male are all negative, in contrast to the result from the

OLS model. In fact, the largest estimated GWR coefficient for percent male is around -3, where the OLS estimate is approximately 2.72. There is no justifiable substantive reason for all the estimated percent male coefficients to become negative in the GWR model. At this point in the analysis, it appears ill-advised to keep the intercept and percent male in the GWR model, as interpreting the percent male coefficient marginally will not be useful, and moreover, will be misleading.

To be certain, however, it is beneficial to apply and evaluate the diagnostic tools described earlier to the GWR model. Using the same kernel bandwidth estimated previously in the GWR model, the condition indexes and variance-decomposition proportions are calculated from the kernel weighted design matrix using R software. The distribution of the condition indexes associated with the largest variance component is plotted in Fig. 21.5. Most condition indexes are greater than the conservative threshold of 30, which indicates a problem with collinearity in the model. A useful way to display the condition indexes and the variance-decomposition proportions together is with a parallel coordinate plot. This allows one to identify where both the condition index and the variance-decomposition proportions indicate a problem in the model simultaneously. Figure 21.6 is a parallel coordinate plot of the condition indexes associated with the largest variance component and the variance-decomposition proportions for the two worrisome terms, the intercept and percent male. The variance-decomposition proportions for both terms are over 0.95 for all block groups and this indicates a major problem with shared variance for the intercept and percent male. Effectively, the variance in the two sets of regression

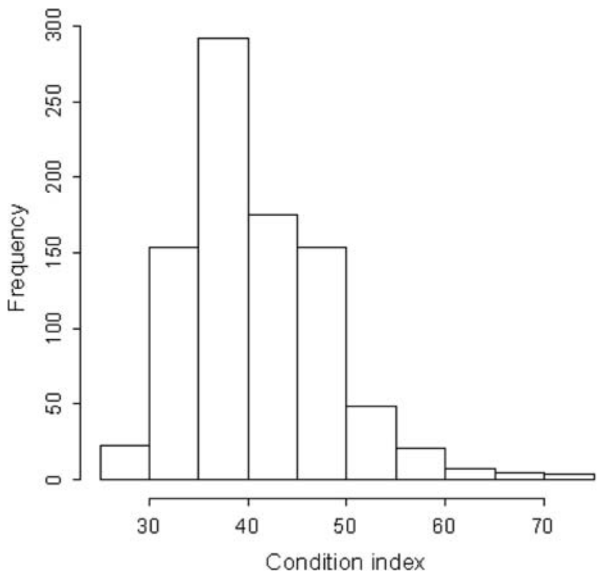



**Fig. 21.5** Histogram of the condition indexes for GWR model 1

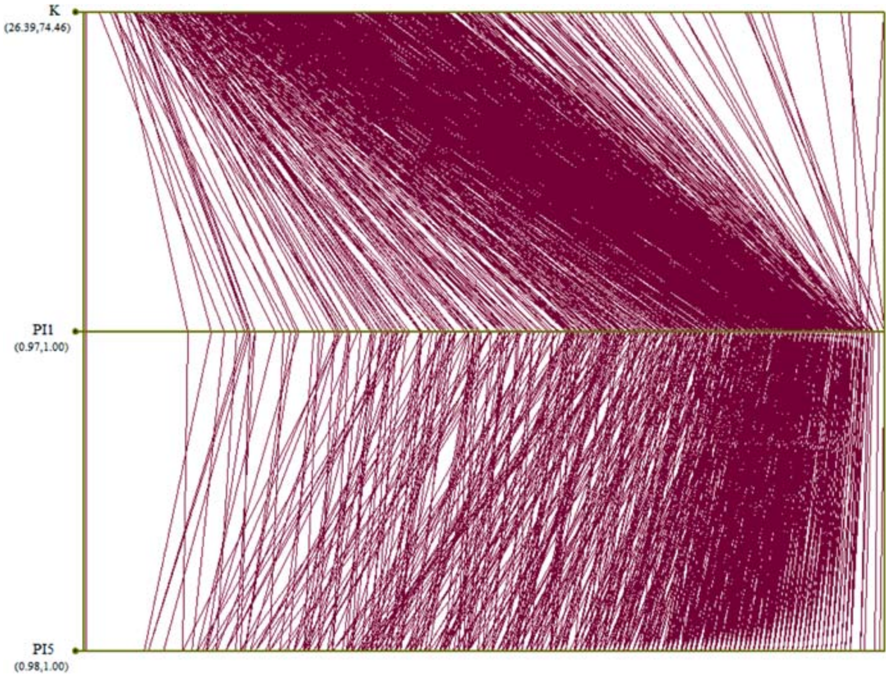



**Fig. 21.6** Parallel coordinate plot of the condition index (K) and variance-decomposition proportions for the intercept (PI1) and percent male (PI5) for the GWR model

coefficients is explained by one variance component. There is essentially one piece of information from two different sources. Hence, attempting to separate the two sources for inference is not possible.

In addition to the intercept and percent male, there are three other terms for which to investigate diagnostic values, and an efficient way to visualize the relationships in the coefficients for the five terms together is through parallel coordinate plots. Figure 21.7 is a parallel coordinate plot for the estimated GWR coefficients for all five terms. The lighter lines are a selection set of approximately half of the block groups with the lowest intercept coefficients. The darker lines are the remaining observations. It is clear in the parallel coordinate plot that the lowest values of the intercept correspond with the highest coefficient values for percent male. Therefore, this plot conveys similar information as the scatter plot in Fig. 21.4; however, it is multivariate in nature in that it provides additional information for the relationships between other regression coefficients. For example, it is apparent from the plot that several block groups have small values for percent poverty and small values for the intercept and large values for percent male. While visualizing the relationships among coefficients is helpful, linking a parallel coordinate plot of the diagnostic values to the parallel coordinate plot of the regression coefficients offers more insight into the nature of the relationships. Figure 21.8 is a parallel coordinate

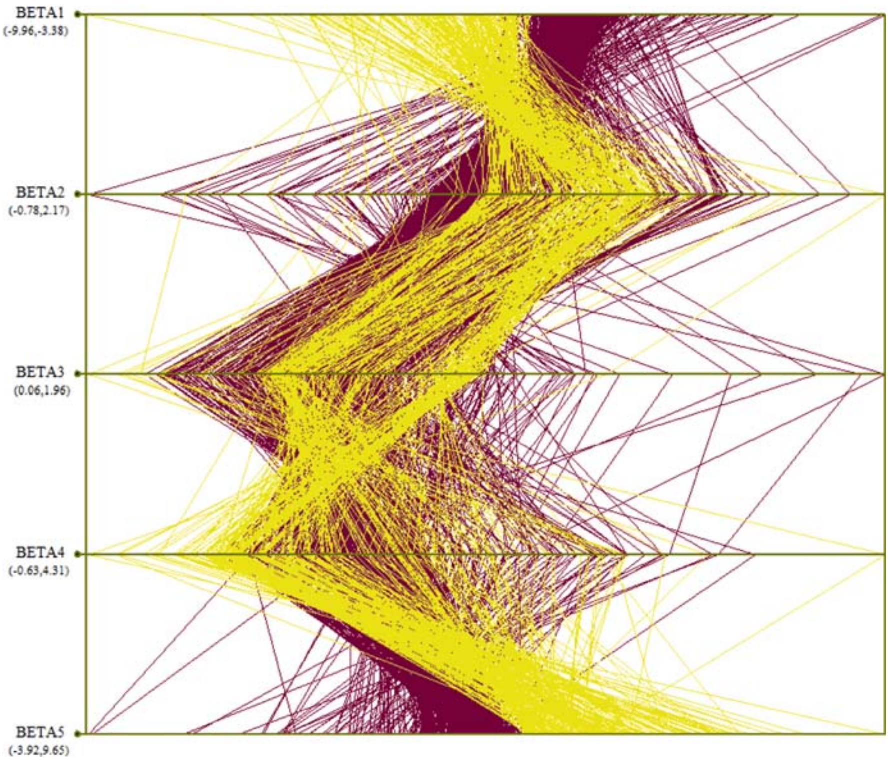



**Fig. 21.7** Parallel coordinate plot of the estimated GWR coefficients for the intercept (Beta1), percent minority (Beta2), percent renter (Beta3), percent poverty (Beta4), and percent male (Beta5)

plot of the condition index and variance-decomposition proportions for all five regression terms for the largest variance component. In this plot, it is clear that this variance component is dominated by the first and fifth regression terms, the intercept and percent male.

It is also informative to link the parallel coordinate plots of estimated GWR coefficients and diagnostic values with maps of the GWR coefficients and investigate where selected observations for outstanding coefficients or diagnostic values are located in the study area. Specifically, linking the observations with troublesome diagnostic values to estimated coefficient maps allows the user to identify spatially where the coefficients are not reliable. Figure 21.9 shows the GWR coefficients for the intercept and percent male with the selected block groups from the parallel coordinate plots in Figs. 21.7 and 21.8 shaded in a cross-hatch pattern on the maps. There are two distinct groupings of block groups (northwest and southeast) with low values of the intercept and high values of the percent male effect. These are examples of where the GWR coefficients should not be interpreted due to the large diagnostic values. The

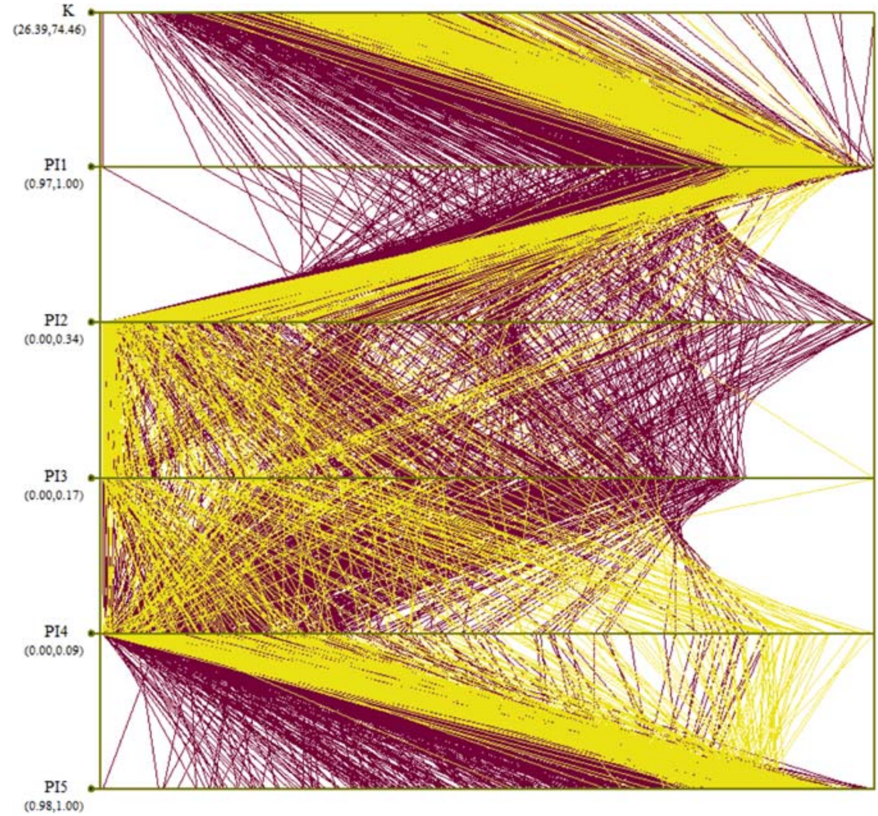



**Fig. 21.8** Parallel coordinate plot of the condition index (K) and variance-decomposition proportions for the intercept (PI1), percent minority (PI2), percent renter (PI3), percent poverty (PI4), and percent male (PI5) for the GWR model

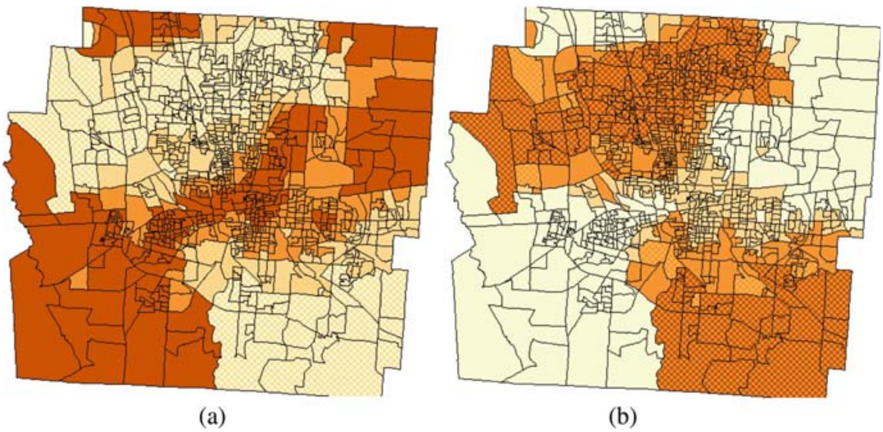



**Fig. 21.9** Estimated GWR coefficients for the intercept (**a**) and percent male (**b**) with selected observations from the parallel coordinate plot in Fig. 21.8

ability to select observations based on patterns of diagnostic values in the parallel coordinate plot and immediately visualize them in the coefficient map is a powerful tool in assessing the trustworthiness of model output.

Overall, the scatter plot of coefficients for the intercept and percent male term, the histogram of the condition indexes, and the parallel coordinate plot of the variance-decomposition proportions and the condition indexes indicate the presence of extensive and debilitating collinearity in the current GWR model. The next logical step, if remaining in the GWR framework, is to remove the percent male term and evaluate whether that alleviates the problem. This new GWR model without percent male is

$$y(s) = \beta_1(s) + \beta_2(s) \cdot \mathrm{MNRT}(s) + \beta_3(s) \cdot \mathrm{RENT}(s) + \beta_4(s) \cdot \mathrm{PVRT}(s) + \varepsilon(s). \quad (21.13)$$

For clarity, we refer to this model as GWR model 2 and the first model as GWR model 1. The estimated kernel bandwidth for this model is $\hat{\gamma} = 1.49$. The coefficient of determination for the OLS version of this model is $R^2 = 0.57$. The condition indexes are plotted in the histogram in Fig. 21.10. The largest condition index is just over 8; hence the condition indexes alone do not indicate a problem with collinearity with this model. It appears that removing the percent male term was adequate to alleviate the problem in the model.

For illustrative purposes, it is useful to visualize the variance-decomposition proportions along with the estimated regression coefficients to explore any patterns that may exist. Figure 21.11 has a scatter plot of the variance-decomposition

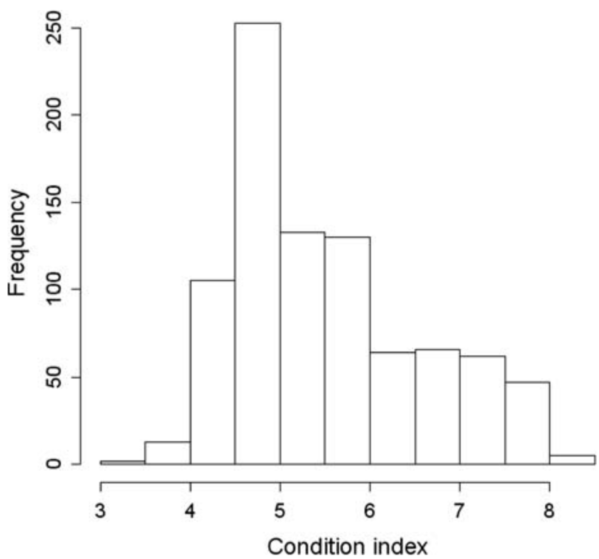



**Fig. 21.10** Histogram of the condition indexes for GWR model 2

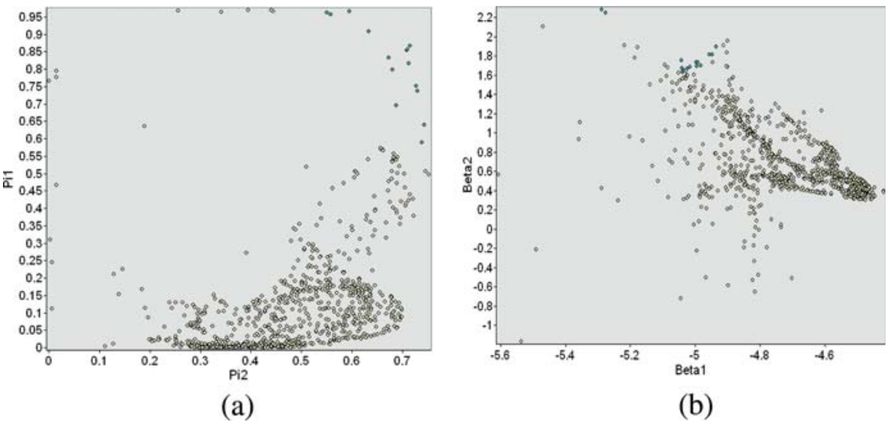



**Fig. 21.11** Variance-decomposition proportions for the intercept (Pi1) versus percent minority (Pi2) for the largest variance component (**a**) and the estimated GWR coefficients for percent minority (Beta2) versus the intercept (Beta1) for GWR model 2 (**b**)

proportions for the intercept and percent minority GWR coefficients for the highest variance component from GWR model 2 and a scatter plot of the estimated coefficients for these terms. There is a selection set in the scatter plots (aqua-colored points) for observations where the variance-decomposition proportions for both the intercept and percent minority are greater than approximately 0.55. These are points in the upper right of the variance-decomposition proportions scatter plot. As shown in the coefficients scatter plot, these are observations that tend to have higher values for the percent minority effect and lower values for the intercept. The estimated regression coefficients for these two terms are mapped in Fig. 21.12. The

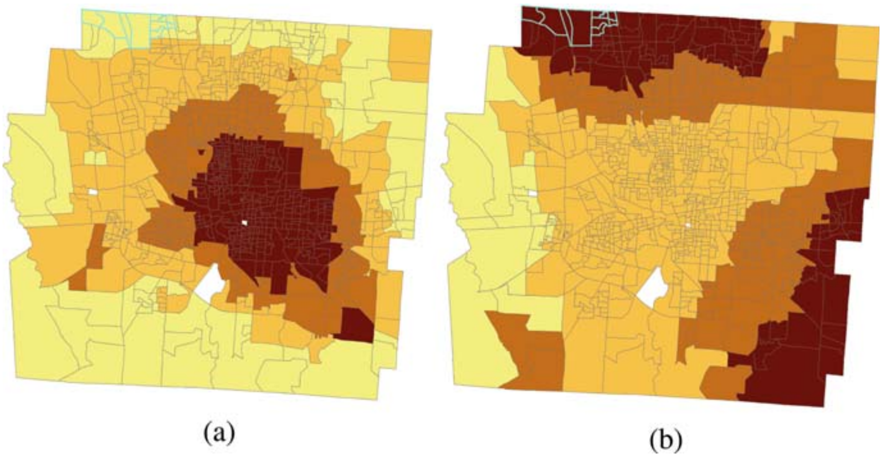



**Fig. 21.12** Estimated GWR coefficients for the intercept (**a**) and percent minority (**b**) for GWR model 2

selected observations are outlined in a light shade and are visible in (b). These observations with the highest variance-decomposition proportions for the intercept and percent minority are co-located in the northwest corner of the county. This cluster of observations has a substantial shared variance in the intercept and percent minority term, and the regression coefficients should be interpreted with caution if the condition indexes were higher for the block groups in the cluster. Reinforcing the theme of the power of linking graphs of diagnostics and coefficients, we see that by linking the observations in the scatter plot to the maps of coefficients, the underlying spatial structure that was not evident in the two-dimensional graphs becomes apparent.

Often the concern with collinearity in GWR models is between two or more covariates and not with the intercept. Therefore, one might be interested in exploring the relationship between two covariate effects in GWR model 2. To do so, one can start with Fig. 21.13, which is a scatter plot of the estimated

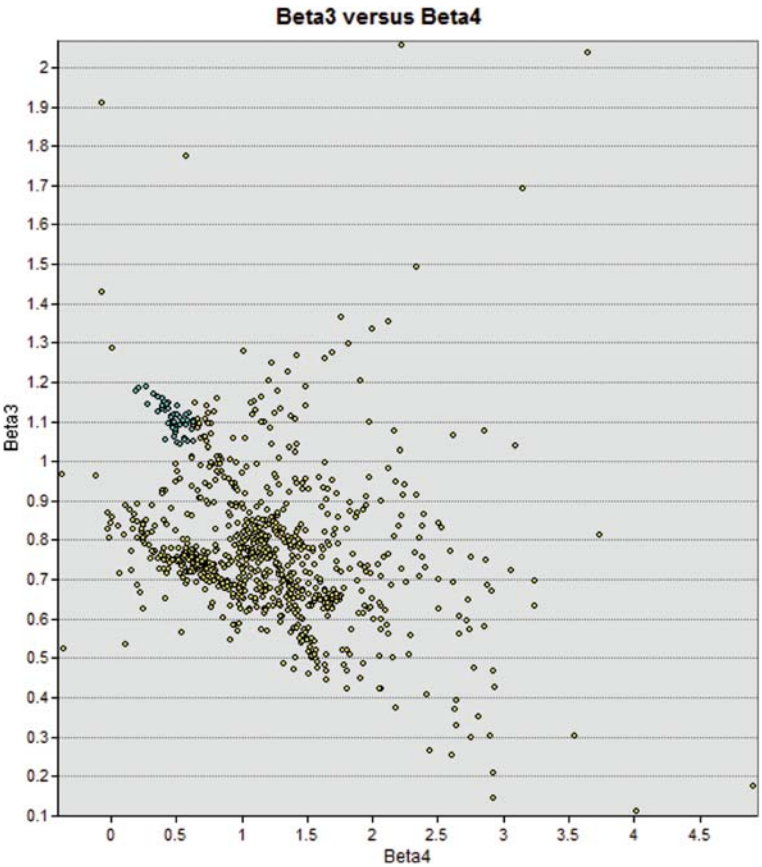



**Fig. 21.13** Estimated GWR coefficients for percent renter (Beta3) versus percent poverty (Beta4) with a selected set of block groups in aqua

GWR coefficients for percent renter versus percent poverty. In this plot, there are lines or ribbons of coefficients that indicate a negative relationship between these two effects. There is a selection set in the scatter plot for one of these linear groupings of coefficients with negative correlation, where coefficients for percent renter tend to be high and coefficients for percent poverty tend to be low. In linking these selected block groups to the coefficient maps for percent poverty and percent minority in Fig. 21.14, spatial structure in the coefficients is apparent in the form of a cluster of block groups in a north-central part of the county with generally high values for percent renter coefficients and low values for percent poverty. From the scatter plot, it is clear there is negative correlation in some coefficients for percent renter and percent poverty, and linking the coefficient maps to the plot makes it clear where these coefficients are in the study area.

The graphs presented so far have been produced using GeoDa, ArcGIS, and R software. The parallel coordinate plots and some maps were made in GeoDa. The histograms and some scatter plots were made in R. Some scatter plots and maps were produced in ArcGIS. It is straightforward to create selection sets across several linked graphs in both GeoDa and ArcGIS. An alternative to these software packages for visualizing GWR coefficients and diagnostics is the Java-based software ESTAT (http://www.geovista.psu.edu/ESTAT/), which stands for Exploratory Spatio-Temporal Analysis Toolkit. ESTAT was created at the GeoVISTA Center at Penn State University to visualize multivariate spatial and temporal data. By default, ESTAT has a scatter plot, parallel coordinate plot, bivariate map, and time series plot that are all dynamically linked. As an example of using ESTAT for visualizing GWR coefficients and diagnostics together, Fig. 21.15 shows an ESTAT exploration of the estimated

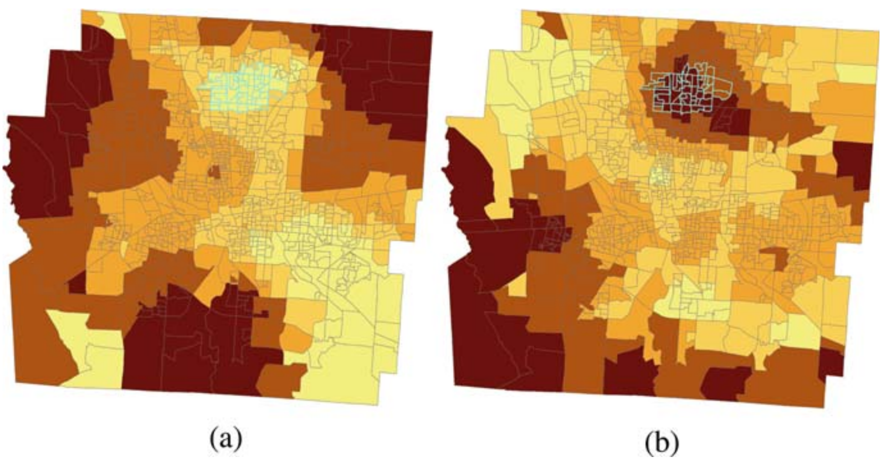



**Fig. 21.14** Estimated GWR coefficients for percent poverty (**a**) and percent renter (**b**) with a selection set of block groups outlined

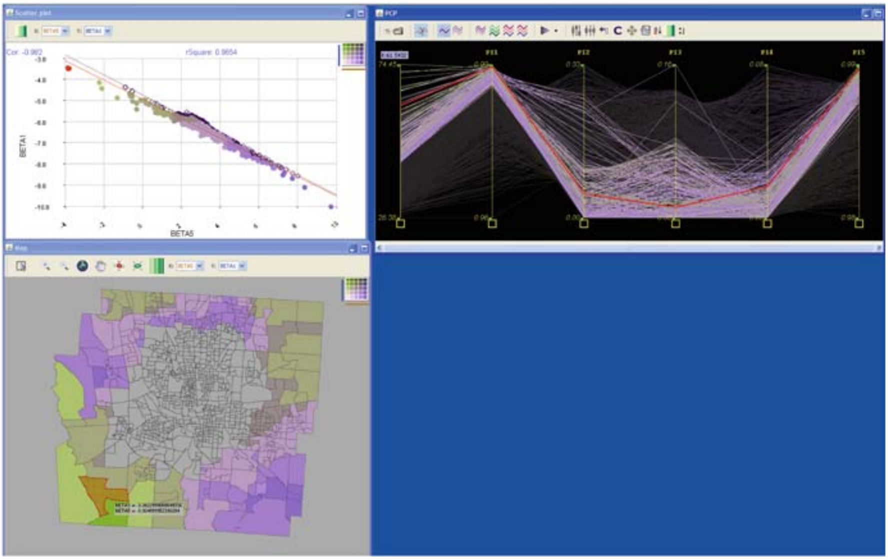



**Fig. 21.15** Estimated coefficients and diagnostic values for GWR model 1 using ESTAT

coefficients for the intercept and percent males from GWR model 1 and the condition indexes and variance-decomposition proportions for the largest variance component. The time series plot is omitted due to the static temporal nature of the census undercount data. In the linked graphs, there is a selection of observations shaded lighter in the scatter plot and parallel coordinate plot where the variance-decomposition proportions for the intercept and percent males are both high. In the bivariate map, the selected block groups are shaded according to their coefficients for both covariates and the unselected observations are shaded gray. This example is useful in showing that ESTAT provides a user-friendly visualization system with integrated graphics as the default.

## 21.4 Conclusion

This paper has described an approach using a combination of visualization tools and statistical diagnostics to highlight estimation issues in geographically weighted regression coefficients. An example demonstrating the utility of the approach was presented for exploring census undercount in Franklin County, Ohio. The estimation issues to draw attention to include inflated variances and strong dependence in estimated coefficients for different regression terms, at least partly due to model collinearity. These are problems for the inference on relationships between variables represented in a regression model, which is often of primary interest in the application of GWR. Visualization tools offer utility in evaluating and assessing model output because it can be difficult to digest the amount of information generated by a GWR model, as well as the additional information provided by model diagnostic tools. More specifically, visualizing and linking estimated regression coefficients and model diagnostic values provide the data analyst with information about the reliability of local model parameter estimates as well as the location of troublesome estimates. Exploring the diagnostics and coefficients together in dynamically linked graphics, which several software options offer, enables the user to easily and immediately find and select areas where more investigation is needed prior to interpreting model results. The work presented here focuses on geographically weighted regression; however, one can apply similar visualization tools to coefficients from other spatially varying coefficient regression models, such as Bayesian regression models. The diagnostic tools are different in the Bayesian setting, but linked graphical tools such as scatter plots, parallel coordinate plots, and maps are applicable. Future work will pursue visualizing diagnostics for Bayesian models using local measures, such as the partial deviance information criterion (Spiegelhalter et al. 2002).

## References

Anselin, L. (2003). *GeoDa 0.9 user's guide*. (Urbana-Champaign, IL: Spatial Analysis Laboratory (SAL), Department of Agricultural and Consumer Economics, University of Illinois

Belsley, D.A. (1991). *Conditioning diagnostics: collinearity and weak data in regression*. (New York: John Wiley & Sons)

Chambers, J.M., Cleveland, W.S., Kleiner, B. & Tukey, P.A. (1983). *Graphical methods for data analysis*. (Pacific Grove, CA: Wadsworth & Brooks/Cole)

Cleveland, W.S. & McGill, M.E. (1988). *Dynamic graphics for statistics*. (Monterey, CA: Wadsworth)

Darga, K. (1998). Straining out gnats and swallowing camels: the perils of adjusting for census undercount. Office of the State Demographer, Michigan Information Center, Michigan Department of Management and Budget

ESRI (2005). *ArcGIS 9.1 Users Guide*

Fotheringham, A.S., Brunsdon, C. & Charlton, M. (2002). *Geographically weighted regression: the analysis of spatially varying relationships*. (West Sussex: John Wiley & Sons)

Griffith, D. (2008). Spatial filtering-based contributions to a critique of geographically weighted regression (GWR). *Environmental and Planning A*, 40, 2751–2769

Inselberg, A. (1985). The plane with parallel coordinates. *The Visual Computer,* 1, 69–91

Moustafa, R. & Wegman, E. (2006). Multivariate continuous data – parallel coordinates. In A. Unwin, M. Theus & H. Hofmann (Eds.), *Graphics of large datasets: visualizing a million* (pp. 143–155). (Berlin: Springer)

Neter, J., Kutner, M.H., Nachtsheim, C.J. & Wasserman, W. (1996). *Applied linear regression models*. (Chicago: Irwin)

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B,* 64, 583–640

Steffey, D.L. (1997). A review of the census undercount issue. Faculty Fellows Program, Center for California Studies, California State University

Tierney, L. (1990). *LISP-STAT: an object oriented environment for statistical computing and dynamic graphics*. (Chichester: Wiley)

Tukey, J.W. (1977). *Explanatory data analysis*. (Reading, MA: Addison-Wesley)

Venables, W.N. & Smith, D.M. (2002). *An introduction to R*. (Bristol: Network Theory Limited)

Waller, L., Zhu, L., Gotway, C., Gorman, D. & Gruenewald, P. (2007). Quantifying geographic variations in associations between alcohol distribution and violence: A comparison of geographically weighted regression and spatially varying coefficient models. *Stochastic Environmental Research and Risk Assessment,* 21(5), 573–588

Wegman, E. (1990). Hyperdimensional data analysis using parallel coordinates. *Journal of the American statistical association,* 85, 664–675

Wheeler, D.C. (2009). Simultaneous coefficient penalization and model selection in geographically weighted regression: The geographically weighted lasso. *Environment and Planning A*, 41, 722–742

Wheeler, D.C. (2007). Diagnostic tools and a remedial method for collinearity in geographically weighted regression. *Environment and Planning A,* 39(10), 2464–2481

Wheeler, D.C. & Calder, C. (2007). An assessment of coefficient accuracy in linear regression models with spatially varying coefficients. *Journal of Geographical Systems,* 9(2), 145–166

Wheeler, D.C. & Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systemsm,* 7, 161–187

# Epilog

**Michael F. Goodchild**

## Why Does Location Matter?

The 21 chapters of this book all share one defining theme – they are all in one way or another about location. In looking back over the collection it is tempting therefore to ask why? Why is location considered so important that it defines the contents of a book of almost 500 pages? I had a somewhat related reaction over 40 years ago when I first encountered SYMAP, the first software package for computer mapping (by today's standards incredibly crude computer mapping) developed by a group at Harvard University: "Why would a scientist want to put a map in a computer?" It is a reaction I still encounter frequently, particularly among social scientists who are not geographers: "What possible role can location play in explaining anything?"

Over the years many researchers, including all of the editors and chapter authors of this book, have come to realize the importance of location, and thus of geospatial analysis and modeling. Some years ago some colleagues and I wrote about six arguments for location (Goodchild et al. 2000) in our rationale for establishing the Center for Spatially Integrated Social Science, a center based on the notion that location could serve as an integrating theme and basis for conversation and collaboration across the social sciences:

- *Integration*: the value of location as a common key linking otherwise unconnected items of information, allowing them to be compared and correlated. This is the basis for the familiar "layer-cake" view of GIS that dominates many textbook covers.
- *Cross-sectional analysis*: the ability to draw inferences from the patterns displayed by phenomena on the Earth's surface. While it is difficult to confirm

M.F. Goodchild (✉)
Department of Geography, Center for Spatial Studies, University of California, Santa Barbara, CA, USA
e-mail: good@geog.ucsb.edu

B. Jiang, X. Yao (eds.), *Geospatial Analysis and Modelling of Urban Structure and Dynamics*, GeoJournal Library 99, DOI 10.1007/978-90-481-8572-6,

the presence of a specific process without the benefit of longitudinal data, it is often possible to use cross-sectional data to eliminate other process options.

- *Spatially explicit modeling*: the appearance of location-related variables (distance, adjacency, connectivity, area, *etc*.) in models of processes.
- *Place-based analysis*: forms of analysis that provide local rather than global summaries of pattern, perhaps the most radical of the six in relation to traditional scientific method.
- *Linking science to policy*: location is essential if general scientific knowledge is to be applied to the development or examination of policy options in specific locations.
- *Location as an index to information*: the use of location (and perhaps time) as a basis for searching, as increasingly employed in on-line search services.

All of these, with the possible exception of the last, are exemplified, sometimes in combination, by the research described in the pages of this book.

It is important, I think, not to underestimate the degree to which many of these challenge the orthodox view of scientific method, which is after all grounded in the traditional disciplines of physics, chemistry, and biology. Place-based analysis in particular suggests that the *nomothetic* view – that discoveries are of greatest value when they apply everywhere at all times – may be overly restrictive in studies of the complex geographic world, since phenomena may be far from homogeneous across the Earth's surface. Techniques such as Geographically Weighted Regression (see for example Chapters 19 and Chapters 21) abandon the notion that the only results worthy of the label scientific must be true globally, preferring a more nuanced view in which the parameters of global models are allowed to vary locally. This is not quite the *idiographic* view of geographic description – the view that everywhere is unique – that was so denigrated by the scientific geographers of the 1960s, but it is not quite the nomothetic view either.

Geospatial analysis and modeling also require a more nuanced view of statistical methods, which were inherited largely from disciplines such as psychology and agronomy in which controlled experiments are the norm. Cross-sectional data are rarely collected by investigators through rigorous experimental design, but instead are acquired from agencies such as the US Bureau of the Census. Sampling, if it takes place at all, is rarely random and independent, and instead spatial dependence is the almost universal norm. Yet our educational system invariably trains students to deal with the conditions of controlled experiments, not the *natural* experiments that dominate geospatial analysis.

## The World is Flat (and Rural)

Despite the potential benefits, to put a map in a computer proved to be an enormously tedious, time-consuming, and error-prone affair. Early users of SYMAP were issued with special rulers, marked off in 1/10 inch intervals on one side and 1/6 inch on the other, to be used to measure the coordinates of points so that these could later be punched on cards and read into the computer. Even well into the 1990s it was still customary to train students in the use of large map *digitizers* that could be employed to capture point locations using a hand-manipulated mouse. Today the task is more likely to be carried out on-screen, using a cursor to point at a scanned image. Nevertheless the paper map remains an important source of the data that find their way into GIS and into geospatial analysis and modeling.

Maps capture a view of the world at a particular scale or representative fraction, and the agencies that have traditionally produced maps have emphasized complete coverage of their respective jurisdictions. For example there are more than 55,000 different map sheets in the US Geological Survey's National Topographic Series covering the 48 conterminous states at 1:24,000. Because maps are expensive to produce, they have tended to emphasize information that is relatively static. Urban dynamics, which are complex, detailed, and transitory, have never fared well in the mapping business. Moreover maps are two-dimensional, and have paid little if any attention to the complex three-dimensional worlds of urban structures. Finally, mapping has always been strongly linked to warfare, and indeed in many countries mapping still managed entirely as a military enterprise – and wars in the past have been fought largely in open, rural spaces. Recently military doctrine has been shifting, driven by the belief that the future of warfare is urban, and that it is the *human terrain* of urban dynamics that will be important, not the natural terrain of rural stasis.

This rural, two-dimensional bias is exacerbated by GPS, which is only useful outdoors and away from tall buildings and dense tree cover. GPS-based navigation is a technology for the car, failing when the car moves into a structure or into an urban canyon, and useless to a pedestrian trying to navigate the complex three-dimensional geography of urban environments. Most of us spend almost all of our lives in such environments, and find today's GPS-based navigation to be somewhat like tobacco – to make use of it one must go outside.

## Looking Forward

If the geospatial approach to urban dynamics and modeling is important, as the chapters of this book have argued and as I have tried to summarize, and if geospatial data for the study of urban dynamics are currently in woefully short supply, then what developments can we expect in the future?

First, there is every reason to believe that the data problem will ease, through the development of new technologies and as a result of the abundance of important applications. A variety of forms of tracking are becoming available, based on cellphones, RFID (radio-frequency identification), GPS, and video surveillance, and raising crucially important issues of privacy and confidentiality. Several technologies are under development to solve the indoor positioning problem, and in time one of them will prove the most useful and successful and will come to dominate the field. Individual citizens are now empowered to create geospatial data for themselves and to make it available to others, and the phenomenon known as volunteered geographic information, community mapping, or user-generated geospatial content will become increasingly important as an alternative to traditional sources (Goodchild 2007). In future, we should envision a world in which it is possible *to know where everything is at all times*, and begin the important work of considering the societal costs and benefits of such a world, and the applications that will be enabled.

Second, there is likely to be a significant improvement in the supply of data on the three-dimensional structure of urban areas. Photogrammetric software for the extraction of external building form from airborne, ground-based, and satellite-based systems, particularly LiDAR, is now advanced to the degree that cost-effective digital representations can be produced for major cities. These technologies are still problematic for non-Western architectural forms, and do not solve the problem of capturing the internal structure of buildings. In the latter case the CityGML project (www.citygml.org) has developed methods for integrating the two-dimensional representations of GIS with the three-dimensional representations of building information management (BIM). These developments will be important and valuable for future work in geospatial urban dynamics.

Perhaps more important than these developments, however, is what I perceive as the growing conflict between research practice in fields such as urban dynamics and the traditional methodological principles of science, and the need to reconcile this conflict. Scientific philosophy is still dominated by the image of the rugged individual investigator – the Darwin or Einstein or Rutherford – and increasingly out of touch with the reality of contemporary research. Only 7 of the 21 chapters of this book are by single authors, collaboration between researchers with different skill sets being increasingly important to the solution of today's scientific problems. The term *black box* is still pejorative, and yet few if any of us can claim to understand fully all of the ways in which our data are manipulated by research software. Science is still dominated by discovery, and yet more and more of our time is spent in tying science to policy, and in developing the tools that make science accessible to its end users. All of our data are uncertain and perfect explanation is virtually unachievable in fields such as urban dynamics, and yet our scientific philosophy gives us little guidance to working in this world, preferring an earlier world in which perfect prediction was still possible, within the bounds of experimental error. How valuable it would be to have access to a book that described the realities and pitfalls of working in a field such as the geospatial analysis and modeling of urban structure and dynamics – something akin to David Harvey's 1969 *Explanation in Geography* (Harvey 1969) but updated to contemporary practice.

## References

Goodchild, M.F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal,* 69(4), 211–221

Goodchild, M.F., Anselin, L., Appelbaum, R.P. & Harthorn, B.H. (2000). Toward spatially integrated social science. *International Regional Science Review,* 23(2), 139–159

Harvey, D. (1969) *Explanation in Geography*. (New York: St Martin's Press)

# Index