Spring 2012

Spatial Analysis of Travel Behavior and Response to Traveler Information

Xin Wang
Old Dominion University

SPATIAL ANALYSIS OF

TRAVEL BEHAVIOR AND RESPONSE TO TRAVELER INFORMATION

by

Xin Wang
B.S. July 2000, Southwest Jiaotong University, China
M.S. April 2003, Southeast University, China

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

CIVIL AND ENVIRONMENTAL ENGINEERING

OLD DOMINION UNIVERSITY
May 2012

Approved by:

Asad J. Khattak (Director)

Mecit Cetin (Member)

R. Michael Robinson (Member)

ABSTRACT

SPATIAL ANALYSIS OF
TRAVEL BEHAVIOR AND RESPONSE TO TRAVELER INFORMATION

Xin Wang
Old Dominion University, 2012
Director: Dr. Asad J. Khattak

Transportation planners have long recognized that it is urgent to integrate emerging spatial analysis with travel behavior studies. A clearer understanding of the spatial interactions among travelers and the complex environment they face has the potential to reap benefits of the ongoing technologies of travel behavior, spatial analysis and Advanced Traveler Information Systems (ATIS).

Considering that spatial patterns have been overlooked in the literature of travel behavior and ATIS, the main objective of this research is to use robust methods of spatial analysis to enhance the understanding of how the associations between traveler decisions, built environment and socio-demographic characteristics are organized spatially. This dissertation takes a significant step towards filling this gap by using innovative spatial data description methods, e.g. geo-imputation, dynamic buffer analysis, spatial statistics to model the travel behavior of both the general population and university students.

This study starts by developing a unique database from extensive behavioral data combined with a variety of spatial measurements, taking advantage of increased GIS capabilities. Five different activity-based databases from different regions are used, combined with their related socio-demographic and land use data. Among them are two general population travel surveys from North Carolina, which were conducted in

Charlotte and at the Greater Triangle in 2003 and 2006, respectively. The Virginia Add-on for the general population was conducted in 2008, while two waves of the Virginia University Student Travel Survey (USTS) were conducted in 2009 and 2010. The general population and the university students are compared with each other in terms of how they traveled and responded to ATIS.

Issues addressed in this dissertation include two aspects. The first one is how to describe data in space more accurately. When there is a need to know the exact locations of residences (geo-coordinate), but such information is unknown, geo-imputation is used as a fundamental method of assigning synthetic locations randomly to these residences based on available zonal information. After locating the residences by using geo-imputation, dynamic buffer analysis is used to capture locally built environment characteristics around residences, which place emphasis on capturing accessibility.

The second issue is modeling travel behavior in space. Particular emphasis is placed on modeling associations between trip making, trip decision changes and their associated explanatory variables. The general population is compared with the university students who represent an energetic and technology-savvy subgroup of the population. Different spatial scales are used for these two groups: the regional level is used for the general population; the university campus is used as a special trip generator for the university students.

At the regional level, a unique model structure, i.e. Geographically Weighted Regression (GWR), is used to allow associations to change across space, referred to as spatial heterogeneity. Significant spatial heterogeneity is found in the associations between trip-making and built environment, as well as in the model of travelers' information acquisition behavior and their travel decision adjustments. The spatial heterogeneity in the trip-making models suggests that there is higher spatial variability in favor of the statement that better land use design can help reduce auto trips. It is important to note that these potentially useful insights would have remained uncovered if using a non-spatial model that does not take spatial heterogeneity into account.

At the special trip generator level, when local models don't work well, the university campus is studied as a case which represents a combination of livable environments and a group of people who have different life cycles compared with the general population. Particular spatial analysis is applied to capture the association between trip-making and students' residential proximity to campus. The models confirm there are rings of mobility around the campus. Different from the traditional travel demand model for the general population, this varied level of mobility of students based on their residential proximity of campus is important and must be considered in the students' travel demand model.

This thesis is dedicated to my family.

ACKNOWLEDGMENTS

I would like to express my deepest thanks to my advisor, Dr. Asad J. Khattak. This expression is not enough to him regarding his guidance, mentorship, encouragement and support. I am always amazed at his creativity, sharpness and power of determined will to tackle research problems. Appreciation also goes to Dr. R. Michael Robinson, Dr. Mecit Cetin, Dr. Guoqing Zhou for their time and productive comments which are incredibly detailed and informative. I would also like to express gratitude to Ms. Kathy Moulton for giving me her warm advice on how to present and communicate in English.

I want to thank my peers in the transportation program, Hongbing, Sanghoon, Semuel, for providing me the chance to learn from each other, and for their help and encouragement. Also, my appreciation also goes to my old classmates, Yingling, Fang Lei, Wang Qian, Wen Ya, Chen Qian by showing me that friendship in college benefits your whole life. I would also like to express my sincere thankfulness to my friends in ODU, Libin, Yanhong, Yanxiao, for their friendship and support.

I would like to express my warmest gratitude to my family, especially, Tang Kun, my husband, for his love, patience, courage and support during my doctoral study in the past five years. This support is priceless. I couldn't have done this without it.

I would also like to acknowledge the Transportation Research Institute at Old Dominion University that provided the opportunity and necessary financial support during my graduate studies. I would like to thank to all those who I may have forgotten.

TABLE OF CONTENTS

LIST OF TABLES

Page

Table

LIST OF FIGURES

Page

Figure

1. INTRODUCTION

1.1 Motivation

The goal of travel, as a derived demand, is to satisfy peoples' need for activity participation in the context of their socio-demographic factors and space and time. The impacts of socio-demographic factors have been largely captured in studies. However, a stumbling block in traditional travel behavior models is that the geographical nature of trips is rarely considered, and there is a lack of systematic framework to explain the spatial phenomena that emerges from complex interactions among travelers. As urban areas are the space where the majority of people live, work and participate in activities, intangible urban spatial constraints may be imposed upon transportation systems, imposing spatial patterns on how people travel and participate in activities. Therefore, exploring the geographical nature of travel behavior is important.

Interest in analyzing spatial data has grown considerably in the scientific research community, especially with emerging technologies of geographic information system science (GISc) and applied spatial statistics. Over the last decade, there have been a number of developments in the field of spatial data analysis, theoretical and practical, which are capable of providing indicative details of information hidden in space by providing solid supports for spatial data mining and spatial statistics. However, current GIS applications in the travel behavior field are still focused on data assembling and visualization using thematic maps, which do not explain travel behavior well. Meanwhile, transportation is moving from a data-poor to a data-rich environment (Miller and Shaw, 2001). Increasingly, available travel behavior surveys have provided detailed thematic and geographically referenced data for disaggregated activity participations and travel behaviors, coupled with the widespread availability of open source and proprietary tools (Buliung and Remmel, 2008) for spatial analysis. The availability of high quality travel survey data, the growth of geographical eco-demographic databases, and the emergence of new quantitative techniques have the capability to generate considerable research activity.

In this context, a research field integrating spatial analysis and travel behavior can enhance the understanding of their mutual interdependencies. By doing so, we won't limit our long-term capability to reap the benefits of the ongoing technologies of both travel behavior and spatial analysis.

Several issues related to combining spatial analysis and travel behavior study using behavior data with detailed geographic references are addressed and discussed in greater detail in the following sections.

1.2 Spatial Data Representation

How various components of travelers' behavior are properly represented is the fundamental question of spatial analysis and modeling. Most currently available travel behavior survey data and socio-demographic data are represented by using predefined geographical boundaries, e.g. TAZ (Traffic Analysis Zone), Census Tract, and Census Block. These boundaries have been used as standard units of delivering survey results and are consistently used by researchers and governmental agencies to estimate and predict travel demand and to inform travel behavior studies. As a simplified representation of urban space, the predefined boundaries have provided an easy way to contain and store socio-economic data by clustering information based on geographically boundaries.

However, problems exist. Studies have documented that spatial data aggregation based on a pre-defined boundary can cause the Modifiable Areal Unites Problem (MAUP) and affect the outcomes of transportation planning models significantly (Shawn Turner, 1997, Páez and Scott, 2005, Fotheringham and Wong, 1991). MAUP occurs when the spatial zoning system used to collect and/or report geographic data is "modifiable" or arbitrary. Variation of spatial effect can occur when data from one scale of areal units are aggregated into larger or smaller areal units, similar to "ecology fallacy." When the MAUP problem exists in spatial analysis, it introduces significant variability in the coefficient estimates as well as standard errors (Kwigizile and Teng, 2009). The MAUP is especially critical for trip generation since spatial aggregation into zoning systems can substantially affect the results from travel demand analysis (Miller, 1999). The trip generation results depend on how the TAZs are defined, and the result would be different if delimitation of TAZs are changed, even for the same region with the same socio-economical inputs if aggregated or partitioned in different ways (Kwigizile, 2007). Moreover, using predefined boundaries can cause the "edge effect"– ignoring the similarity or interdependencies that occur among locations within and outside of the boundaries (Miller, 1999, Griffith, 1983). The definition of the boundary of those sub-classifications will influence the estimation of the coefficients, referred to as the (undesirable) "boundary effect" (Miller, 1999).

To alleviate the above problems, there is a need to find a better way to represent spatial data, which can overcome the predefined boundary and use data which is free of predefined boundaries, e.g. disaggregate data. However, unfortunately, both the current travel behavior data and widely used socio-economic data, including NHTS (National Household Travel Survey), ATUS (American Time Use Survey), Census and CTPP (Census Transportation Planning Programs) data, do not release the data with detailed location information due to privacy concerns. NHTS provides detailed information for each trip made by household members in their travel days, but the specific locations of where they live, work and travel to are usually masked by being aggregated into a bigger geographic unit, e.g, census block group. Also, socio-economic information, such as densities of population and employment, is based on a predefined geographical unit. Hence, there is the need to extract useful information on the disaggregated level from available aggregated level data. By doing so, more local measurements (built environment) in the neighborhood of residences, working locations and travel destinations can be obtained.

1.3 Capturing Spatial Characteristics

The relationship between built environment and travel demand has attracted much attention in past years, especially recently with the development of the neo-traditional movement, the so-called New Urbanism, with the statement that proper neighborhood design can impact travel behavior and reduce automobile dependency and use. However, no consensus exists. Further, the conclusions of these studies usually cannot be compared directly due to different studies using different methodologies to capture built environment. An agreed-upon conceptualization of "built environment" is lacking (Handy, 2005). In most cases, proxy variables such as local accessibility with certain distance to residences or destinations are used to represent local environment characteristics (e.g. roadway characteristics such as length) and connected node ratio within buffer areas around residences. However, it may not capture the built environment accurately since the round buffers with fixed sizes may cover unusable land use in space, e.g. water bodies. In general, how to capture built environmental measurements more effectively and accurately is an important issue.

Besides the local neighborhood measurements, from a regional perspective, special generators are introduced in the travel demand modeling to represent certain types of facilities whose trip generation characteristics are not fully captured by the standard trip generation module or whose travel pattern cannot be easily captured by the standard travel survey. Regional travel demand analysis is somewhat rough in the treatment of activity participation in a metropolitan area with significant diversity. There are some spatial generators which combine special subgroups of population and special land use. They are concentrations of unusual activities in urban area, which merit special consideration in travel demand modeling. Such generators might include university campuses, military bases, big hospitals, large scale transportation hubs, special recreational sites such as sports stadiums, and large regional shopping centers. These special generators may be relatively few in number, but can impact regional traffic since they may produce or attract mass daily trips, and they represent a significant portion of trips and include special travel patterns with both temporal and spatial features. Their influence on the nearby roadway network system could not be adequately captured in regional model. Therefore, they justify special surveys as well as particular analysis. Doing so allows the travel demand model to better replicate the real scenario of the study area. However, although most MPOs have incorporated special generators in their travel demand models and many appear to be interested in developing more effective special generator procedures (Mamun et al., 2010), most of the travel demand models for special generators are tinkering with trip generation models, similar to traffic impact analysis. Also, the emphasis is placed on obtaining better trip generation rates, e.g. those suggested by NCHRP 365, instead of concentrating on understanding the specialty of travel behavior relative to those generators and the subgroup of the population bound by those generators.

1.4 Spatial Modeling

Issues have arisen with regard to describing travel behavior's spatial nature and how to model its spatial relationship effectively. Spatial analysis considerations are seldom recognized or accommodated in travel modeling (Bhat and Zhao, 2002), e.g. conventional statistical models fail to capture certain important properties of spatial data, such as clustering, dispersion, and systematic variability across space. All of these violate basic assumptions of independence and homogeneity implicit in conventional statistical analysis. Violation of these assumptions, in turn, leads to information loss, biased and/or inefficient parameters and the possibility of seriously flawed interpretation, conclusions and policy prescriptions (Griffith and Layne, 1999). Given these potential pitfalls, spatial effects, though often regarded as nuisances, can be perceived as opportunities to obtain deeper insights (Páez and Scott, 2005).

In reality, the utility of spatial units which are close to each other can be correlated due to commonly unobserved spatial elements. For instance, residents living in the same spatial cluster may share similar life attitudes or life styles due to personal preference, social or environmental reasons. Thus, people/households living in close proximity exhibit parallel behaviors, giving rise to similar observations for trip outcomes (e.g. duration, activity frequency, vehicle-mile travelled) (Buliung and Kanaroglou, 2007). This situation can cause spatial dependence, which is defined as situations where high variable values cluster near other similarly high values and low values cluster near similarly low values (Fotheringham et al., 2000).

At the same time, most models for trip frequency and activity participations assume the association between dependent variables, and explanatory factors are fixed across space. For instance, in the traditional trip generation model, the amount and type of trips in a TAZ are functionally associated with household characteristics such as income, vehicle ownership, family size and other socio-economic factors, e.g. density and type of development. These associations are assumed to be uniform for every spatial location. Both suggested trip rates based on cross classification and regression models are average trip rates which do not consider spatial variance, although urban or suburban areas are differentiated in some cases, e.g. NCHRP 365 suggested different trip rates for urban vs. suburban areas. However, this segmentation is still very rough, especially in the last few decades when American cities experienced vast physical sprawl and fragmentation (Chorus et al., 2007b). Therefore, ignoring spatial heterogeneity in travel behavior modeling may be problematic from both statistical and real condition perspectives.

1.5 Research Objectives

Considering that spatial patterns have long been overlooked in the travel behavior literature and amongst practitioners as well, this dissertation is meant to take a significant step towards filling this gap.

The purpose of this study is to detail important elements of spatial analysis and travel behavior methodologies and discuss their applications. When considering whether spatial pattern is needed to be included into activity participation and travel decision analysis, there are two basic questions as, Cliff and Ord (1981) stated regarding this aspect: 1) Is the spatial pattern displayed by the phenomenon significant in some sense and therefore worth interpreting? 2) Can we obtain any information on the process which has produced the observed pattern from an analysis of the mapped distribution of the phenomenon? Therefore, this study is mean to develop an integrated research capable of incorporating the spatial analysis into travel behavior models by considering these issues:

- How to take advantage of the current available behavior data with masked location information to conduct spatial analyses based on disaggregated level.
- How associations between activity participations and built environment are organized spatially and the ultimate consequences of such patterns.
- Other than targeting a capturing built environment in a very local geographical scale, how we can learn from analyzing a special trip generator- Is there a centripetal travel behavior style around special generator?
- How travelers respond to Advanced Traveler Information Systems or ATIS from a spatial perspective. Does their information acquisition behavior and travel decision adjustment change across space?

This dissertation, as integrated research of spatial analysis, travel behavior and travelers' response to ATIS, predominantly focuses on increasing our understanding of the complex spatial aspects of travel behavior and decision change based on ATIS. It covers a diverse array of topics and provides an in-depth analysis of a broad set of case-studies. The contents range from behavior data processing to spatial statistical modeling; from exploring the general impact of built environment on travel behavior to case study of the special generator in space; and from traveler's travel behavior to their responses to the diffusion of innovative and "high-impact" advanced traveler information systems. The unifying theme is to enhance the connections between spatial analysis and travel behavior modeling by applying a collection of spatial analysis techniques, e.g., geo-imputation, spatial heterogeneity, buffer shapes, etc., to unique behavioral datasets, which place heavy reliance on combining empirical travel behavior data with spatial contextual effects. The use of spatial analysis methods in this dissertation will reflect and emphasize the significance of spatial contextual effects in explaining travel decisions.

1.6 Summary of Data Sources

The study required extensive data collection on activity-travel data at micro level and related land use, as well as socio-demographic data with GIS support. Various travel behavior datasets from different areas are used in this dissertation. All of them have used trip diary to collect travel data, which satisfied the research. These data include:

- Regional travel survey of Charlotte, North Carolina conducted in 2003;
- Regional travel survey of the Research Triangle, North Carolina conducted in 2006;
- National Household Travel Survey (NHTS) Virginia Add-on survey, conducted in 2008;
- University Student Travel Survey (USTS), sponsored by Virginia Department of Transportation (VDOT), part of Virginia University Student Travel survey

as Virginia Add-on complement, which were conducted in 2009 and 2010. The target populations of the survey were university undergraduate and graduate students. Four universities in Virginia were involved in the first USTS, including two universities in urban areas – Old Dominion University (ODU) and Virginia Commonwealth University (VCU) and two universities in suburban areas – the University of Virginia (UVA) and Virginia Tech (VT). ODU and VT were also involved in the second USTS.

The first three travel surveys utilized standard household travel survey methods, in which all of the household members were asked to record their trips for a specified 24-hour period using a specially designed travel diary. The sampling plan included both geographic and demographic goals to ensure that the survey is representative of the region's population and activity-travel patterns. The travel survey database usually contains four different levels of data: personal data, household data, vehicle data and trip data. The surveys relied on the willingness of regional households to 1) provide demographic information about the household, its members and vehicles; 2) have all household members record all travel-related details for a specific 24-hour period, including trip purpose, mode, and travel time information for each trip. These detailed travel information are used to capture the daily travel behavior of travelers.

The USTS survey is the transformed-to-internet version of National Household Travel Survey (NHTS) and regional surveys, which was used partly because online surveys offer an efficient means of collecting data for college students. The USTS instrument was designed to resemble the NHTS. However, the first round survey conducted in 2009 suffered from the problem of underreporting and incompleteness due to response burden (average finishing time is over 40 min). Therefore, the second round of surveys was conducted in 2010 at ODU and VT. The second survey obtained a higher response rate and higher trips reported due to a substantial reduction of response burden by revising and refining the survey instructions (Khattak et al., 2011, Son et al., 2012).

The Triangle survey, Charlotte survey and USTS provide the exact location (latitude and longitude) information for all the respondents, their trip origins and destinations. This exact location information is critical for creating built environment variables and estimating spatial models. Given that the Triangle and Charlotte survey were conducted earlier, the NHTS Add-on provides the latest national household travel survey for Virginia residents. The USTS provides data for a unique environment, the university campus, which was traditionally underrepresented in regional travel demand models and not well understood. Therefore, USTS can serve as a case study for the special trip generator. Only the Triangle survey and first USTS provides ATIS usage and responding information regarding traveler information sources, information acquisition frequencies and travel choice changes based on the information received, which allows the analysis of how travelers respond to ATIS.

Therefore, the Charlotte and Triangle data are used to evaluate the accuracy of geo-imputation, a method used to create point based location given the accurate location information is known. Then the geo-imputation method is applied to assign exact location for residences in Virginia add-on data. Based on the assigned synthetic locations, the built environment characteristics can be calculated and used as regressors in activity models. University data is used to show the unique travel behavior of university students.

Finally, the Triangle data is used again to show how travelers respond to ATIS in a spatial context, and USTS first wave data is used again to be compared with the general population from the Triangle area.

1.7 Chapter Structure

It should be noted here that some chapters consists of papers that have been published, forthcoming, or submitted for publication in a scientific peer-reviewed journal. Some of the contents may be related, and some methodological overlaps between chapters exist. Therefore this dissertation is organized by topics. The structure of this dissertation is shown in Figure 1:

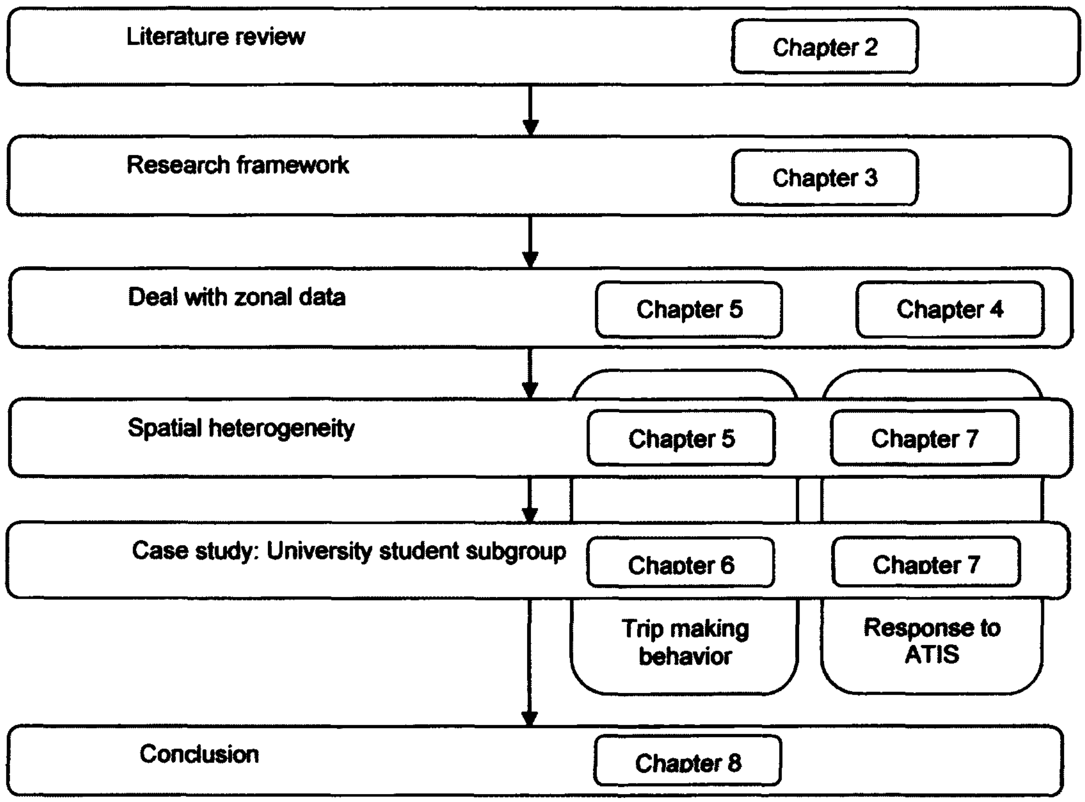


Figure 1. Chapter structure

Chapter 2 provides a synthetic review of the literature that is relevant for the study of the spatial pattern of activity participation, travel behavior and how traveler response to ATIS (Advanced Traffic Information Systems). Considering the dissertation covers a relatively wide range of topics, only reviews of relevant studies were kept. Chapter 3 presents the conceptual framework and design of this study. Chapter 4 elaborates on geo-imputation as a method to assign synthetic location randomly, which can present spatial data fairly by using available aggregated zonal data. The method presented in Chapter 4 is used in Chapter 5 as a fundamental method to deal with spatial data with aggregated geographical unit. Both Chapter 5 and Chapter 6 discuss trip making behavior; while Chapter 5 focuses on general population, Chapter 6 focuses on university students instead. In Chapter 6, the university campus is also studied as a case study of special generator. Chapter 7 analyzes the role of ATIS to support travel decision, and general population and university students are compared with each other. Spatial heterogeneity is addressed in both Chapter 5 and Chapter 7. While Chapter 5 emphasizes the spatial heterogeneity of the associations between built environment and trip making, Chapter 7 emphasizes several issues related to the role of traveler information to support travel decision, including the two-stage process of information delivery, the spatial heterogeneity of how traveler response to ATIS and comparison between the general population and university students in terms of how they acquire and respond to traveler information. In Chapter 8, the results are discussed from an overall perspective and suggestions for further research are also given.

2. LITERATURE REVIEW

The objective of this chapter is to provide a brief summary of the methodologies used by various researchers and to present state-of-the-art efforts to combine spatial analysis with travel behavior.

2.1 Spatial Data Representation

The spatial data used in transportation studies has long been focused on using aggregated data, such as socio-demographic data based on predefined TAZ level. Using zonal data is a “container” view of space, in which space is reduced to a receptacle or carrier points of spatial phenomena (Fotheringham and Wegener, 2000). When spatial data is aggregated into larger zones, some important properties within zones, e.g. spatial distributions and spatial interactions are lost.

Studies have shown that using aggregate data based on zonal level can introduce spatial aggregation error to statistical models. The agglomeration of individual, geo-referenced observations into larger spatial zones can smooth local variation, leading to errors in measurement of geographical variables, and this error in turn affects the estimation of statistical models that incorporate spatially-aggregated variables (Luo et al., 2010). Ecological fallacy (shows that the model coefficients estimated based on aggregated data differ from those at the individual level) and the well-known modifiable areal unit problem (MAUP) (Páez and Scott, 2005, Fotheringham and Wong, 1991) are two major problems related to aggregation error (Luo et al., 2010). Also, studies show that TAZ and census-tract-based analyses are too large to correctly reflect patterns of development on the ground, thus they yielded poor measures of nucleation and land-use mix measurement (Moudon et al., 1997). In addition, spatial aggregation was found to increase the magnitude of correlation coefficients in early studies (Gehlke and Biehl, 1934, Hillsman and Rhoda, 1978, Openshaw and Taylor, 1979).

The literature also indicates that the impacts of aggregate error may depend on different geographical levels and the study area (Luo et al., 2010, Hewko et al., 2002). Generally, the literature has suggested that using finer resolution data is the best way to reduce aggregate level errors (Hewko et al., 2002).

The capability to capture the exact location of either residence, work place or other travel destinations has played an important role in examining the spatial characteristics of daily household activity-travel behavior. Knowledge of exact geo-locations can facilitate:

- Geometric measurement to capture built environment characteristics or individual accessibility, e.g. urban design or connectivity measures focused on local network geometry (Brownson et al., 2009, Buliung, 2004, Dill, 2003).
- Accurate calculation of vehicle miles traveled or daily household kilometers traveled (DHKT) (Buliung 2006).
- Activity space calculation, e.g. Household Activity Space (HAS) (Buliung 2006), spatial footprint analysis (Fan and Khattak, 2008a), spatial dispersion measurements such as Standard Distance Circle (SDC) and Standard Deviation Ellipse (SDE) (Buliung, 2004, Schönfelder and Axhausen, 2003).

- Exploration of spatiotemporal behavior of individuals and calculating space-time accessibility measurements, e.g. household trajectories in space–time (Miller, 2005) and space-time prisms (Kim and Kwan, 2003).

As geocoded data are increasingly used to link to socioeconomic, demographic, or built environmental data to meet the needs of the analyses mentioned above, an important preliminary step is the assignment of a point-level location to the address of each entity — a process known as geocoding. However, researchers have mentioned potential positional error introduced through geocoding, and these geocoding errors can result in considerable bias in spatial analyses (Rushton et al., 2006, Zinszer et al., 2010, Hay et al., 2009). In particular, the existing geocode method, like street geocoding (interpolated method), is rarely completely successful in practice. In fact, it is common for 10%, 20%, or even 30% of the addresses to fail to geocode using standard software and street files, and this proportion can be even higher for particular sub-regions of interest (Zimmerman, 2008).

Despite the geocode errors, the more important issue is that the exact geocode information is usually unavailable in the public use database due to disclosure avoidance. This is the case for widely-used surveys such as the US Census and American Community Survey (ACS), including the Census Transportation Planning Package (CTPP) (Krenzke et al., 2011). Additional privacy concerns arise as a consequence of being able to accurately represent the location of individuals using geocoding as well as the ability to link disparate data sources. Geographic identifiers support such linkages because data are easily combined when common identifiers such as names, phone numbers, driver's license numbers, or home or work addresses are present in different databases (Rushton et al., 2006). Therefore, various technologies have been used to provide masked data to avoid violating the pledge of confidentiality. These technologies include data swapping, rounding (Fienberg et al.), collapsing categories, applying thresholds, table suppression, and generation of synthetic data (Zayatz et al., 2010). Also, census surveys often release more detailed information based on larger geographic units, e.g., more information can be found at the census tract level than the census block level (Zayatz et al., 2010). Due to changes in technology and the continued concern for loss of privacy, the Census Bureau has instituted a Disclosure Review Board (DRB), which reviews all tables before their release to ensure confidentiality of responses. Data dissemination rules initiated by the DRB have caused significant loss of detailed spatial data for CTPP 2000, and are expected to cause even further losses when applied to products related to the American Community Survey (Cambridge Systematics et al., 2009). For the National Household Travel Survey (NHTS) and its add-on surveys, the location information has been masked and only aggregated zonal information is typically available to researchers instead of exact lat/long point-based information. Generally, disclosure avoidance measures pose significant challenges for exploring activity participation and travel behavior at a micro-scale spatial level, despite the substantial improvement in spatial analysis methodologies.

To overcome the problem of lacking exact location, several methods can be applied to disaggregate spatially aggregate data within a spatial unit such as an urban district or a census tract. Raster cells or pixels can be used as addresses by considering different densities within the zone (Fotheringham and Wegener, 2000), but this method requires density parcel data and cannot represent point address directly.

Besides using raster or pixel data, geographic imputation can be applied to assign point based location, which has been applied in various fields, especially in epidemiology and the medical field. Specifically, geo-imputation is usually used to deal with data that is missing location information. Researchers exercise options to exclude cases with missing location information or to include them and assign locations with a lower level of spatial precision (Henry and Boscoe, 2008). However, excluding cases may cause geographic selection bias (Henry and Boscoe, 2008). One method used for geo-imputation is to assign the persons or households in proportion to a geographic unit's (e.g., town's) population or other variables that include race, age and gender-specific population distributions within the tolerances of available information on geography (e.g., postal ZIP code or county) and demography (Henry and Boscoe, 2008). However, this method can only disaggregate larger zonal level data to finer zonal level, but it still does not provide exact geo-location information. To circumvent this, the centroid (the geometric center of a polygon) or weighted centroid point can be used for geo-imputation (Hewko et al., 2002). Furthermore, synthetic assignment has been used to give a random location to the entity. For example, random locations can be assigned to observations within polygons, and the process can be repeated many times using Monte Carlo simulation to estimate associated uncertainty (Luo et al., 2010). Generally, the literature has suggested that using finer resolution data is the best way to reduce aggregate level errors (Hewko et al., 2002). However, the key question of whether geo-imputation results in statistically significant errors, in general, has not been answered.

2.2 Spatial Statistics in Transportation

Spatial models in the social sciences have a long tradition, which date back to 1820s (Fotheringham and Wegener, 2000). However, because spatial models are data-hungry, the real rise of spatial modeling occurred in the 1960s with the general availability of large, fast computers (Fotheringham and Wegener, 2000). Interest has increased in how to link spatial analysis, GIS technology and transportation (Miller, 1999, Buliung and Kanaroglou, 2007) since transportation data often has spatial attributes, while conventional database systems cannot make much use of the spatial or location attributes of a data set (Taylor et al., 2000).

The general philosophy of capturing spatial patterns in transportation focuses on capturing spatial dependence and spatial heterogeneity. Bhat and Zhao (2002) highlighted the need to accommodate spatial issues in travel modeling, and proposes a specific spatial model formulation (mixed ordered logic model) in the context of activity stop generation. Their results underscore the importance of accommodating and testing for the presence of unobserved heterogeneity in the modeling of stop-making decisions. Significant heterogeneity in the response to some factors was found. For instance, propensity to stop for shopping relates to the level of accessibility, but this effect may be important only when accessibility levels are low. However, this study used a zonal based data; therefore, it only captured the heterogeneity across zones (Páez and Scott, 2005).

Spatial dependence, referred to as the spatial autocorrelation problem, describes the situation when there is a tendency for variables to display some degree of systematic spatial variation (Páez and Scott, 2005). Furthermore, the spatial dependence violates the independence assumption of regression, which can cause bias and misspecification in the model. For example, all the cases that are very poorly fitted by the model might be in one part of the map. A common specification in the spatial analysis literature for capturing such spatial correlation is to allow alternatives that are contiguous to be correlated (Bhat and Guo, 2004). Spatial lag operator (contiguity weight matrix which represents the contiguity of observations) was also included in the error term. If the coefficient of this spatial lag operator is statistically significant (5% level), the spatial dependence is important to address. A positive coefficient means that similar observations are clustered. Alternatively, significant negative spatial correlation indicates that neighboring observations are more dissimilar. Significant research (Garrido and Mahmassani, 2000, Bhat and Guo, 2004, Kwigizile and Teng, 2009) has been conducted studies to capture this spatial dependence.

Spatial heterogeneity describes another situation when the mean, variance or covariance structure changes over space (Páez and Scott, 2005). Traditional regression models assume that the dependent variable has the same variance for all correlates, which is often not the case in real-life transportation situations. If spatial heterogeneity exists and is not accounted for, then the model may have biased parameters, misleading significance levels and/or inaccurate forecasts (Páez and Scott, 2005). Geographically Weighted Regression (GWR) is a tool that captures spatial heterogeneity by estimating model parameters locally instead of globally (Fotheringham et al., 2002, Páez et al., 2002, Lloyd, 2007). In GWR, the estimated parameters, which capture associations of variables (e.g., association of congestion or socioeconomic factors with information acquisition), can vary over space. The local parameters are estimated for each variable in a spatial context. In doing so, more detailed local associations of variables are provided and the key assumption of global models, where "one size/model fits all" is relaxed. Furthermore, GWR is often interpreted as a "smoother", which can be used to approximate the observed variable surface to a higher level of accuracy, a feature that makes it attractive for various aspects of urban analysis (Páez and Scott, 2005). From a policy maker's angle, GWR can improve regional analysis and policy making since the subsequent policy inferences would be poorly suited to many local settings (Ali et al., 2007).

Although the GWR model has been used in other fields such as social science, environmental, investigation of industrialization, etc., studies of GWR applications in transportation are less common. Zhao and Park (2004) applied GWR to estimate AADT (Annual Average Daily Traffic) on non-expressway roads based on available AADT information on similar roads; the results indicated that the GWR models are able to better explain variations in data and to predict AADT with smaller errors than the OLS (Ordinary Least Squares) model. Chow et al. (2006) investigated the spatial variations by using the GWR model to estimate the relationships between transit use and potential ridership predictors. Results indicated that the GWR model improved accuracy in predicting transit use for HBW (Home Based Work) purposes over linear regression models. Du and Mulley (2006) looked at the relationship between transport accessibility and land value with the implication of a local model and GWR, which revealed that nonstationarity existed in the relationship between transport accessibility and land value. Clark (2007) found the local model produced by GWR is more accurate in estimating the relationship between income and car ownership. Hadayeghi et al. (2003) found that GWR was a significant improvement over the global model when using GWR in accident prediction models to test spatial variations in the estimated parameters from zone to zone.

Overall, several studies have applied GWR successfully, and improvements have been found in goodness of fit and forecasts over other traditional global models.

2.3 Spatial Analysis and Travel Behavior

The relationship between accessibility, urban form, built environment and travel behavior has had a rich history over the past decade. Examining the spatial characteristics of daily household activity-travel behavior has important implications for understanding and addressing urban transportation issues (Buliung 2006). Understanding spatial patterns, activity participation, and their relationships is a primary objective in the travel behavior research agenda. There are several studies which review relevant work (Krizek, 2003, Chatman, 2005, Ewing and Cervero, 2001, Ewing and Cervero, 2010, Joh, 2009, Crane, 2000, Van Wee, 2002). However, there is debate about whether land use or specially designed community with certain characteristics of built environment is associated with special travel behavior. Some studies support the notion that connections between land use and travel behavior exist (Shay and Khattak, 2005, Khattak and Rodriguez, 2005, Kockelman, 1997, Parthasarathi, 2011, Fan, 2007), while others challenge them (Boarnet and Crane, 2001, Crane, 2000). Despite the clear evidence of a connection in those studies supporting the connection, some argued that the association between built environment and travel behavior is not direct, even weak (Boarnet and Sarmiento, 1998, Crane and Crepeau, 1998, McNally and Kulkarni, 1997). To a great extent, personality, attitudes, and socio-economic factors are stronger correlates with travel behavior than land use variables (Stead, 2001, Cao, 2009, Handy et al., 2005). Specifically, results from Boarnet and Crane (2001) indicated that urban form influenced travel behavior (if the influence exists) not directly, but through altering the price of travel. These inconsistent findings further indicate a complex relationship between urban form and travel behavior, which may be sensitive to different geographical scales, types of data, alternative behaviors and statistical assumptions.

Still, many believe that through better design, the built environment will impact travel behavior and reduce automobile use. Srinivasan (2001) found that spatial characteristics measuring nonwork opportunities (nonwork accessibility and commercial-residential balance) were significant in affecting incremental travel time and trip linking during home-based nonwork tours, as corridors rich in nonwork opportunities and close to highways tend to encourage trip linking, but destinations or home locations with such characteristics tend to discourage trip linking. Also, pedestrian and transit-oriented design and developments that encourage proximity of commercial land uses (such as banks and shops) to residences have been found to be negatively associated with single-occupant vehicle use (Kockelman, 1997). Many cities, especially metropolitan regions, have implemented land use policies to guide long-term transportation plans with the purpose of VMT reduction, congestion mitigation and air quality improvement.

Geographic Information Systems (GIS) are convenient platforms for theoretical and applied transportation and urban analysis (Miller, 2003). A specific branch of GIS applied to transportation issues, commonly known as GIS-T, has emerged as a new area in the last couple of decades (Miller and Shaw, 2001). New GIS technology is being applied to create new spatial variables to capture the built environment, e.g., network connectivity (Dill, 2003, Parthasarathi et al., 2011), network topology (Xie and Levinson, 2007), accessibility (Xie and Levinson, 2007, Thill and Kim, 2005, Fan and Khattak, 2009), land use mix (Kockelman, 1997), monocentric and polycentric urban structures

(Veneri, 2010, Schwanen, 2001), and spatial structure of neighborhoods and transportation corridors in metropolitan areas (Srinivasan, 2001, Srinivasan, 2002). Generally, collecting spatial data is expensive and time-consuming. Therefore, combining built environment characteristics with travel demand is not common in practice—although finer spatial detail (than TAZ level) is increasingly being captured to model in some of the regional travel demand models, e.g. NYMTC (New York), ARC(Atlanta) (Vovsha et al., 2004).

2.4 Spatial Analysis and Travelers' Response to ATIS

Studies regarding travel information acquisition and the impact of information on travelers and transportation system are abundant (Polak and Jones, 1993, Levinson, 2003, Toppen et al., 2004, Toledo and Beinhaker, 2006, Chorus et al., 2007a, Chorus et al., 2007b, Wang et al., 2009, Zito et al., 2011, Choo and Mokhtarian, 2007, Dia and Panwai, 2010). Given the multitude of studies in this area, there are also comprehensive reviews of the literature (Lappin and Bottom, 2001, Chorus et al., 2006a).

Not all travelers seek traffic information to facilitate their travel decisions. The 2006 Greater Triangle Household Travel Survey data showed that about one-half of the respondents (49%) reported that they did not acquire travel information from electronic sources and never seek regional travel information (Khattak et al., 2008). Investigation in the San Francisco Bay Area (Khattak et al., 2003a) also suggested that a significant gap exists between access and use (100% vs. 66.4%) . A panel survey of the Seattle-area in 2000 (Peirce and Lappin, 2003) showed that travel information was used by 12% of the survey respondents and 3.2% of all the trips conducted by the respondents. The survey also found that the most common source of travel information is radio traffic reports.

Existing evidence shows that travel information usage is associated with various factors, including the traveler's knowledge (Peirce and Lappin, 2003), owning electronic devices such as mobile phones and being willing to use internet (Polydoropoulou et al., 1996, Peirce and Lappin, 2003, Yim et al., 2002). However, personal characteristics such as gender and income are not always found to be significantly associated with information usage (Peirce and Lappin, 2003, Petrella and Lappin, 2004). Trip characteristics are found to influence travel decisions substantially especially when: (1) the distance and duration of the trip is longer (Kitamura et al., 1994, Englisher et al., 1996, Peirce and Lappin, 2003, Fan and Khattak, 2008b) (2) the trip is arrival-time sensitive (Peirce and Lappin, 2003) and (3) substantial variability or uncertainty exists about travel times(Peirce and Lappin, 2003, Chorus et al., 2007b).

Among travelers who seek traffic information, only a subset adjusts travel decisions. Investigation in the San Francisco Bay Area showed that 33.1% of respondents changed their decision (Khattak et al., 2003a). The number in Greater Triangle area was 34.6 % (Khattak et al., 2003b), and in Seattle, about 37 % of the information-using trips also involved some resultant change in travel behavior (Peirce and Lappin, 2003). A study of "SmarTraveler" users in the Boston area found that about 30% of the users changed their travel behavior "frequently" in response to information and 96 % changed their trips "occasionally"(Englisher et al., 1996).

Researchers have found multiple factors associated with travel decision changes: journey related attributes such as unexpected delays or travel time congestion, e.g., due to incidents, roadway construction, or special events (Khattak et al., 1996, Polydoropoulou et al., 1996, Chorus et al., 2007a); different information form (Wang et al., 2009, Polak and Jones, 1993, Khattak et al., 2008); personal attribute such as gender, age and income (Polak and Jones, 1993, Mannering et al., 1994, Abdel-Aty et al., 1997); cognition or travelers' knowledge (Adler and Blue, 1998); contextual and national factors (Polak and Jones, 1993); and the frequency of information and travel information usage (Khattak et al., 2008).

Some of the models available in the literature have considered spatial characteristics (Fan and Khattak, 2008b) by including categorical variables of land use type in the correlation, such as land use type or density measurements as regressors. However, most of the land use measurements are still based on predefined land units, e.g., traffic analysis zones. The edge effect still remains, that is to say the geographical patterns between the correlations may not be consistent with these defined land units. The spatial measurements based on a finer scale were barely considered as an explanatory variable in traveler information delivery mechanism literature.

2.5 Summary of Literature Review

Much of the focus has been placed on understanding links between non-spatial travel activities and socio-demographic factors by researchers as well as practitioners. Research on both spatial analysis and activity based travel behavior model has been insightful, yet combining these methods still needs exploration:

- Although there is consensus that using disaggregate data to estimate discrete models can obtain better results than aggregate models, modeling the spatial data based on accurate point location is not easy due to the lack of point-based data, which involves private information disclosure issues. To alleviate this problem, researchers have used centroid of geographic boundary to represent the locations within that boundary when there is need to capture built environment based on the locations. However, no one has evaluated how accurate this method is, since it seems to be the best solution given the present circumstances. Furthermore, to the best of the author's knowledge, no imputation method has been proposed as a better data representation method to create synthetic residences with geographic information.

- As stated previously, current models relevant to studying associations between built environment and travel behavior as well as studies of traveler information delivery mechanism have relied on traditional statistical models where the associations are fixed in the study region. This lacks spatial interpretation and cannot be visualized as an interactive thematic map. More importantly, it hides possible important implications of these associations, e.g. spatial heterogeneity. It is not clear whether considering spatial heterogeneity is necessary when exploring the factors impacting travel decisions, including trip making and travel plan changes in response to ATIS.
- Using aggregated zonal characteristics such as population density, average characteristics of a geographical boundary, or circular buffers around a residence to measure land use variables does not accurately capture the built environment characteristics around a residence. Moreover, no practice that incorporates spatial variations in built environment with trip generation can be found. There are substantial gaps and potential for improvements in knowledge of how built environment correlates with travel.

- It is very common to conclude, in general, by using travel behavior surveys based on a relatively large scale, e.g. city or using the survey based on local community. However, comparative study is less common, e.g. how a certain group of population looks different from the general population in terms of how they travel and respond to ATIS, especially when they are exposed to different spatial contexts.

3. RESEARCH DESIGN

The purpose of this chapter is to provide a conceptual framework for discussion. The questions of interest here are the links between the spatial patterns, travelers, and their travel behavior, which not only refer to their travel decisions, but also includes their responses to the advanced traveler information systems.

3.1 Overview of Conceptual Structure

Two concepts are fundamental to both spatial analysis and transportation: site and situation. Site refers to the geographical characteristics of a specific location, while situation concerns the site's relationships with regard to other locations. Therefore, modeling the site and situation lies in the core of integration of spatial analysis and travel behavior. Meanwhile, emphasis is also placed on addressing the spatial pattern of different scales. Figure 2 presents the study objects of different scales. On the residence level, emphases have been placed on modeling the site from a microscopic perspective, e.g. capturing the built environment around the residences. On the district level, consideration has been given to both site and situation, while on the city level, modeling the situation of its components is highly significant and should be focused upon.

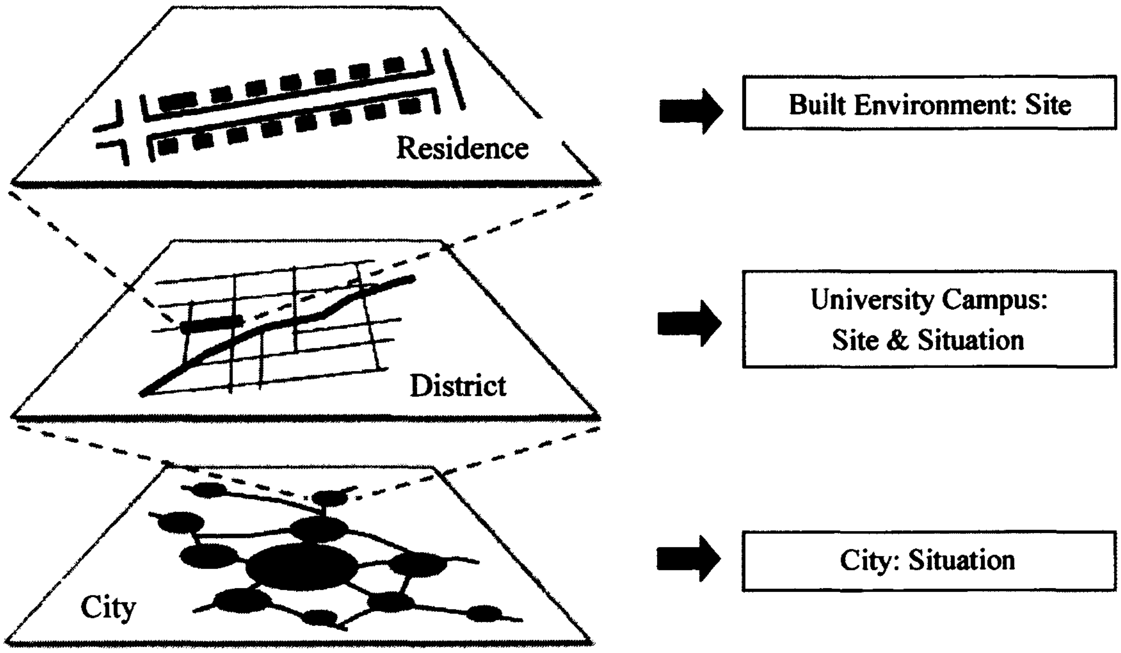
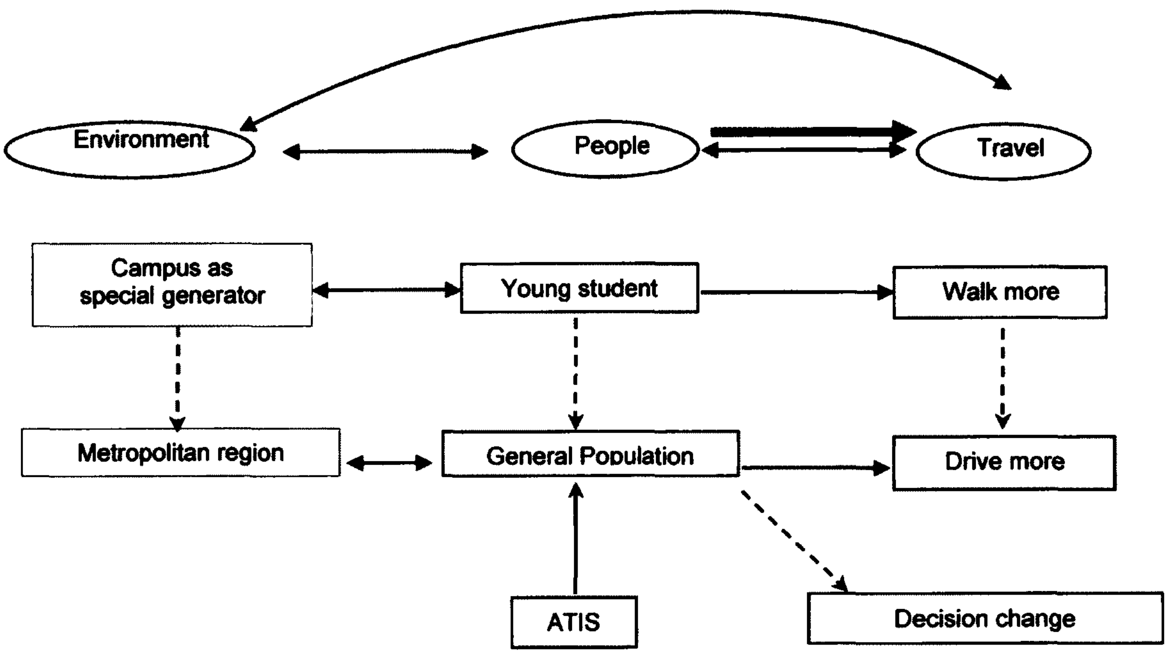


Figure 2. Study objects of different scales

3.1.1 The Model Structure

In an urban transportation system, people who travel, their travel behavior, and the environment of space in which they travel are three important dimensions. To better understand their interrelationships, the Social Cognitive Theory (SCT) (Bandura, 1986) is borrowed here. It hypothesized that there are reciprocal relationships between the characteristics of a person, the behavior of a person, and the environment in which the behavior is performed. However, the two-way influence between those three elements does not mean the strength of the associations between each pair is perfectly symmetrical, nor does it mean that the interactions happen simultaneously (Bandura, 1986). Similarly, it is a hypothesis that travelers, their travel behavior, and the environment (includes not only the physical environment but also the soft environment, e.g. information technology development) are intertwined with each other. Figure 3 is a simplified representation of the interplay between the elements. Since the causality of these relationships cannot be effectively captured in statistical models, correlation can instead be captured. The emphasis is to study the travel behavior and the factors related to it.



⟶ denotes association ----> denotes evolvement

Figure 3. Conceptual framework

More interestingly, there is a sequential evolvement in travelers' life stages and their travel behaviors. As the life stage changes, the environment surrounding the traveler also changes, which is accompanied by a different mobility level, and consequently, the different travel behavior. Assume there is a traveler. When he was still a single undergraduate student, he lived on-campus in order to have greater access to educational and other activities and also to interact socially with his peers. The campus acted as the core of his daily life, akin to a relatively insulated and physically bounded environment. As most of his trips were of a short distance, he usually walked or bicycled. Later on, he enrolled in a graduate program and got married. He no longer lived close to campus, considering the need to seek balance between his study and work. The campus then acted more like a routine anchor, not necessarily being the core of his daily life. He had to drive much more due to the distance between school and home, but he still walked when he was around the campus. Later on, like other people in the general population, he had a job and a family after graduation, prompting his decision to move to a suburban community. He needed to find a balance between travel, work, and personal business, e.g., parenting responsibilities. Driving alone and carpooling with his family became his dominant trip mode. In short, from a traveler's life changes, travel behavior with substantial evolution over time can be observed. Similarly, in the same urban region with a mixed population, it is important to study the disparity of travel behavior among different subgroups of the population. A case study for a certain subgroup of travelers may improve understanding of special travel needs and subsequently benefit the accuracy of the regional travel demand model.

Besides the long-term travel behavior evolvement based on personal life changes mentioned above, another evolvement of travel behavior is short-term travel decision adjustment based on the traveler information received. This change has been brought by the unprecedented development of traveler information distribution in recent years. Compared with the last decade, both personal electronic devices and information technologies have advanced considerably. Today, people can quickly access the internet or connect with a GPS service through portable devices such as smart phones and tablets, via which traveler information can be acquired almost anywhere, anytime. This soft environment represented by the rapid availability of information has changed people's travel decisions by distributing both pre-trip and real-time en-route travel information more quickly, widely, and effectively. To sum up, the accessibility to travel information and its possible impact to travel behavior is worth studying.

3.1.2 A Special Trip Generator

The conceptual structure of a special trip generator with an alternative friendly environment is presented in Figure 4. The hypothesis is that a ring of mobility associated with this special trip generator exists. If a university campus is the special trip generator here, the figure shows three rings, each characterized by specific mobility considerations.

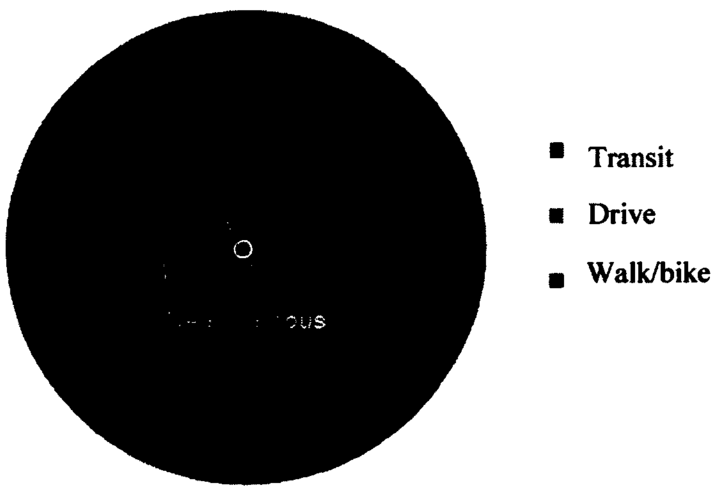


Figure 4. Conceptual structure of a special trip generator

Figure 4 shows three different rings according to how far it is from the center. The center of the special trip generator (A) is the core area, similar to an on-campus environment, which often related to a mixed land use and alternative mode friendly environment. In such a context, the pedestrian space is the dominant trip mode, as most origins and destinations are accessible by sidewalks and bicycle facilities including bicycle path and parking racks, which favor non-motorized trips. The ring area around A is the near-generator area (B), which represents areas adjacent to core areas. The walking and bicycling space may lose some of its importance compared to the core area, but it is still dominant in this area. The outer ring area is the peripheral area (C), where mobility is dominantly provided by motorized transportation, with walking and cycling servicing very few residual functions, which are often leisure-oriented.

3.1.3 The Bispace Model

The premises in the previous section are only a mono-dimensional model if only non-spatial attributes such as social-economic properties without a spatial label are considered in the model. In a mono dimensional model, there is no need to consider how the dependent and independent variables are distributed on a map; they are both unmappable. However, if the location information is added into this mono-dimensional model, it becomes bi-dimensional; as for every variable, how it looks from space is stressed. That is to say, in the bispace model, both the site and situation are captured. The bispace model is necessary, because if observing from space, all the three elements show spatial patterns: land use clusters partially due to zoning systems; the population clusters based on people's economic status, e.g. communities with properties of different price level; and the environment and personal characteristics that may influence the travel decision together or separately, which shape travel behavior clusters in space. This tendency usually is more distinct with a conglomeration of certain groups of populations or special land use in space, such as colleges or military bases. As all three elements are embedded in space, spatial impacts are imposed on the site and its relationships. To this end, the mono-dimensional model seems plain as it discards the spatial projection of the data and the spatial relationships among the data. The mono-dimensional model yields an incomplete analysis of the data and the related factors.

Therefore, a bispace model must satisfy one or both of the following criteria: 1) capture the site-specific characteristics of the sample (site); 2) capture the spatial distribution of the samples and their associations (situation). Therefore, a spatial model actually becomes a model of the object investigated in bispace (space, attribute). Its outcomes are dependent on the geographical positioning of objects within the model. That is to say, the results from a spatial model will not remain the same when we rearrange the sample to change the spatial distribution of values under investigation.

3.2 Overview of Methodologies

The methodological scope of this study is relatively detailed and extensive as the dissertation is an interdisciplinary study, integrating advanced GIS technology and spatial statistics into travel behavior analysis. A wide variety of different research methods are applied for answering the research questions, including:

- Literature reviews;
- GIS data mining and data processing: Analyzing quantitative spatial data, including econo-demographic data from census surveys, travel survey data, land use data. Behavior data with geographical information are stored and managed by GIS software. GIS software is used to combine data from different sources based on geographical location despite the different format from their original sources. This can aid the data processing step in the later

analysis, e.g. error and outlier identification, proximity analysis (buffer area analysis), built environment variable creation, network analysis, data clustering and extracting. Geo-imputation is used to assign synthetic locations to samples whose exact locations are unknown but are needed for research purposes;

- Cross-sectional study: it is used to capture the various associations in this study. These associations include those between built environment and trip making; those between students' personal and residential characteristics and trip making; and those between various factors and people's response to ATIS, etc. A shortcoming of using cross-sectional surveys is that only a picture of associations at one specific point in time is obtained, and it may not reveal long-term variance clearly. Though effort is made to compare young college students with the general population to emphasize the difference in travel preference due to different life stages, the data is not a time-series and demonstrates only a very general tendency of travel decision made by younger college travelers versus travelers from the general population.
- Spatial model: it is used to define the relationship between travel behavior in geographic reality and how that reality is captured in the form of a statistical model containing both associations between variables and emphasizing these associations on locations. To meet this need, the spatial models are combined with a cross-sectional study.

- Spatial analysis: includes the development and application of statistical techniques for the proper analysis of spatial data which, as a consequence, makes use of the spatial referencing in the data.
- Visualization of modeling results: instead of showing the associations obtained from models merely by mathematical equations, it is important to display the spatial pattern using GIS visualization. Interactive GIS visualization also allows for the ability to explore different layers of the map, to easily target research region by zooming in or out, to do a query or to change the visual appearance of the map based on the need for different themes. Since visualization can provide a more direct picture of how factors or associations change in space, it can benefit policy or decision making also.

In combination, these methods aim to provide a coherent and integrative answer to the research questions. In summary, two relationships are emphasized in this dissertation: 1) the association between travel behavior and built environment (physical environment); 2) the association between travel decision and ATIS (soft environment). Two cases are discussed around the topics: 1) the general population and 2) the university student. Three scales are explored: 1) residence level; 2) generator level; and 3) regional (city) level. A summary of different perspectives and relative methodology is presented in Figure 5.

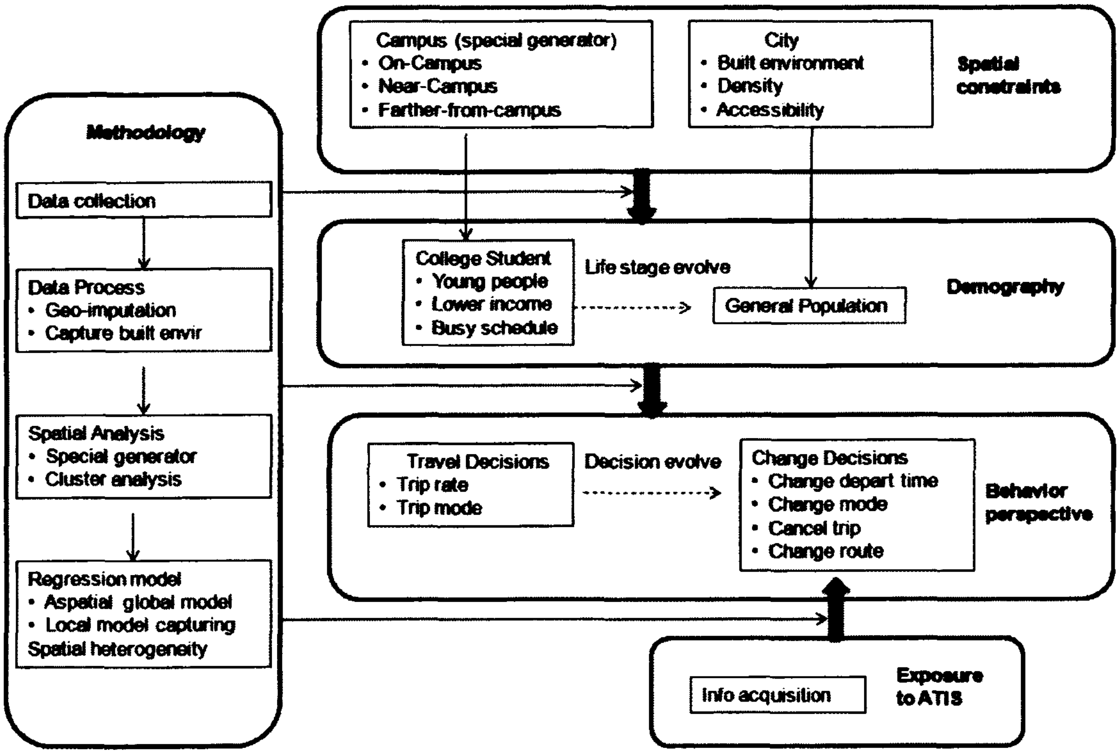


Figure 5. Summary of methodology

3.3 Methodology Specifications

The methodologies used in the following chapters are presented here:

3.3.1 Geo-Imputation: a Synthetic Method

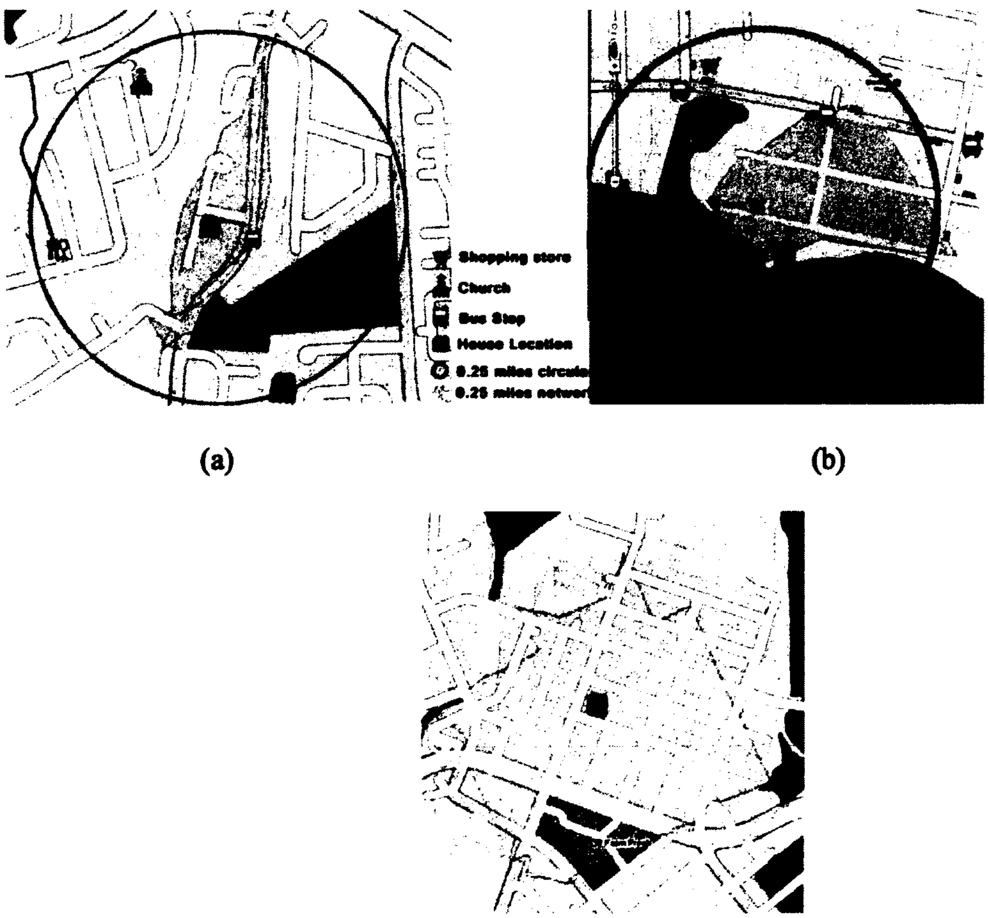
Geographic imputation is a method used to assign synthetic point locations to geographical data which have only zonal information instead of exact point information (latitude/longitude). Therefore, it is an effective method to disaggregate data from a larger geographic unit. Random points are first assigned within the zone given the zone ID is known. Random assignments are repeated several times, similar to a simulation procedure. Later on, the random assignments are compared with each other to check whether different assignments lead to similar results. If similar results can be obtained, it means the random error could be neglected, and the synthetic points can be used to replace the actual location. Since geo-imputation is a fundamental method used by this study to process spatial data before modeling, determining its accuracy is important; therefore a special discussion is provided in Chapter 4. The data used by later chapters will be processed by applying this method.

3.3.2 Network Based Buffer: Capturing Built Environment

Studies of spatial analysis often use circular buffers around residences to measure land use around a residence, including buffers with a radius from 0.25 miles to 1.0 mile (Brownson et al., 2009). However, a circular buffer around a residence is not accurate enough to capture built environment as it may contain area and roadway segments which are not accessible to this residence, even contain unusable land use such as a large body of water.

Furthermore, with a fixed radius buffer, a circular buffer does not fully represent the local accessibility of a residence, since accessibility mainly depends on the network instead of how far it can be covered by direct distance. Therefore, different from the circular buffers with a fixed distance, a network buffer around a residence is created by connecting the farthest points along the roadway which areaccessible to the residence within a fixed travel distance. A benefit of the network buffer is that it is adaptive to the roadway around the residence with no fixed shape.

Figure 6 shows some examples of using circular buffers versus a line-based network buffer, all with a buffer size of 0.25 miles (equivalent to 15 minutes walking distance). From the graphs, although the church and restaurant are within the 0.25-mile circular buffer of the residence, they actually belong to the other community which does not have a direct road connection with the residence. Similarly, there is only one bus stop within 0.25 miles in Figure 6 (b) but a circular buffer will cover two bus stops. Thus, it can be seen that using a circular buffer cannot capture the actual built environment accurately. Also, a larger area of the network buffer usually represents a grid network around the residence with higher local accessibility since the network buffers are closer to a circle when a residence has accessibility in all directions with a grid style neighborhood, as shown in Figure 6(c).



(a) (b)

(c)

Figure 6. Examples of using different buffers

Several measurements can be calculated to capture the neighborhood type within the network buffer, including total length of all the roadways, number of intersections, number of cul-de-sacs, and the area of the network buffer. Land use characteristics have shown greater associations with walking using line-based road network buffers than circular buffers (Oliver et al., 2007). Also, researchers need to carefully consider the most appropriate buffer with which to calculate land use characteristics. Literature in epidemiology has applied network buffers to explore correlations of built environment with physical activity and health issues such as obesity (Brownson et al., 2009). However, network buffers have not been used widely in transportation for measuring spatial characteristics.

3.3.3 GWR: Spatial Heterogeneity

Spatial regression methodology is used in this study to capture spatial heterogeneity, also known as spatial variance of association. Geographically Weighted Regression (GWR) is a non-parametric methodology used for the investigation of geographical drifts in regression parameters. Specifically, GWR relaxes the assumption that estimated parameters in traditional regression models hold globally. Note that GWR performs a regression for each residence i using a subset of the residences that are spatially proximate to i; this nearby area is named "kernel," similar to a buffer area (shown in Figure 7). The size (distance in space) of the kernel is termed "bandwidth". If fixed bandwidth is used for every regression location, then the number of residences (local sample size) for each regression will be different as residences are not usually distributed evenly in the space. In areas with higher residential density, the local sample size will be larger, while in areas with sparse residences, the local sample size will be smaller (demonstrated by Figure 8). This causes a problem for some residences, as there are no other residences in its kernel when a fixed bandwidth is used. To fix this problem, an adaptive kernel can be used, which ensures the bandwidth is selected so that each residential location in the sample has the same local sample size.

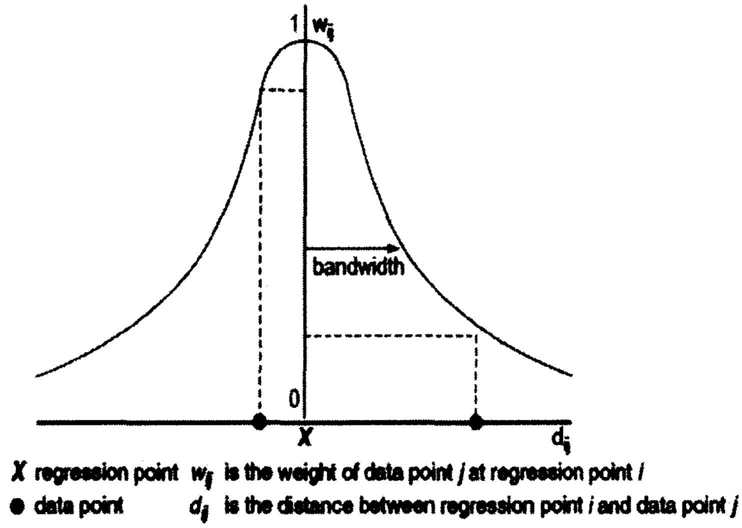


Figure 7. Definition of kernal and bandwidth
Source: (Fotheringham et al., 2002)

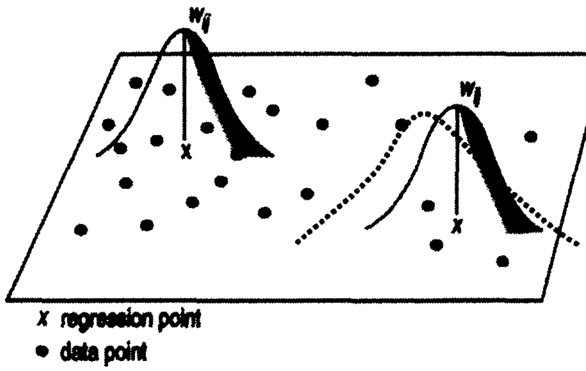


Figure 8. Demonstration of local regressions
Source: (Fotheringham et al., 2002)

After the kernels are chosen, the local models are weighted and estimated by using sub-samples (households) within the kernel of each regression location, where the weights of each household are inversely proportional to their distance from the regression location. This means that the households near the regression location (household i) contribute more in the local model. For each location, both the sub sample within the kernel and the weights are different so that the results of local calibration are unique to a particular location. By plotting the results of these local calibrations on a map, surfaces of parameter estimates, or any other display which is appropriate, can be generated. Therefore, when GWR is applied, key decisions must be made regarding 1) a weighting function (the shape of the kernel), and 2) the bandwidth of the kernel (the size of the kernel). The weighting function usually has a minimal effect on results, while bandwidth may affect results markedly (Lloyd, 2007). Only if there is little variation in the local observations do the global observations provide reliable information on the local areas.

For each residential location, both the sub sample within the kernel and the weights are different so that the results of local calibration are unique to the particular location. In its most basic form, the GWR model is described as follows (Fotheringham et al., 2002):

$$y_i = \beta_{i0} + \sum_{k=1}^{p} \beta_{ik} x_{ik} + \varepsilon_i \qquad \text{(Equation 1)}$$

y_i = dependent variable for sample i (i = 1, 2, . . . , n, where n is the number of observations);

β_{i0}= constant for sample i;

β_{ik}= the parameter at location *i* for explanatory variable x_{ik};

x_{ik} = explanatory variables of the kth parameter for residence *i*,

ε_i = error term at location *i*,

p = number of estimated parameters.

Considering different type of dependent variables, e.g. trip frequency is a non-negative count variable, whether access ATIS or not is a binary variable, different forms of GWR are provided here.

For Logistic GWR:

$$\ln(\text{odds ratio of Probi}) = \ln(\beta_{i0} + \sum_{k=1}^{p} \beta_{ik} x_{ik} + \varepsilon_i) \qquad \text{(Equation 2)}$$

For Poisson GWR:

$$\ln(y_i) = \beta_{i0} + \sum_{k=1}^{p} \beta_{ik} x_{ik} + \varepsilon_i \qquad \text{(Equation 3)}$$

Note that for each location, the β parameter can be different. The model is fitted using a technique known as iteratively reweighted least squares (Fotheringham et al., 2002). Adaptive kernel, the bi-square kernel, is used to calculate the weights. Adaptive kernels are useful when there is a large variation in the geographical density of the observed data (Fotheringham et al., 2002). The kernels have larger bandwidths where the data are sparse in space and have smaller bandwidths in locations where samples are plentiful. The weights are defined by:

$$w_{ij} = \begin{cases} \left[1 - \left(||u_i - u_j|| / G_i\right)^2\right]^2 & if ||u_i - u_j|| < G_i \\ 0 & otherwise \end{cases} \qquad \text{(Equation 4)}$$

The parameter G (the bandwidth) regulates the kernel size, and $||u_i - u_j||$ is the distance between residential locations *i* and *j*. When calibrating the model, the kernel bandwidth is determined by minimization of the Akaike Information Criterion (AIC), which is a measure of goodness of fit and allows finding a model that best explains the data with fewer estimated parameters. The AIC of the model with bandwidth G is given by:

$$\mathrm{AIC(G)} = \mathrm{N}\ln(\mathrm{RSS(G)}) + 2\,\mathrm{K(G)} \qquad \text{(Equation 5)}$$

$$\mathrm{RSS(G)} = \sum_{i=1}^{N} \hat{\varepsilon}_i^2 \qquad \text{(Equation 6)}$$

Where RSS(G) is the residual sum of squares with bandwidth G; N is the number of observations; K(G) denotes the effective number of parameters in the model with bandwidth G, respectively. However, since the degrees of freedom for GWR models are typically small, a small sample bias adjustment in the AIC calculation is appropriate. The Corrected Akaike information Criterion (AICc) is then used to address this bias (Chow, 1976, Fotheringham et al., 2002). AICc is defined as follows:

$$\mathrm{AICc(G)} = \mathrm{AIC(G)} + 2\,\frac{\mathrm{K(G)}\,(\mathrm{K(G)}+1)}{N-\mathrm{K(G)}-1} \qquad \text{(Equation 7)}$$

Given the characteristics of non-parameter models (no fixed model for the whole sample), the likelihood ratio test cannot be done to evaluate these models. The current statistical tests that answer whether the GWR model is better than a global regression include Monte Carlo simulations and the Leung test. However, they are still questionable (Páez et al., 2002). While the Lagrange Multiplier (LM) test is reported by some researchers as a better test (Páez et al., 2002), it cannot be done using available commonly-used estimation software. AIC is widely used to compare global models with local models, or to assess local models with local models with different bandwidths (Fotheringham et al., 2002, Lloyd, 2007). Improvements in the AIC that are larger than 2 or 3 are typically used in relevant literature as criterion to judge whether the improvements due to local models are large enough (Fotheringham et al., 2002, Lloyd, 2007). If the effective number of parameters K is small relative to the number of observations N, then the difference between AIC and AICc is negligible (Chow, 1976).

4. THE ACCURACY OF GEO-IMPUTATION

As stated in previous chapters, understanding travel behavior and its connection with infrastructure and land use is critical for travel demand modeling. Especially given the new developments in spatial analysis, integrating built environment variables in estimation of disaggregate travel demand models is gaining momentum. To better understand associations between travelers' behavior and their residential location and surrounding land uses/infrastructure, researchers first calculate built environment characteristics surrounding residential locations and use the variables as correlates in behavioral models. To create new variables in buffers surrounding a residence, exact geo-coordinates of survey respondents' residences are needed. This spatial information can then be integrated and analyzed to explain travel behavior of survey respondents. Unfortunately, conventional travel behavior surveys such as NHTS and census surveys do not publicly reveal the exact residential location of respondents due to confidentiality concerns. In such situations, if the respondent's zone of residential location is known, then zonal average socio-demographic measurements can be used as correlates in models to represent the average land use variables surrounding a specific residential location. However, using zonal averages can create measurement errors and reduce the local variation that may exist in reality.

This chapter presents a method: geo-imputation, which can overcome the problem of not knowing the exact geo-coordinate location of a household. It can assign household to an exact geo-coordinate location (lat-long). Analyses are conducted to evaluate whether such assignments of geo-coordinates are relatively accurate and if so what are the implications.

4.1 Data Description

Spatial information was extracted from the exact geo-coordinate level data from Charlotte, North Carolina (N=3,310 households), and the Research Triangle, North Carolina (N=4,724 households). Not all samples from the surveys are used for this analysis. Only those with TAZ information are used; therefore, the final dataset is composed of 4,724 households from the Research Triangle area and 3,310 households from the Charlotte area.

Figure 9 shows the study areas and sampled household locations in both regions. Other data used in this analysis include boundary files, e.g. TAZ, census block, tract, public maintained roadway, transit stop and other GIS based files.

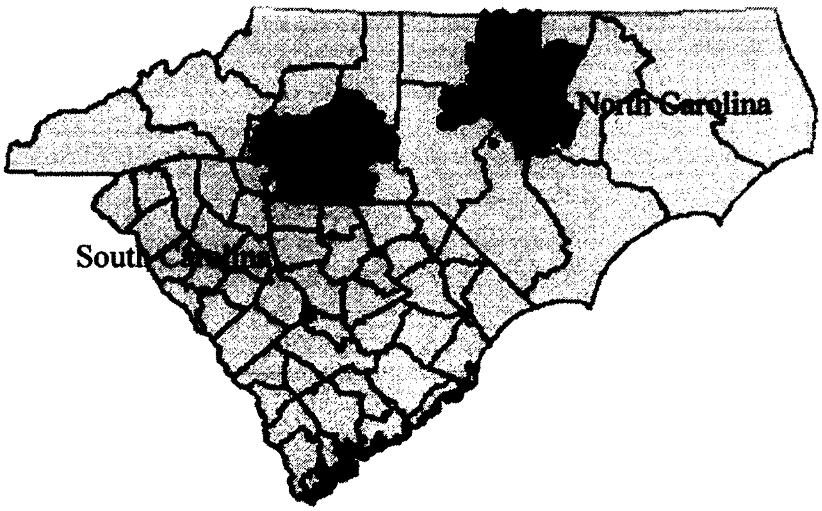


Figure 9. Study areas and actual residences

Among the 1,408 TAZs in the Research Triangle area that have at least one household sample, the number of sampled households in TAZs varies from 1 to 38, with the lower number in suburban regions and higher samples in urbanized regions. The central part of the Greater Triangle Area is highly populated, while the western part is mainly farmland with a few residential areas. Among the 1,422 TAZs in the Charlotte Area that have at least one household sample, the number of sample households in TAZ varies from 1 to 19, with the highest sample density concentrated in the central part.

4.2 Methodology of Geo-imputation

4.2.1 Framework

The methodological framework for the geo-imputation is presented in Figure 10.

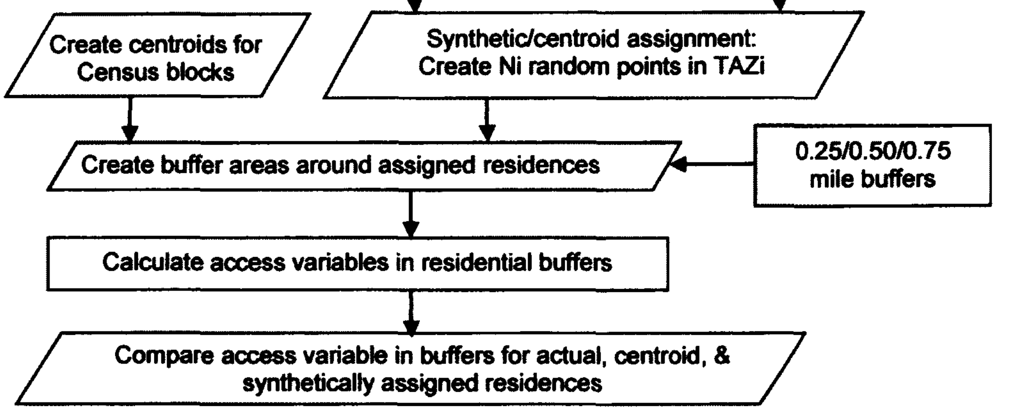


Figure 10. Methodology framework of geo-imputation

A key issue is that of agreement or concordance between accessibility measures based on actual geo-coordinates and those created with synthetic semi-random assignment. Actual residences are more likely to be clustered together in locations with a higher density of roadways, whereas synthetic semi-random assignment will provide more scattered residences. Therefore, the synthetically-based accessibility measures may have a relatively higher random error, reflected in higher dispersion, i.e., greater variance or standard deviation. However, the extent of systematic errors is unclear. Systematic errors can be observed if the means of the calculated synthetic accessibility variables are consistently above or below the means based on actual residential locations.

Both TAZ and census block are used as the base to assign synthetic residence. The reason to use the census block is that, if the TAZ level assignment cannot obtain equivalent synthetic residences as actual one, further assignment would be conducted to test whether using smaller geographic units can reduce the error in the accessibility measure and obtain a more realistic distribution of roadway length in buffers. If the TAZ level assignment can obtain reasonable synthetic residences, using a smaller level to assign is not needed due to substantial heavier calculation burden.

The households in both databases are firstly aggregated to the zone (TAZ or census block) level, the most commonly used geographic unit in transportation analysis. Then the total number of residences in each zone can be obtained. Two assignment methods (block centroid and synthetic semi-random) are applied with the condition that the total number of randomly assigned residences for each zone equals the aggregated number of sampled residences in that zone. The synthetic assignment is constrained to residences on local or arterial roadways, avoiding freeways or ramps as physical locations of residences. Next, buffers of various sizes (0.25, 0.5, and 0.75 miles) are created around these the centroid and synthetically assigned residences. Accessibility measures of roadway length and transit stops within each buffer are calculated. The analysis is repeated using the finer level of census block instead of TAZ. Then the calculated average roadway length and transit stops for each TAZ and different buffer sizes are compared using statistical tests. Specifically, the roadway miles in buffers are analyzed for the exact residential locations, block centroid locations, and synthetic roadway-based randomly assigned residential locations. While the synthetically assigned locations/addresses will not, in all likelihood, be the true addresses, comparisons with the true addresses will allow for determining the extent of the errors in accessibility variables. Lastly, inferences are drawn about the geo- imputation.

4.2.2 Synthetic Assignment

A sample assignment screenshot is provided in Figure 11.

Figure 11. Assigning residences to each TAZ along roadways

Synthetic residences are created in relevant TAZs using ArcGIS. An ArcToolBox called "Create Random Points" is used to randomly place a specified number of points within an identified area. If Ni households are located in TAZ i based on the travel survey, then Ni synthetic random points are created within TAZ i. A previous study (Khattak and Wang, 2011) has shown that synthetic semi-random assignment, which adds a constraint that all residences must be located along roadways, is a more realistic method than using completely random assignment of residences. Synthetic semi-random assignment is the preferred method to randomly assign residence in TAZs. Therefore, the roadway network is used as the constraining feature. All the residences are assigned along roadways, and with a further constraint that they cannot be located on freeways, bridges and ramps.

4.2.3 Creating Zonal Centroids to Represent Residences

Instead of semi-randomly assigning the residences to certain roadways, researchers usually use the centroid of a zone (the geometric center) to represent residential locations (Sultana, 2002, Jones et al., 2010). Using a centroid is a convenient way of obtaining synthetic locations of residences. However, the location of a centroid in a zone is not necessarily a reasonable place to locate residences, e.g., residences can end up in unusable land, away from roadways. In this study, the census blocks which have at least one household sample are selected for comparison of actual residences, census block centroid assignment and synthetic assignment. Buffers of various sizes are created around census block centroids to calculate accessibility measures.

4.2.4 Buffer Sizes

Different buffer sizes are used to test their associations with calculated roadway length. Three different sizes (shown in Figure 12) are used to create circular buffers around each household's residential location in the sample. A residence is highlighted, showing the roadway surrounding the residence and two transit stop locations in the circular buffer. The buffer sizes are respectively 0.25, 0.5, and 0.75 miles. Using a walking speed of 3 mph, 0.25 miles will be a 5-minute walk and 0.75 miles will be a 15-minute walk. These buffer sizes are selected based on the fact that micro-factors in the neighborhood are well captured within these thresholds. Evidence in the literature indicates that travelers typically are willing to walk to public transit a distance of 0.25 to 0.75 miles (O'Sullivan and Morrall, 1996). Furthermore, built environment research also uses these buffer sizes to analyze walkability or transit access around residences (Brownson et al., 2009, Kligerman et al., 2007).

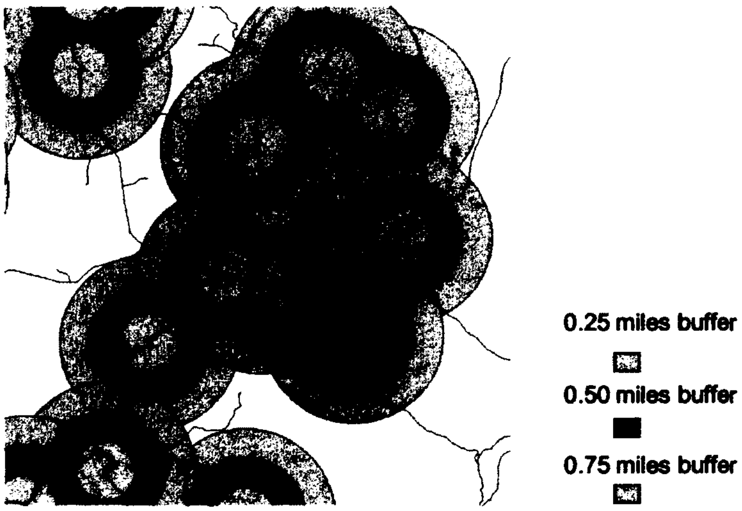


Figure 12. Three buffer sizes used for each residence in the dataset

4.2.5 Comparison of Accessibility Measurements

<u>*Comparing average roadway length within buffers*</u>

Roadway length in a buffer is a measure of network connectivity and network accessibility. Student t-statistics were used to test the statistically significance of difference (5% level) between accessibility measures in buffers when using synthetic assignment and actual residential locations. Only the means of samples in the geographic unit are compared. This is because we cannot directly compare actual residence with a particular synthetic residence since there are multiple synthetic residences in most of the zones. That is, one-to-one equivalence (matching) of a particular synthetic residence with an actual residence is not possible within the scope of this study.

<u>*Comparing roadway length within buffers by zone*</u>

To compare roadway length within buffers by zone, average roadway length is calculated for each zone. Figure 13 demonstrates the calculation of average roadway length in three hypothetical zones.

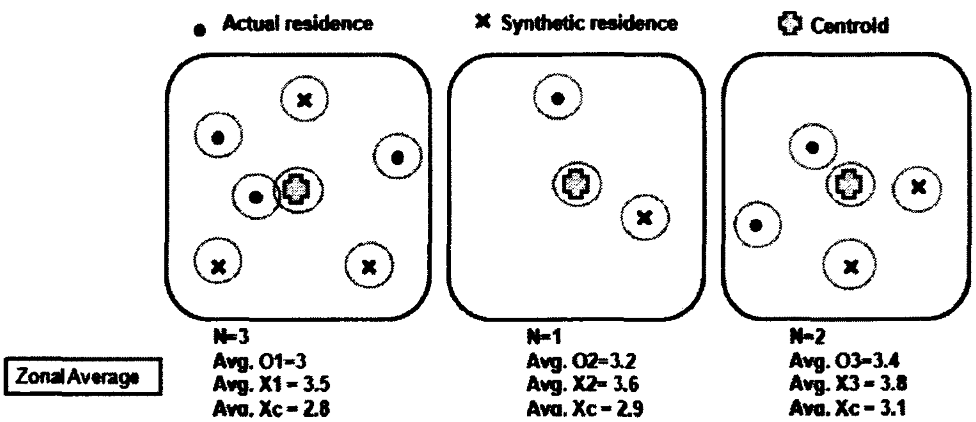


Figure 13. Demonstration of calculating zonal average

Each zone has three different average roadway lengths in buffers that are associated with the actual (observed) residence, and the synthetic residence if all the sampled residences are coded at the zone centroid. A sample calculation of the mean of roadway length for the actual residence (O_i), synthetic residence (X_i), and the centroid (X_{ci}) is shown.

After obtaining the means for the zones, several goodness-of-fit measures are calculated to compare the actual values with synthetic values. These included:

Percent root mean squared error (PRMSE) is a widely used measure of accuracy to gauge the differences between values predicted by a model or an estimator and the values actually observed in the field.

$$\text{PRMSE}= \frac{\left(\sum_j (AvgX_j - AvgO_j)^2 / (N-1)\right)^{0.5} * 100}{\left(\sum_j AvgO_j / N\right)} \qquad \text{(Equation 8)}$$

Where:

$AvgX_j$ is the mean of accessibility variables in buffers around synthetic residences in zone j;

$AvgO_j$ is the mean of accessibility variables in buffers around actual residences in zone j;

N is the number of zones.

Mean absolute percentage error (MAPE) is the absolute difference between observed and synthetic value, divided by the actual value. This measure is used to overcome the disadvantage of PRMSE heavily weighting extreme values by squaring them.

$$\text{MAPE} = \frac{\left(\sum_j (| AvgX_j - AvgO_j | / AvgO_j)\right) * 100}{N} \qquad \text{(Equation 9)}$$

4.3 Comparison Results

4.3.1 Roadway Length Comparison on Different Geographical Unit

The descriptive statistics of roadway length in buffers around synthetic residences (TAZ level and block level), block centroid and actual residences are listed in Table 1.

Table 1. Descriptive statistics for roadway length at different levels

a) for the Research Triangle, NC

Buffers	Assignment		Mean	Min	Max	SD	Variance	%SD/ Mean	Variance/ Mean	% difference of means
0.25 Miles	TAZ level Synth. Assgmt.	1st	[redacted]	0.26	7.02	1.05	1.10	51	0.53	-4.63%
		2nd		0.26	7.84	1.05	1.09	51	0.53	-5.09%
		3rd		0.27	8.27	1.06	1.12	51	0.54	-4.17%
		4th		0.26	8.09	1.06	1.12	52	0.55	-5.09%
	Block Centroid			0.00	8.16	1.26	1.58	67	0.83	-11.57%
	Block level Synth. Assgmt.	1st		0.25	8.24	1.04	1.07	50	0.51	-1.41%
		2nd		0.25	8.20	1.03	1.07	49	0.51	-1.41%
		3rd	2.11	0.27	8.18	1.04	1.07	49	0.51	-0.94%
		4th	2.12	0.25	8.27	1.04	1.08	49	0.51	-0.47%
	Actual Res.		2.16	0.00	8.20	1.04	1.08	48	0.50	Base
0.50 Miles	TAZ level Synth. Assgmt.	1st	[redacted]	0.54	25.60	3.97	15.73	55	2.18	-3.73%
		2nd		0.72	26.21	3.99	15.89	55	2.21	-4.26%
		3rd		0.54	25.91	3.98	15.85	55	2.20	-3.86%
		4th		0.55	26.25	3.97	15.75	55	2.18	-3.73%
	Block Centroid			0.02	26.84	4.19	17.57	62	2.48	-5.73%
	Block level Synth. Assgmt.	1st	7.32	0.51	26.93	3.90	15.24	53	2.08	-2.53%
		2nd	7.34	0.54	26.69	3.93	15.46	54	2.11	-2.26%
		3rd	7.34	0.65	26.65	3.91	15.31	53	2.09	-2.26%
		4th	7.36	0.59	26.80	3.93	15.41	53	2.09	-2.00%
	Actual Res.		7.51	0.51	26.67	3.96	15.71	53	2.09	Base
0.75 Miles	TAZ level Synth. Assgmt.	1st	15.23	1.19	45.98	8.46	71.60	56	4.70	-1.93%
		2nd	15.19	1.06	45.57	8.49	72.11	56	4.75	-2.19%
		3rd	15.24	1.08	46.40	8.47	71.70	56	4.70	-1.87%
		4th	15.26	1.04	44.27	8.44	71.24	55	4.67	-1.74%
	Block Centroid		15.11	0.04	45.77	8.58	73.69	59	4.88	-2.70%
	Actual Res.		15.53	1.48	45.19	8.32	69.28	54	4.46	Base

b) For Charlotte, NC

Buffers	Assignment		Mean	Min	Max	SD	Varianc e	%SD/ Mean	Variance/ Mean	% difference of means
0.25 Miles	TAZ level Synth. Assgmt.	1st	[redacted]	0.26	5.98	0.96	0.92	46.83	0.45	-3.76%
		2nd	[redacted]	0.28	5.96	0.96	0.92	47.06	0.45	-4.23%
		3rd	[redacted]	0.15	6.41	0.96	0.93	46.83	0.45	-3.76%
		4th	[redacted]	0.26	6.53	0.96	0.93	46.60	0.45	-3.29%
	Block Centroid		[redacted]	0.00	6.20	1.11	1.29	59.36	0.69	-12.21%
	Block level Synth. Assgmt.	1st	2.09	0.29	6.24	0.95	0.91	45.45	0.44	-1.88%
		2nd	2.08	0.30	6.14	0.95	0.90	45.67	0.43	-2.35%
		3rd	2.10	0.29	6.26	0.94	0.88	44.76	0.42	-1.41%
		4th	2.08	0.26	6.02	0.94	0.89	45.19	0.43	-2.35%
	Actual Res.		2.13	0.01	6.25	0.96	0.93	45.07	0.44	Base
0.50 Miles	TAZ level Synth. Assgmt.	1st	[redacted]	0.83	20.37	3.55	12.61	49.58	1.76	-3.89%
		2nd	[redacted]	0.81	20.09	3.53	12.48	49.16	1.74	-3.62%
		3rd	[redacted]	0.83	21.48	3.57	12.74	49.58	1.77	-3.36%
		4th	7.21	1.00	21.13	3.56	12.64	49.38	1.75	-3.22%
	Block Centroid		[redacted]	0.10	21.42	3.79	13.53	53.46	1.91	-4.83%
	Block level Synth. Assgmt.	1st	7.31	0.76	21.42	3.52	12.38	48.15	1.69	-1.88%
		2nd	7.28	0.86	21.59	3.51	12.30	48.21	1.69	-2.28%
		3rd	7.29	0.77	21.27	3.50	12.25	48.01	1.68	-2.15%
		4th	7.27	0.68	21.51	3.52	12.39	48.42	1.70	-2.42%
	Actual Res.		7.45	0.54	21.38	3.59	12.89	48.19	1.73	Base
0.75 Miles	TAZ level Synth. Assgmt.	1st	15.22	1.54	38.27	7.51	56.38	49.34	3.70	-3.24%
		2nd	15.22	1.76	37.71	7.49	56.05	49.21	3.68	-3.24%
		3rd	15.21	1.84	38.27	7.53	56.67	49.51	3.73	-3.31%
		4th	15.25	1.57	39.03	7.50	56.32	49.18	3.69	-3.05%
	Block Centroid		15.29	1.27	38.88	7.74	56.16	50.62	3.67	-2.80%
	Actual Res.		15.73	1.41	41.41	7.69	59.05	48.89	3.75	Base

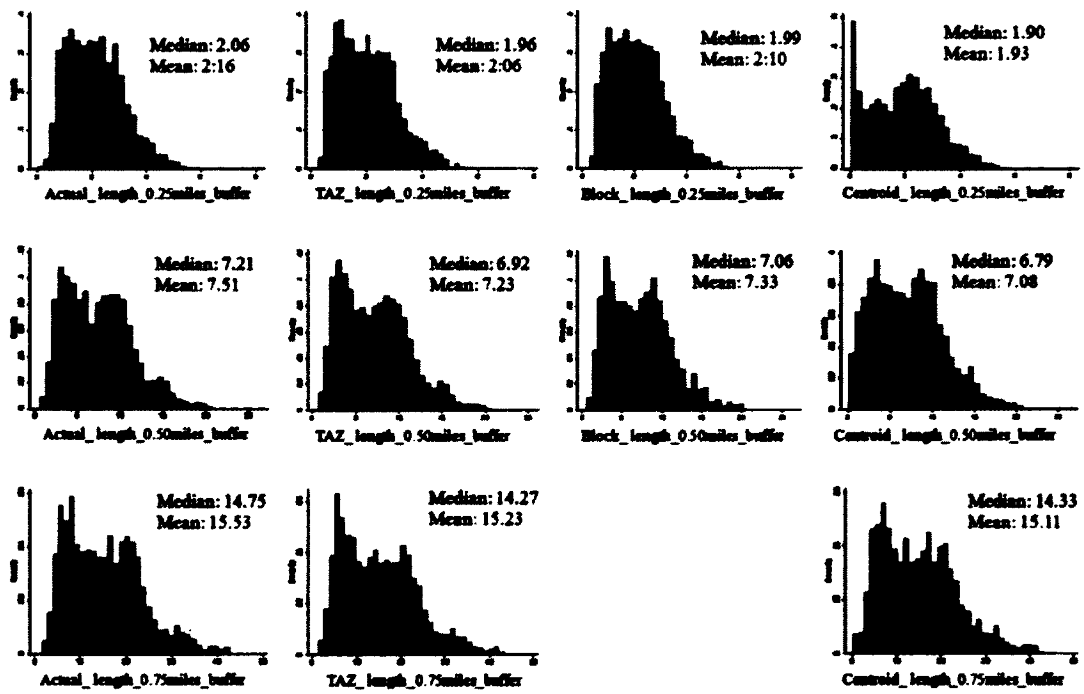
Note: Centroid values are the descriptive statistics weighted by number of samples in each block

Grey cell indicates the sample mean is statistical significantly different (5% level) from the actual mean

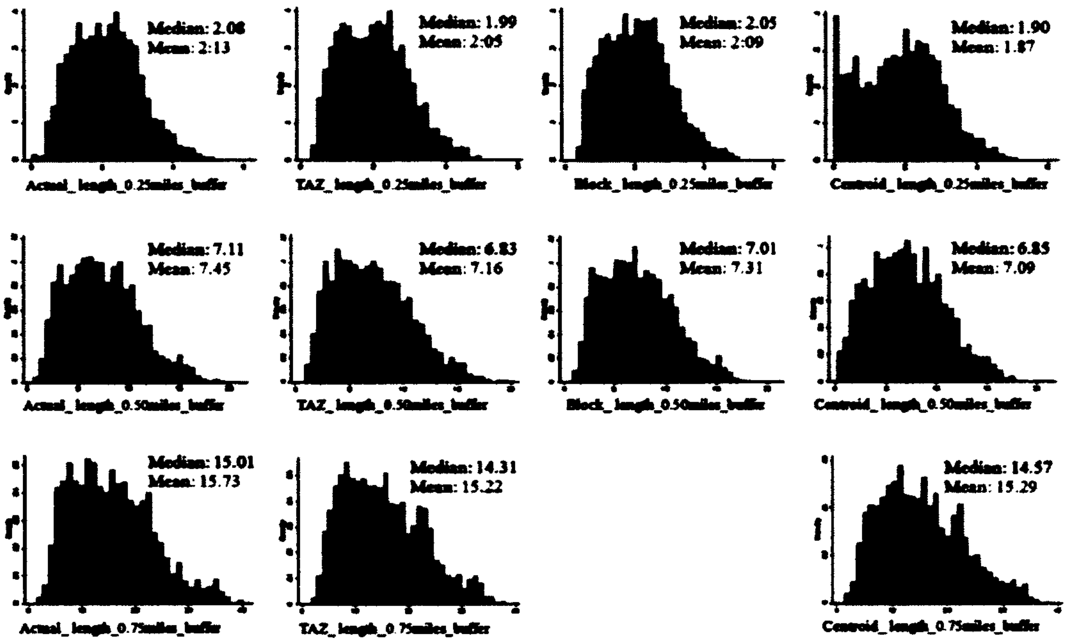
The four synthetic semi-random assignments yielded similar results. This indicates that synthetic assignments are reasonably stable. Furthermore, synthetic semi-random assignments have consistently smaller means for roadway length in buffers compared with the actual residences. This is partly because in synthetic assignments residences are not as closely clustered as in real-life. Hence the synthetically assigned residences have a greater chance of being in areas with lower density of roadways. The differences between random assignment and actual residences in terms of roadway length in buffers are statistically significant (5% level) for smaller buffer sizes (0.25 and 0.50 miles). Furthermore, the difference in mean, which measures the extent of errors in the variable, is larger for smaller buffer sizes (5% for 0.25 miles buffer but about 2% for 0.75 miles buffer). However, systematic errors may be limited as reflected in the means of synthetic assignment being lower and higher for 0.25 and 0.5 mile buffers respectively compared with the actual residences. There is empirical evidence for random errors reflected in larger standard deviations of the synthetic assignments compared with actual residences, especially for TAZ level assignment. However, the random errors are alleviated when random assignment is based on the census block level.

Standard Deviations (SD) capture variations in the distribution of roadway accessibility. They are quite close for the four synthetic semi-randomly assigned residences and the actual residential locations. However, the block centroid shows larger standard variations than actual and synthetically assigned residences. To compare buffers with different means, the Coefficient of Variation (CV) is provided, which is a normalized measure of dispersion (also known as Relative Standard Deviation or RSD). The percentages of CV are below 100% indicating relatively low variance distributions and under-dispersion. Furthermore, for the same buffer sizes, synthetic semi-random assignment has slightly higher percent CV compared with actual residential locations. This implies that households tend to scatter more in space than in reality when assigned semi-randomly. Alternatively, clustering of households in real-life may reduce the variability in the accessibility measure. Also, this may be due in part to the TAZ being a relatively large geographic unit, which gives higher levels of spatial freedom in semi-randomly assigning residences. Variance divided by mean captures the dispersion level in the distribution of roadway accessibility. If the sample variance is greater than the sample mean, the data shows over-dispersion; otherwise if the sample variance is smaller than the sample mean, it shows under-dispersion. The measurement of dispersion indicates that the roadway length is under-dispersion for small buffer size (0.25 miles) but it is over-dispersion for larger buffer sizes (0.50miles and 0.75 miles).

The histogram distributions of roadway length in buffers around actual residences and synthetically assigned TAZ level, Census block level and centroid level residences are presented in Figure 14.



a) For the Research Triangle, NC



b) For Charlotte, NC

Figure 14. Distribution of roadway length variable for different level

Note: The centroid-based distribution is weighted by number of samples in each block

The horizontal axis represents the roadway length within the buffers, and the vertical axis represents the density.

The distribution of roadway length in the above figure shows a bimodal distribution of roadway lengths for actual residences with a positive (right) skew. However, the distribution does not seem to be normal or lognormal. The differences between the mean and median are substantial for the centroid assigned residences and actual residences. This is partly because the block centroid may be located in unusable land. This may be the major reason for relatively large differences in block centroid and actual residences.

If block centroids are used as residential locations, as is the case in current practice, then the value of roadway accessibility measure is substantially less compared with the actual locations of residences. Notably, the minimal roadway length for the block centroid is less than the buffer size (it can be zero for 0.25 miles buffer shown in Figure 14), so some of them may be located in unusable land. This may be the major cause of the large difference between the actual residences and block centroids. Also, the differences between block centroid and actual residences are statistically significant for all buffer sizes. Furthermore, block centroid assignments have substantially lower roadway lengths in smaller buffers (about 16% lower for 0.25 miles buffer and 9% lower for 0.5 miles buffer). This indicates that using census block centroids to represent the locations of residences can cause substantial errors. Table 2 presents the comparison of PRMSE and MAPE between assignments on TAZ level and census block level.

Table 2. Error for roadway length in buffers around different levels

a) For the Research Triangle, NC

Buffers	Assignment	TAZ MAPE	TAZ PRMSE	Block MAPE	Block PRMSE
0.25 Miles	1st Synth. Assgmt.	22.15%	23.34%	19.62%	19.95%
	2nd Synth. Assgmt.	21.60%	22.94%	19.71%	20.88%
	3rd Synth. Assgmt.	21.31%	22.16%	19.48%	19.82%
	4th Synth. Assgmt	21.53%	23.33%	19.97%	20.31%
	Block Centroid	29.51%	32.23%	29.51%	32.23%
0.50 Miles	1st Synth. Assgmt.	16.06%	18.08%	13.10%	13.54%
	2nd Synth. Assgmt.	15.37%	17.56%	13.20%	14.06%
	3rd Synth. Assgmt.	15.41%	17.54%	12.79%	13.32%
	4th Synth. Assgmt	15.75%	17.55%	13.08%	13.48%
	Block Centroid	17.34%	24.24%	17.34%	24.24%
0.75 Miles	1st Synth. Assgmt.	12.57%	14.03%	-	-
	2nd Synth. Assgmt.	12.35%	14.09%	-	-
	3rd Synth. Assgmt.	12.31%	13.66%	-	-
	4th Synth. Assgmt	12.34%	13.90%	-	-
	Block Centroid	11.73%	21.67%	11.73%	21.67%

b) For Charlotte, NC

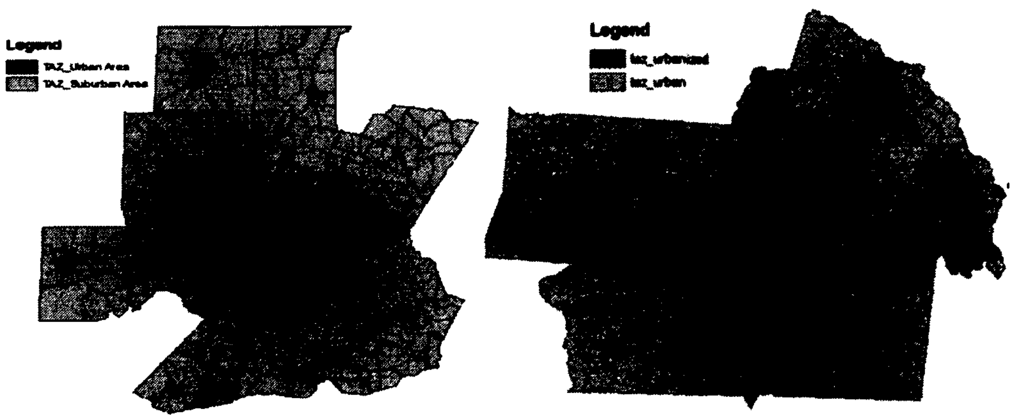
Buffers	Assignment	TAZ MAPE	TAZ PRMSE	Block MAPE	Block PRMSE
0.25 Miles	1st Synth. Assgmt.	22.15%	23.34%	19.62%	19.95%
	2nd Synth. Assgmt.	21.60%	22.94%	19.71%	20.88%
	3rd Synth. Assgmt.	21.31%	22.16%	19.48%	19.82%
	4th Synth. Assgmt	21.53%	23.33%	19.97%	20.31%
	Block Centroid	29.51%	32.23%	29.51%	32.23%
0.50 Miles	1st Synth. Assgmt.	16.06%	18.08%	13.10%	13.54%
	2nd Synth. Assgmt.	15.37%	17.56%	13.20%	14.06%
	3rd Synth. Assgmt.	15.41%	17.54%	12.79%	13.32%
	4th Synth. Assgmt	15.75%	17.55%	13.08%	13.48%
	Block Centroid	17.34%	24.24%	17.34%	24.24%
0.75 Miles	1st Synth. Assgmt.	12.57%	14.03%	-	-
	2nd Synth. Assgmt.	12.35%	14.09%	-	-
	3rd Synth. Assgmt.	12.31%	13.66%	-	-
	4th Synth. Assgmt	12.34%	13.90%	-	-
	Block Centroid	11.73%	21.67%	11.73%	21.67%

A traffic analysis zone or census block can contain more than one household sample. For zones with more than one sample, average roadway length for each TAZ/block is calculated. Averages by zones are then used to calculate the PRMSE and MAPE. The extent of errors from synthetic and block centroid assignments relative to actual residential locations indicates a lower systematic error with larger buffer sizes. The PRMSE and MAPE measures show that the errors for smaller buffer sizes (0.25 miles) are relatively large compared to larger buffers. Note that a value close to zero for these two measures means concordance between the actual and synthetically created accessibility variable. When the buffer size increases, the chance of overlap between actual and synthetic residences also increases, resulting in lower systematic error. These measures can be used for comparative purposes, though a PRMSE above 30% is often considered high for transportation applications, which is beyond the range of acceptable accuracy.

The results indicate that using smaller geographic units to semi-randomly assign residences can increase concordance between actual and synthetic assignments. Specifically, the difference of roadway length in 0.50 miles buffer between random assignment and actual residences is not statistically significant. Also, both the PRMSE and MAPE are relatively lower. However, since the computational burden increases substantially with a census block level synthetic assignment (using 3257 blocks instead of 1400 TAZs); the improvement in concordance is rather marginal.

4.3.2 Roadway Length Comparison on Urban vs. Suburban Area

To explore positional differences between synthetic assignments and the actual residences, the TAZs are grouped into urban and suburban/rural classifications by using the Census 2000 Urbanized Areas boundary files. Figure 15 provides the urban vs. suburban/rural TAZ boundaries used in this analysis.



a) in the Research Triangle b) in Charlotte Area

Figure 15. Urban vs. Suburban TAZ boundary

The descriptive statistics for roadway length in buffers at the TAZ level by area type are provided in Table 3.

Table 3. Descriptive statistics for roadway length at TAZ level for suburban/rural areas

a) For the Research Triangle (unit=miles, N=3012 for urban, N=1712 for suburban)

Buffer		Assignment	Mean	Min.	Max.	S.D.	%SD/Mean	% difference of mean
0.25 Miles	Urban	1st Synth. Assgmt.	[redacted]	0.28	7.02	0.94	37.15%	-2.69%
		2nd Synth. Assgmt.	[redacted]	0.40	7.84	0.94	37.45%	-3.46%
		3rd Synth. Assgmt.	[redacted]	0.32	8.27	0.93	36.76%	-2.69%
		4th Synth. Assgmt	[redacted]	0.31	8.09	0.95	37.70%	-3.08%
		Block Centroid	[redacted]	0	8.16	1.05	42.68%	-5.38%
		Actual residence	2.60	0.00	8.20	0.93	35.77%	Base
	suburban	1st Synth. Assgmt.	[redacted]	0.26	4.57	0.64	52.03%	-10.87%
		2nd Synth. Assgmt.	[redacted]	0.26	5.09	0.65	52.42%	-10.14%
		3rd Synth. Assgmt.	[redacted]	0.27	5.27	0.68	55.28%	-10.87%
		4th Synth. Assgmt	[redacted]	0.26	2.16	0.66	54.10%	-11.59%
		Block Centroid	[redacted]	0.00	5.77	0.91	101.11%	-34.78%
		Actual residence	1.38	0.00	5.60	0.71	51.45%	Base
0.50 Miles	urban	1st Synth. Assgmt.	9.20	1.63	25.60	3.42	37.17%	-2.23%
		2nd Synth. Assgmt.	9.17	1.59	26.21	3.46	37.81%	-2.76%
		3rd Synth. Assgmt.	9.18	1.39	25.91	3.44	37.47%	-2.44%
		4th Synth. Assgmt	9.19	1.52	26.25	3.42	37.21%	-2.34%
		Block Centroid	9.22	0.36	26.84	3.50	37.96%	-2.02%
		Actual residence	9.41	1.41	26.67	3.47	36.88%	Base
	suburban	1st Synth. Assgmt.	[redacted]	0.54	15.85	1.95	52.00%	-10.71%
		2nd Synth. Assgmt.	[redacted]	0.72	15.88	2.02	54.01%	-10.95%
		3rd Synth. Assgmt.	[redacted]	0.54	16.04	2.03	54.13%	-10.71%
		4th Synth. Assgmt	[redacted]	0.55	14.87	2.01	53.46%	-10.48%
		Block Centroid	[redacted]	0.00	16.77	2.42	69.14%	-16.67%
		Actual residence	4.20	0.51	17.49	2.19	52.14%	Base
0.75 Miles	urban	1st Synth. Assgmt.	19.57	3.61	45.98	7.26	37.10%	-0.56%
		2nd Synth. Assgmt.	19.53	3.21	45.57	7.28	37.28%	-0.76%
		3rd Synth. Assgmt.	19.58	3.92	46.40	7.23	36.93%	-0.51%
		4th Synth. Assgmt	19.59	4.07	44.27	7.22	36.86%	-0.46%
		Block Centroid	19.73	4.62	45.77	7.21	36.54%	0.25%
		Actual residence	19.68	4.86	45.19	7.23	36.74%	Base
	suburban	1st Synth. Assgmt.	[redacted]	1.19	27.85	3.61	47.63%	-8.45%
		2nd Synth. Assgmt.	[redacted]	1.06	32.89	3.73	49.54%	-9.06%
		3rd Synth. Assgmt.	[redacted]	1.08	24.17	3.73	49.21%	-8.45%
		4th Synth. Assgmt	[redacted]	1.04	27.42	3.71	48.62%	-7.85%
		Block Centroid	[redacted]	0.00	35.29	4.12	53.93%	-7.73%
		Actual residence	8.28	1.48	36.32	4.03	48.67%	Base

b) For Charlotte (unit=miles, N=2672 for urban, N=641 for suburban)

Buffer		Assignment	Mean	Min.	Max.	S.D.	%SD/Mean	% difference of mean
0.25 Miles	Urban	1st Synth. Assgmt.	[redacted]	0.29	5.98	0.91	40.09%	-3.40%
		2nd Synth. Assgmt.	[redacted]	0.28	5.96	0.90	39.65%	-3.40%
		3rd Synth. Assgmt.	[redacted]	0.31	6.41	0.90	39.47%	-2.98%
		4th Synth. Assgmt	[redacted]	0.26	6.53	0.90	39.47%	-2.98%
		Block Centroid	[redacted]	0.00	6.20	1.05	49.30%	-9.36%
		Actual residence	2.35	0.02	6.25	0.90	38.30%	Base
	suburban	1st Synth. Assgmt.	1.14	0.26	5.51	0.59	51.75%	-2.56%
		2nd Synth. Assgmt.	1.09	0.27	4.17	0.52	47.71%	-6.84%
		3rd Synth. Assgmt.	1.11	0.15	5.18	0.54	48.65%	-5.13%
		4th Synth. Assgmt	1.14	0.30	5.59	0.60	52.63%	-2.56%
		Block Centroid	[redacted]	0.00	4.31	0.69	94.52%	-37.61%
		Actual residence	1.17	0.01	4.25	0.53	45.30%	Base
0.50 Miles	urban	1st Synth. Assgmt.	[redacted]	0.99	20.37	3.29	40.77%	-3.47%
		2nd Synth. Assgmt.	[redacted]	1.00	20.09	3.25	40.17%	-3.23%
		3rd Synth. Assgmt.	8.13	0.93	21.48	3.29	40.47%	-2.75%
		4th Synth. Assgmt	8.11	1.01	21.13	3.28	40.44%	-2.99%
		Block Centroid	[redacted]	0.21	21.42	2.01	25.00%	-3.83%
		Actual residence	8.36	0.54	21.38	3.31	39.59%	Base
	suburban	1st Synth. Assgmt.	3.40	0.83	12.42	1.56	45.88%	-3.68%
		2nd Synth. Assgmt.	3.38	0.81	12.19	1.54	45.56%	-4.25%
		3rd Synth. Assgmt.	3.38	0.83	12.60	1.53	45.27%	-4.25%
		4th Synth. Assgmt	3.43	1.00	13.31	1.65	48.10%	-2.83%
		Block Centroid	[redacted]	0.10	11.38	1.73	57.67%	-15.01%
		Actual residence	3.53	0.68	11.43	1.49	42.21%	Base
0.75 Miles	urban	1st Synth. Assgmt.	17.24	1.80	38.27	6.89	39.97%	-2.87%
		2nd Synth. Assgmt.	17.24	1.76	37.71	6.85	39.73%	-2.87%
		3rd Synth. Assgmt.	17.24	1.97	38.27	6.90	40.02%	-2.87%
		4th Synth. Assgmt	17.25	1.87	39.03	6.88	39.88%	-2.82%
		Block Centroid	17.29	2.01	38.88	6.81	39.39%	-2.59%
		Actual residence	17.75	3.28	41.41	6.86	100.00%	Base
	suburban	1st Synth. Assgmt.	6.90	1.54	19.47	2.64	38.26%	-2.27%
		2nd Synth. Assgmt.	6.85	1.96	19.97	2.66	38.83%	-2.97%
		3rd Synth. Assgmt.	6.87	1.84	20.39	2.67	38.86%	-2.69%
		4th Synth. Assgmt	6.89	1.57	19.75	2.69	39.04%	-2.41%
		Block Centroid	6.69	1.27	20.24	2.69	40.21%	-5.24%
		Actual residence	7.06	1.41	20.16	2.62	37.11%	Base

Note: Centroid values are the descriptive statistics weighted by number of samples in each block

Grey cell indicates the sample mean is statistical significantly different (5% level) from the actual mean

The comparison between urban and suburban TAZ based synthetic results show that the differences between synthetic assignments and actual residences are concentrated in suburban/rural areas. Students' t-tests indicate that in suburban areas, statistically significant differences exist (5% level) between synthetic assignments and actual residences in terms of roadway length in all buffers. However, for the urban area, the difference is only significant for a smaller buffer size of 0.25 miles. Furthermore, roadway length in buffers is systematically lower than actual for the suburban/rural area. Also, roadway length in the smaller 0.25 miles buffer for urban area is two times that for suburban/rural areas. Also, the dispersion (standard deviation) in the urban area is higher than that of suburban/rural area. Suburban/rural areas have higher percent CV.

Again, no statistically significant difference is found for larger buffer sizes (0.75 miles) in the urban area of Charlotte. However, significant differences are found for smaller buffer sizes (0.25 miles). Unlike the Research Triangle Area, no significant difference is found in suburban/rural areas of Charlotte. More specifically, roadway accessibility in buffers using synthetic assignments is slightly lower than actual residences in urban areas of both the Research Triangle and the Charlotte region. However, the percentages are rather different for suburban/rural areas for these two regions (10% in suburban/rural area of Research Triangle, and 2% to 6% in suburban/rural areas of Charlotte). Overall, synthetic assignments in Charlotte gave relatively better concordance with actual residences in terms of the accessibility measure compared with the Research Triangle region, especially in suburban/rural areas.

4.3.3 Transit Stops in Buffers Comparison

Due to limited availability of public transportation data, only Durham, Chapel Hill, and Raleigh from the Research Triangle are included in the analysis of transit accessibility measures. The distribution of transit stops in the area is shown in Figure 16.

Figure 16. Transit stops in the Research Triangle Area of North Carolina

Transit stops in buffers of different sizes are counted in the Research Triangle Area. Relevant descriptive statistics are shown in Table 4. Transit accessibility in residential buffers in terms of PRMSE is shown in Table 5. Note that most of the TAZs with one or more sampled residences had no transit stops. Such TAZs are excluded from further analysis transit accessibility.

Table 4. Descriptive statistics for number of transit stops

TAZ level in Research Triangle, NC (N=1863)

Buffer	Assignment	Mean	Min.	Max.	S.D.	SD/Mean %	% difference of mean
0.25 Miles	1st Synth. Assgmt.	4.07	0	32	4.40	108.11%	-0.73%
	2nd Synth. Assgmt.	4.11	0	34	4.40	107.06%	0.24%
	3rd Synth. Assgmt.	4.10	0	34	4.51	110.00%	0.00%
	4th Synth. Assgmt	4.07	0	34	4.37	107.37%	-0.73%
	Actual Res.	4.10	0	25	4.41	107.56%	Base
0.50 Miles	1st Synth. Assgmt.	14.92	0	68	13.58	91.02%	0.61%
	2nd Synth. Assgmt.	15.13	0	71	13.64	90.15%	2.02%
	3rd Synth. Assgmt.	15.00	0	73	13.64	90.93%	1.15%
	4th Synth. Assgmt	15.04	0	72	13.60	90.43%	1.42%
	Actual Res.	14.83	0	67	13.48	90.90%	Base
0.75 Miles	1st Synth. Assgmt.	32.16	0	127	27.39	85.17%	0.78%
	2nd Synth. Assgmt.	32.12	0	127	27.09	84.34%	0.66%
	3rd Synth. Assgmt.	32.15	0	127	27.03	84.07%	0.75%
	4th Synth. Assgmt	32.09	0	128	27.14	84.57%	0.56%
	Actual Res.	31.91	0	126	27.34	85.68%	Base

Table 5. Errors in number of transit stops in Research Triangle Area

Buffer	Assignment	TAZ PRMSE (N=1863)	Block PRMSE (N=1160)
0.25 Miles	1st Synth. Assgmt.	44.56%	44.05%
	2nd Synth. Assgmt.	49.74%	46.26%
	3rd Synth. Assgmt.	46.70%	44.39%
	4th Synth. Assgmt	51.11%	43.22%
	Actual Res.	Base	Base
0.50 Miles	1st Synth. Assgmt.	25.30%	NA
	2nd Synth. Assgmt.	25.83%	
	3rd Synth. Assgmt.	24.47%	
	4th Synth. Assgmt	26.61%	
	Actual Res.	Base	Base
0.75 Miles	1st Synth. Assgmt.	17.29%	NA
	2nd Synth. Assgmt.	17.06%	
	3rd Synth. Assgmt.	17.34%	
	4th Synth. Assgmt	17.61%	
	Actual Res.	Base	Base

Transit accessibility did not differ statistically significantly (5% level) between buffers around synthetically assigned residences and actual residences. The number of transit stops for synthetic residences are slightly higher than that of actual residences. Due to their high dispersion, the percentages of CV are relatively high (more than 100% for the smaller buffer size). For actual residences, the CV is also relatively high, implying over-dispersion in transit accessibility. Furthermore, using a smaller spatial unit to assign synthetic residences did not improve the results, as indicated by PRMSE for TAZ level assignment versus a block level assignment. Nonetheless, synthetic assignments using larger buffer sizes (0.50 and 0.75 miles) can give greater concordance, e.g. PRMSE dropped by nearly one-half going from 0.25 miles buffer to 0.5 miles buffer.

4.4 Summary of Findings

This chapter determines the relative level of accuracy when exact geo-coordinate information is not available. Geo-coordinate imputation can potentially overcome the problem of not knowing the exact geo-coordinate location of a residence. It can synthetically assign a household to an exact location (lat-long). The question then is whether such assignment gives reasonably accurate results. This study explores if synthetic assignments of geo-coordinates are concordant with actual residential locations; and if so, what the implications for travel demand modeling are. Using behavioral travel surveys in North Carolina, 4,724 households in Research Triangle, NC and 3,310 households in Charlotte, NC are extracted. These households are semi-randomly assigned to synthetic point locations within the TAZ or census block of the sampled residence. Three buffer sizes (0.25, 0.50 and 0.75 miles) are used to create boundaries around each residential location, and the roadway length within these buffers is calculated using relatively detailed VDOT roadway centerline data available from ESRI. This is a key characteristic used to measure the built environment around a residence, e.g., a denser roadway network in a buffer implies greater connectivity and accessibility. The roadway length is calculated at both TAZ and census tract levels. Comparison between different assignments, different buffer sizes, and different geographic units are conducted. If the synthetic semi-random assignments can create equivalent residences that do not give statistically significant differences in accessibility measures from real residences, then the randomly assigned lat-long can be used to approximate residential locations and create new variables. By doing so, the confidentiality issue can be overcome to some degree.

Although geo-imputation presented in this report does not produce the same effect on capturing roadway length in buffers around residences for different regions, (i.e., it works better in Charlotte than in the Research Triangle area), the comparison results can be summarized as follows:

- Using TAZ information to synthetically assign residences is not equivalent to having actual household locations when analyzing roadway accessibility measures, i.e., roadway length within 0.25 mile buffers around residences. However, for larger buffer sizes (0.75 miles), the roadway accessibility measures created using synthetic assignment and actual residents are not statistically different (5% level), indicating good concordance with actual residences. For a 0.50 miles buffer, the results are mixed and depended on the study region.

- Assigning residences semi-randomly based on census blocks can provide greater concordance than synthetically assigning residences based on larger geographic units, (i.e., using the TAZ level). Lower PRMSE and MAPE are observed for the accessibility measures (roadway length and number of transit stops in buffers around residences) when census block is used for assigning residences. Census block assignment gave reasonable accessibility measures for 0.50 miles and larger buffer sizes.
- The TAZ-based synthetic semi-random assignment gave shorter (2% to 5% less) roadway length in buffers compared to actual residences. The difference of roadway length between synthetic semi-random assignment and actual residence is larger in smaller buffers than in larger buffers, reflecting greater systematic measurement error when smaller geographic scales are used for analysis.
- The measure of dispersion (variance divided by mean) shows that synthetic residences assigned based on the census block level are statistically very close to the dispersion of actual residences. However, synthetic residences based on the TAZ level have greater random error than synthetic residences based on the census block level. Furthermore, roadway length is under-dispersed for the small buffer size (0.25 miles). However, it is over-dispersed for larger buffer sizes (0.50 miles and 0.75 miles).
- The standard deviation and percent Coefficient of Variation shows that residences have greater random error as they are more scattered in the TAZ boundary when synthetically assigned compared with actual residential

locations. Conversely, the actual residences are more closely clustered together in space, due to agglomeration, and therefore have smaller variance.

- The results clearly indicate that using block centroid to represent residences, as is current practice, gives relatively large measurement errors when calculating accessibility measures of roadway length and number of transit stops in buffers.

5. BUILT ENVIRONMENT ASSOCIATIONS WITH TRIP MAKING

By applying the geo-imputation method in the previous chapter, built environment measurements can be calculated based on the exact location of residences. This provides a basis to understand the connection between built environment and travel behavior. To appropriately implement land use policies, a fundamental question is whether there are associations between built environment and trip making and whether the implications from these associations can be generally used within a metropolitan area.

This chapter addresses the association about the built environment and travel behavior from a spatial perspective by emphasizing spatial heterogeneity which possibly exists but has been rarely captured. The hypotheses intended for examination are whether built environments, including land use mixes and roadway density, are associated with different modes of trip to the same extent in study region. If yes, how to capture this potential spatial variation in association and what its implication is. A unique database using behavioral data combined with a variety of spatial data, taking advantage of emerging GIS technologies, is used for this purpose. Network analysis and geographical regression methods are used in this analysis to help answer the above questions.

5.1 Methodology to Capture Built Environment

Buffer analysis is used to capture the built environments around residences. Instead of using buffers with a fixed size, network based buffers (0.25 miles) are created around residences by using GIS network analysis module. A dynamic network based buffer can effectively represent the accessible area within 0.25 miles of the residence. Then various built environment variables within the buffer are measured.

To capture land use mix, public facilities are counted within network buffers created around residences. They included the number of restaurants, shopping stores, banks, bus stops, and churches. To capture land development density around residences, a satellite image product NLCD land Cover 2001 data for Virginia is used. Specifically, the land cover data is a raster database which has attributes of different land cover categories. For instance, four different land cover types based on percent of impervious surfaces of total land cover are identified in the database, including open space, low intensity, medium intensity and high intensity developments. Open spaces are usually parks, golf course and so on; low intensity developments include single-family housing units and areas with a mix of buildings and vegetation with 20%-49% impervious surfaces; medium intensity development category includes single-family housing units and areas with 50%-79% impervious surfaces; high intensity developments are areas where people reside or work in high numbers, and usually include apartment complexes, row houses and commercial/industrial developments with impervious surfaces accounting for 80% to 100% of the total cover. The area of each land cover category is calculated, and then divided by total area of the buffer to produce a percentage for each category, which is a proxy variable of land use density.

5.2 Data Description

The behavioral data are extracted from the Virginia Add-on survey of NHTS (National Household Travel Survey) conducted in 2008 (survey period was from April 2008 through May 2009). The area studied is Hampton Roads, a region located in southeastern Virginia with a population of approximately 1.7 million. The households' descriptive statistics are shown in Table 6 and are selected based on the typical variables used in household trip production models (e.g., as reflected in the present Hampton Roads regional travel demand model or NCHRP 365).

Table 6. Descriptive statistics for Hampton Roads, VA data (N=3,151)

	Variable	Mean	Std. Dev.	Min	Max
Daily trips	NTRIP (frequency of total trips)	7.92	6.22	0	47
Trip mode	WALK/BIKE (walking or biking trips, binary variable)	0.245	0.43	0	1
	AUTOMOBILE (auto trip frequency)	6.957	5.704	0	45
Household characteristics	HHSIZE (household size)	2.39	1.231	1	10
	HHVEH (vehicles available)	2.143	1.137	0	10
	INCOME (Income, US$)	48.611	17.939	2.5	77.5
Land use in 0.25 miles road buffer area around residence	FOOD (restaurants in buffer)	0.361	1.662	0	40
	SHOP (shopping stores in buffer)	0.084	0.379	0	5
	BANK (bank in buffer)	0.073	0.402	0	8
	BUS (bus stop in buffer)	0.586	1.612	0	17
	CHURCH (church in buffer)	0.123	0.519	0	7
Roadway network condition	LENGTH (roadway length, km)	1.80	1.05	0.05	6.36
	CNODE (number of intersections)	10.49	7.05	0	47
	DANGLE (Number of cul-de-sacs)	1.48	1.81	0	15
	Area (area of buffer, km^2)	0.116	0.063	0.011	0.413
Land cover Density	DENSE (High density, %)	0.014	0.051	0	0.95

Note: Income is based on coding the middle value of income categories to calculate the mean of household income. For instance, if household income is between $10,000 and $15,000, then the income is coded as $12,500 when calculating the sample mean.

3,151 households are contained in the sample for study area, with limited geo-location information due to privacy considerations. They are error checked and given a random lat/long address at their census block level using the geo-imputation method described in the previous chapter.

The reported household daily trip frequency is nearly 8.0; among them, 7.0 trips are made by driving and only 0.2 are made by walking and biking. Automobile trips are the dominant trips in this region; nearly 90% of all trips are automobile trips and more than 75% households reported that they do not walk or bicycle on the travel day.

Socio-demographic variables are used as controls. The household characteristics include: average household size of 2.39, with 2.14 vehicles per household and an average annual household income of nearly $49,000. Overall, these numbers are reasonable and in-line with national statistics for similar urban areas.

To capture the land use mixtures in the surrounding area, the count of public facilities within 0.25 miles (400 meters, equivalent to 15 minutes walking distance) of roadway buffers around sampled residences are calculated. The locations of restaurants, shops, banks, and stores are extracted from local yellow pages, which provide a fairly accurate database. Then they are spatially located (latitude/longitude) using a geocode tool. Within 0.25 miles from the residence, 11.3% (356 out of 3,151) households had restaurants, 6% (189) households had a shop or stores, 8.1% (256) households had churches, and 17.5% (552) had a bus stop.

Roadway characteristics include roadway length, number of intersections and number of dead ends. The average roadway length in 0.25 miles network buffer is 1.1 miles (1.8 km), the average number of intersections is 10, and there are 1.5 dangle points.

A percentage of high density land cover is calculated for each buffer around residences. Land cover density statistics show that on average, only 1.4% of the land cover within the 0.25 miles buffer of residence is developed with high density. However, the maximum number for this percentage is as high as 95% in denser areas.

5.3 Model Result

A-priori, better local road connectivity, mixed land use, and better public spaces will likely be associated with lower automobile trips. To estimate automobile trip frequencies, both the base model (with socio-demographic variables only) and enhanced model (with built environmental variables added) are presented. The global (traditional) Poisson regression models are compared with corresponding GWPR to examine statistical properties of the model. Table 7 presents the results of both global model and local models. All of these models are statistically significant overall, show a reasonable fit and provide similar estimation results (based on marginal effects). The marginal effects are presented to facilitate interpretation of the parameter estimates, i.e., the extent of correlation with daily trip frequency. Due to their high collinearity, the number of intersections and the area of buffer are dropped from the models.

Table 7. Global and local Poisson models for household automobile trip frequency

	Global model (Poisson)				Local model (GWPR)											
	Enhanced model		Base model		Enhanced model						Base model					
	β	MFX	β	MFX	β Min.	β Lwr Quart.	β Med.	β Upr Quart.	β Max.	(Upr - Lwr) >2SE	β Min.	β Lwr Quart.	β Med.	β Upr Quart.	β Max.	(Upr-Lwr)>2 SE
Constant	0.834*		0.753*		0.403	0.725	0.852	0.944	1.139	Yes	0.430	0.650	0.778	0.852	1.106	Yes
HHSIZE	0.220*	1.40	0.220*	1.399	0.111	0.207	0.220	0.234	0.302	Yes	0.112	0.206	0.219	0.233	0.306	Yes
INCOME	0.006*	0.04	0.007*	0.043	0.002	0.005	0.006	0.007	0.011	Yes	0.002	0.005	0.006	0.007	0.011	Yes
HHVEH	0.111*	0.70	0.114*	0.728	0.018	0.105	0.121	0.145	0.187	Yes	0.012	0.108	0.127	0.149	0.194	Yes
FOOD	0.020*	0.13			-0.029	-0.011	0.019	0.066	0.259	Yes						
SHOP	-0.027	-0.17			-0.443	-0.073	-0.009	0.040	0.328	Yes						
BUS	0.0001	0.00			-0.211	-0.017	0.004	0.017	0.044	Yes						
CHURCH	-0.077*	-0.49			-0.474	-0.127	-0.026	0.047	0.184	Yes						
LENGTH	-0.021*	-0.14			-0.224	-0.066	-0.011	0.016	0.090	Yes						
DENSER	-0.619*	-3.93			-7.253	-1.105	-0.738	0.638	1.454	Yes						

Summary Statistics

	Global Enhanced model	Global Base model	Local Enhanced model	Local Base model
Corrected AIC:	10,414	10,463	6,713	6,775
Pseudo R^2	0.170	0.168	N/A	N/A
MAD	3.57	3.58	3.48	3.53
RMSE	4.80	4.82	4.65	4.71
Sample size	3,151	3,151	3,151 (812 local sample size)	3,151 (812 local sample size)
Log-likelihood	-5,197	-5,206	N/A	N/A
Prob. > Chi^2	0.000	0.000	N/A	N/A
LR Chi^2 (9)	LR Chi^2 (9)= 4233.68	LR Chi^2 (3) 4172.82	N/A	N/A
Chi^2 test for model improvement: prob> Chi2=0.0000 *			N/A	N/A

Note: MAD -Mean Absolute Deviation = $\frac{1}{n}\sum|y_i - \bar{y}|$; RMSE-Root Mean Square Error, Standard Deviation of the residuals

* Statistically significant—95% level ; MFX = Marginal effect of variables at the mean of that variable

Not all the built environmental variables show significant correlations with automobile trips in the global model after controlling for household characteristics. Specifically, the number of restaurants within buffers shows a statistically positive association with the number of automobile trips, but the number of churches in the buffer show negative association with the number of automobile trips. As for other facilities such as bus stops and shopping stores, the correlations with automobile trip frequency are not statistically significant. For network and land use variables, both the network and development density show significant negative associations with automobile trips.

The comparison between the base model and enhanced model indicates that the including of built environment variables strengthened the models, reducing unexplained variation in the dependent variables, e.g., R^2 values improved from 0.168 for base model to 0.170 for enhanced model. Also, the enhanced model shows better goodness-of-fit compared with the base models, i.e. less MAD (Mean Absolute Deviation) and RMSE (Root Mean Squared Error). Moreover, a Likelihood Ratio chi-squared test showed that adding built environment variables collectively results in statistically significant improvements to the model fit (5% level).

By comparing AICc between the global and local models, the local model outperforms its counterpart, as the AICc values for local model is substantially lower than the global model. As a general rule, improvements in the AIC that are less than 3 in value could easily arise as a result of sampling error (Fotheringham et al., 2002), while here the difference between the global and local models is substantially greater than 3, indicating that the local models are statistically better than the global model. Furthermore, results show that GWPR model has lower MAD and RMSE, compared with the corresponding global Poisson model.

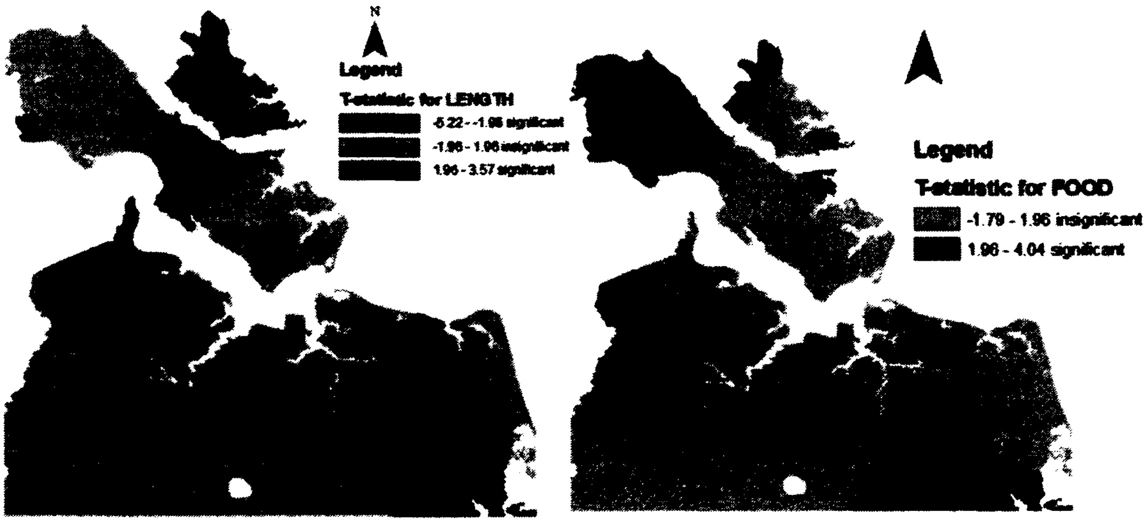
The global enhanced regression model only provides associations between automobile trip frequencies and built environment from an overall regional perspective, i.e., it represents the average relationship between correlates from a regional perspective. The marginal effect shows that on average, one more kilometer of roadway within the 0.25 mile buffer around residence is associated with 0.14 fewer automobile trips. However, the spatial variation of this association is unknown, e.g., whether the association between automobile trips and roadway length in buffer is always negative in this region is unknown. GWPR can be used to detect spatial non-stationarity, exploring if parameters vary across space. It can uncover the possible local spatial deviations of explanatory variables.

The R^2 value for the global regression is 0.17 indicating that it still leaves about 80% of the variance in auto trips. Some of this unexplained variance may result from the general assumption that relationships in the model are constant over space. Suppose, for instance, it is very likely that two similar households behave differently in terms of how they travel, even if both of them have the same roadway length in the neighborhood, but one is located in downtown area, while another one is located in an area close to beach. It is quite possible the household in downtown made fewer driving trips considering the congested traffic nearby, while the one close to beach made more driving trips to the beach. These variances based on different geographic features are not fully captured by global model, and they cannot be easily captured by simply creating a certain variable. One solution is to allow the association to change in space and let the local model search for spatial pattern of associations which yield the optimal specification. If such variations in relationships exist over space, then the global trip-making model will clearly be a misspecification of reality because it assumes these relationships to be constant.

Table 7 also provides the parameter summary for the GWPR model. The distinct difference between global and local estimation is that the global estimation has one set of model parameters for all observations in the sample, while the local model estimates a set of parameters for explanatory variables for each location. Thus the ranges of parameters are provided to show the range of parameters spatial variation. Theoretically all parameters can vary in space. However, it is important to determine if the spatial variance is significant enough to be captured by using the more complicated GWPR model. Estimation of a GWR model with more than 3,000 samples and nine variables is computationally intensive (e.g. it takes more than 20 hours on powerful personal computers). Global models will be appropriate if spatial variations (stationarity in space) are modest. To decide whether the spatial variation is statistically significant, the difference between the lower quartile and upper quartile of a parameter is compared with the standard error of estimate. If the difference is larger than two standard errors, the parameter is considered non-stationary in space (Fotheringham et al., 2002), implying that spatial variance is statistically significant. All explanatory variables show significant spatial variance.

Based on the local parameter estimates for 3,151 households, an Inverse Distance Weighted (IDW) interpolation algorithm is used to assign values to unknown points in space. Thus a continuous parameter surface covering the whole region can be created. Also, a contour parameter surface is generated based on estimation which varies continuously in space. This can give a better picture of where the coefficient is higher in space. Similarly, a local t-test graph can be generated to show whether the association is statistically significant across the study area.

Note that only the socio-demographic characteristics are significant in the entire region. Figure 17 shows the variation in local t-statistic for number of LENGTH and FOOD within 0.25 miles buffer around residence, respectively. Although both LENGTH and FOOD are significant in the global model, there is a large portion in the study area, where they are not statistically significant (95% level). Figure 18 shows the magnitude of LENGTH and FOOD within 0.25 miles buffer association with automobile trips.



a) LENGTH b) FOOD

Figure 17. Local t-statistic in enhanced GWPR model.

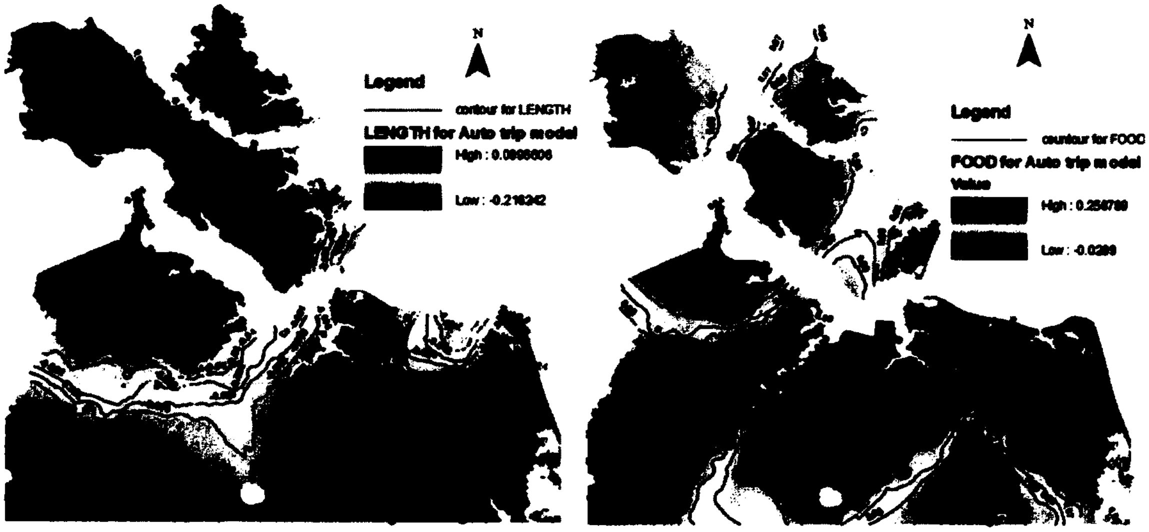


Figure 18. Parameter estimated for LENGTH and FOOD in enhanced GWPR model

Although roadway length within 0.25 buffers has a negative association with automobile trips in the global model, the local model shows conflicting signs for this variable, depending on the area. In the southeast of study region, this correlation becomes positive, while in the northwest of the coefficient map it shows a negative sign. The net effect of LENGTH still remains negative, which is reflected in the global model. For number of restaurants within a 0.25 miles buffer, all the areas with significance show positive association with automobile trips. Since the distribution of t-statistics showed spatial clusters, the model results are verified by estimating two unrestricted Poisson regression models using sub-samples that showed significant associations versus samples that showed insignificant association. The results from unrestricted models confirmed that indeed associations and their significance levels vary in space.

Comparisons between the global and local models can be obtained by computing the residual of the predicted results using the models calibrated in this study. For the global model, the fixed equation is applied to every household; for the local model, different equations are applied for each residence considering parameters vary by location. Figure 19 shows the residual level by comparing the results of global model and local model.

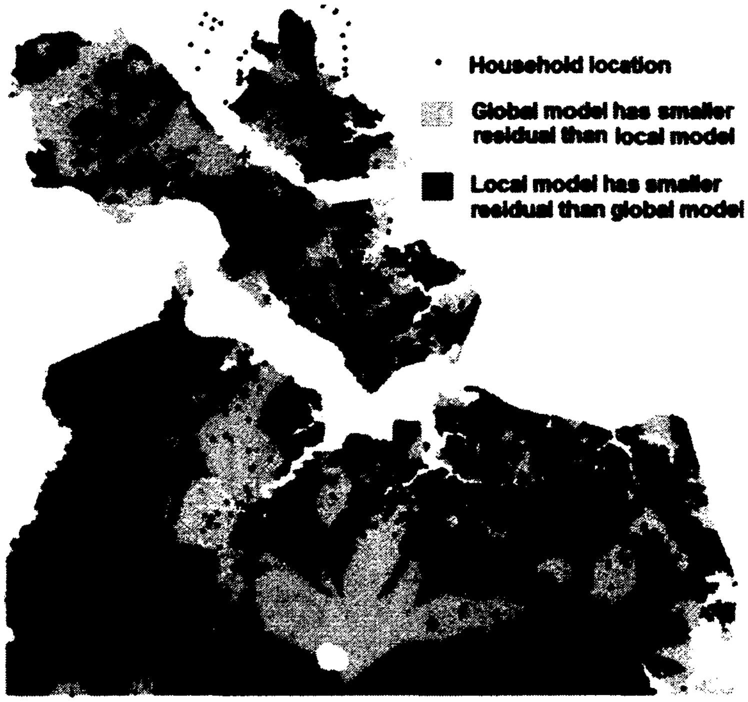


Figure 19. Global model vs. local model – goodness of fit

The darker region from Figure 19 is where the local model has smaller residual than global model, which represents better estimation. This indicates that using the global model for prediction can bring errors in certain areas, e.g., the downtown Norfolk. Relocating the trip productions will change the trip distribution in this region significantly, and it will have a substantial impact on subsequent steps, i.e., trip distribution, mode split, and traffic assignment.

5.4 Applications

The model results show that simply defining an area as urban, suburban or rural and assigning it a higher or lower trip rate can be arbitrary. Figure 20 and 21 show the variation in parameter for household vehicles owned and household size respectively. Substantial spatial variation of these two parameters is found. Thus, the commonly used cross-classification tables by urban or rural area can be considered archaic when richer data are available for spatial analysis. Practically, by using the contour parameter map created by the local model (shown in Figure 21, Figure 22, Figure 23). Customized cross-calculation table can be created for each location, providing a finer forecast of trip productions.

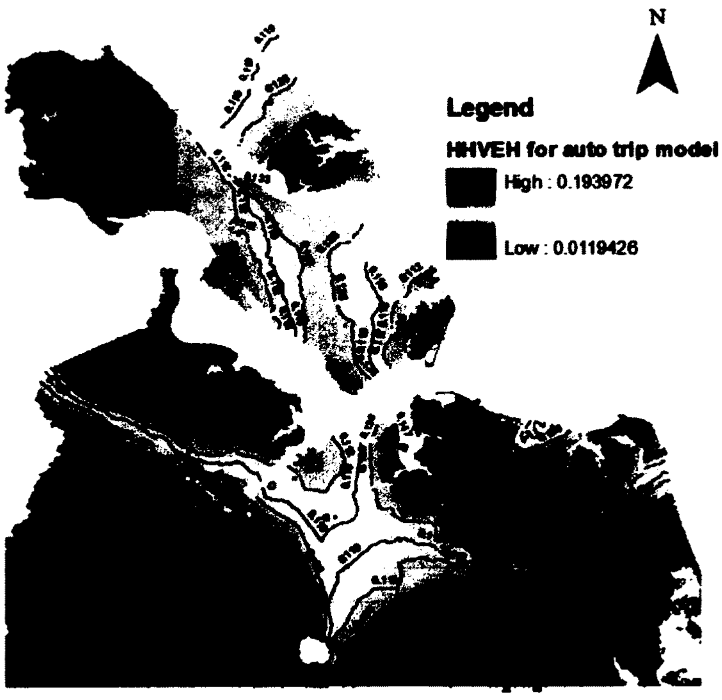


Figure 20. Parameter estimates for HHVEH in GWPR base model

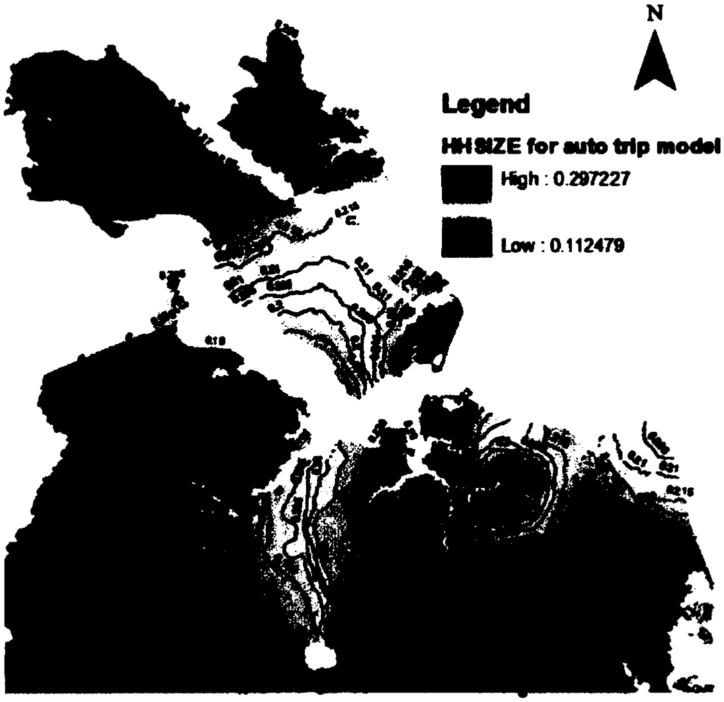


Figure 21. Parameter estimates for HHSIZE in GWPR base model

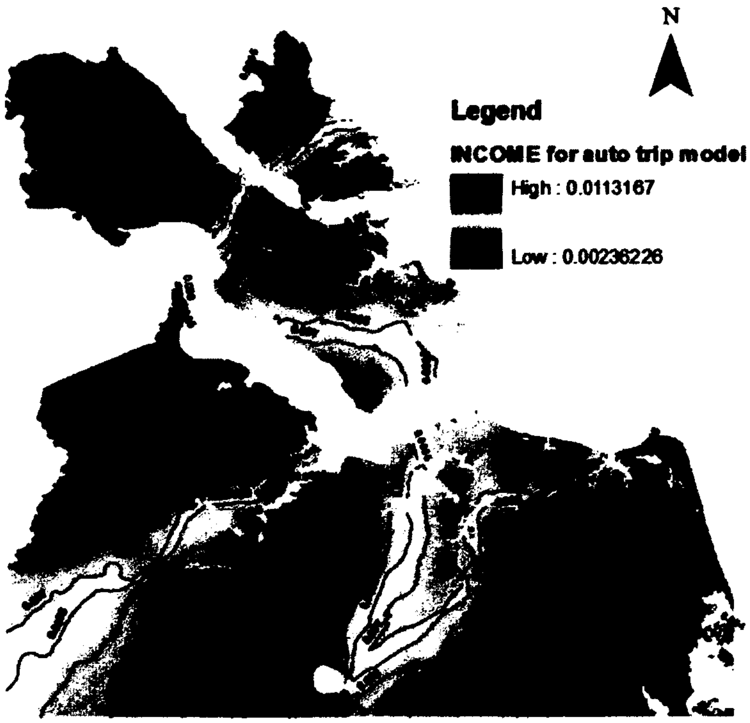


Figure 22. Parameter estimates for INCOME in GWPR base model

Comparisons between the global and local models can be obtained by computing the difference in predicted trip frequency using different models. For the global model, the fixed equation is applied to each TAZ; for the local model, different equations by TAZs are applied. Figure 23 shows the difference between predicted daily trips from these two models, using the same 2030 population, vehicle ownership, and other household characteristics available in the Hampton Roads travel demand model. Since the global model is a spatial average of local models, both of these models, when applied for prediction purposes, should predict the same trip totals. Nevertheless, there is a small difference (less than 5%) in the total number of trips predicted by the global Poisson model and GWPR model.

Figure 23. Differences in daily trip productions between global and local models

A positive value in Figure 23 implies that more trips are predicted using the local model. This indicates that using the global model for prediction can overestimate the travel demand in certain areas, e.g., the city of Norfolk, while the demand in areas such as northern Virginia Beach and areas north of Hampton Roads Bridge Tunnel (HRBT) is underestimated. Relocating the trip productions will change the trip distribution in this region significantly, and it will have a substantial impact on subsequent steps, i.e., trip distribution, mode split, and traffic assignment.

5.5 Discussion

Interpreting associations based on a global model can obscure spatial variations by providing the net result only. This chapter shows that a global model can provide a positive significant relationship for a coefficient that is only significant in a relatively small portion of the study region. In some cases, variables can show opposite signs, depending on the region. Therefore, global models have a tendency to mask complex information behind the average association, and in some cases provide misleading conclusions. These results suggest that 1) accounting for spatial variations in associations between built environment and automobile travel can help identify areas where focusing land use policies can have the highest impacts (by checking the map of parameters in local models), and 2) for regional level transportation models that attempt to integrate land use, analysis should take into account spatial variations, especially for metropolitan areas with substantial variations in socio-demographics and built environments. Simply using a pooled model without considering spatial heterogeneity can be misleading. Finding proper levels of spatial clusters, e.g. using neighborhood databases within a regional survey, should be considered carefully when exploring the associations of built environmental variables with travel demand.

It also demonstrates how new methods that capture spatial heterogeneity can be applied to improve travel demand models. The visualized coefficient maps obtained from the local models are valuable in quantifying spatial variations and in an easy-to-understand format for policy makers, engineers, and planners. Moreover, this method can be of interest to policy makers who rely on travel demand forecasting models for decision making. The research can help advance the state-of-the-art in travel behavior research by using rigorous analysis techniques to incorporate spatial variations into travel demand models. Practitioners may capitalize on the greater spatial variability in parameters to develop locally-based strategies for trip reduction, especially where trips are particularly numerous.

6. UNIVERSITY CAMPUS: A CASE STUDY OF SPECIAL GENERATOR

The previous chapter showed evidences from the local models that spatial variances exist in the associations between built environment and trip-making, but these models are estimated based on data from relative large scale. For a relatively small scale such as university campus, the local models used to capture spatial heterogeneity may not be suitable. One reason is that GWR uses the moving windows regression method to estimate the data, which means a subgroup of samples around each location is used to estimate local models. This method works well in a relatively large region when samples are scattered in space, but it does not work well when samples are extremely clustered, which is the case for university students. Therefore a different spatial analysis should be applied, especially considering the university campus has its own characteristics, e.g. mix of population and alternative friendly environment. As a special location in space, campus serves as both a trip generator and a trip attractor, which has strong centripetal force to daily traffic. Especially in urban areas, the university students commute from a wide range of areas to the school; therefore, it's reasonable to speculate that there may be rings of mobility around the university campus. Moreover, university students are usually uncovered by traditional travel behavior surveys and are underrepresented in the travel demand model. To address this shortcoming, this chapter is to model the student travel behavior using particular spatial analysis. Consideration is given to their unique stage in their lives, the special nature of university students' personal characteristics, lifestyle (both working and studying) and the spatial factors of where they live and study/work. The insights gained from this study can serve as the basis for trip generation in regional travel demand models, where university-dominated zones are treated as special generators. It can also shed light on how a university campus environment, which has a mix of land uses (e.g., office/classes, residential, and commercial), is alternative mode friendly and higher density, which is associated with students' driving and walking/bicycling behavior.

6.1 Data Description

6.1.1 Spatial Analysis of Trip Making

On-campus and off-campus students usually show different travel behaviors, which may be due to the unique context in which universities campus provides, e.g., land use mix (academic buildings and students activity centers, shops), sidewalks, bicycle paths and bicycle parking facilities, etc. (Khattak et al., 2011). To obtain a more comprehensive view of how students' travel behavior varies by their residential status, a spatial analysis is conducted to group them based on residential proximity to the campus. The ODU campus is selected to conduct this spatial analysis since it represents a more complex situation in urban area.

In ODU, due to limited dormitory space provided by the university and no dedicated graduate student housing, most students (81%) live off-campus. However, some of their residences are physically close to campus, even if they do not reside in campus dormitories. These near-campus students share a similar built environment as on-campus students. Therefore, their travel behavior is expected to be similar, due to their proximity to the campus, but still different from the on-campus students, due to possible differences in socio-demographic characteristics. To better understand the travel behavior of near-campus students and those students who live farther from campus, GIS analysis is conducted. While some of the campus buildings are intermingled with privately owned properties, a synthetic campus boundary is defined by using all campus buildings to create a standard ellipse (the shape of the ODU campus is better captured using the ellipse). Near-campus students are identified as those who live within 1 mile from the standard ellipse of campus buildings. The rest of students who live outside of the 1 mile buffer area are termed "farther-from-campus" students.

Figure 24 shows the range of the synthetic campus boundary and sampled students' residences. The statistics shows that 19% of students live on-campus, another 15% of students live near-campus, and the rest (66%) reside farther from campus, outside the influence area of the main campus. The finding from the spatial analysis provides input into the university student travel demand modeling.

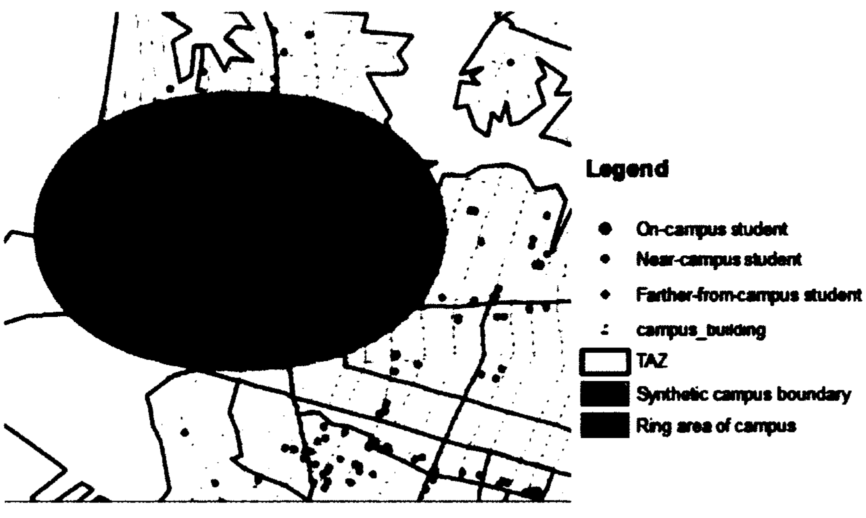


Figure 24. Residential locations of on-campus and near-campus university students

6.1.2 Descriptive Statistics

Table 8 shows the descriptive statistics of personal and travel characteristics and breakdown statistics for on-campus, near-campus and farther-from-campus students.

Table 8. Descriptive statistics of ODU students compared with the general population Hampton Roads, VA population

		All students Mean (SD. min/max)	On-Campus students	Near-campus students	Farther-from campus students	General population of HR
Sample Size (N)		1,468	275 (19%)	216 (14.7%)	977 (67%)	6,543
Personal Property	MALE (binary variable)	0.40 (0.49, 0/1)	0.39	0.50	0.38	0.47
	AGE (years)	24.78 (7.81, 17/81)	19.37	22.54	26.8	48
	INCOME ($1,000)	16.54 (20.04, 5/100)	7.53	10.65	22.38	N/A
	NVEH (no. of vehicles)	1.73 (1.12, 0/7)	0.99	1.36	2.02	2.14**
	WORK (binary variable)	0.54 (0.50, 0,1)	0.20	0.49	0.64	0.52
Academic Property	FULL TIME (binary variable)	0.83 (0.37, 0/1)	0.98	0.95	0.76	-
	UNDERGRAD (binary variable)	0.79 (0.41, 0/1)	0.99	0.77	0.74	-
Trip Property	NTRIPS (daily frequency)	5.26 (2.81, 0/15)	6.24	5.34	4.96	4.58
	NTRIPS-AUTO	3.09 (2.52, 0/15)	1.03	2.09	3.88	3.29
	NTRIPS-WALKBICYCLE	2.04 (2.74, 0/15)	4.94	3.09	1.00	0.34
Travel day	WEEKEND	0.37 (0.48, 0/1)	0.25	0.36	0.41	-

Notes: †Results in the last column are for the Hampton Roads, VA general population based on NHTS (National Household Travel Survey) Virginia add-on sample collected in 2009.

* SD means standard deviations.

**Number of vehicles owned by the household is equivalent to the number of vehicles available for use by students.

In line with expectations, university students are young, busy (going to school and work) and have relatively low incomes. Most students are between 23 to 26 years old, with an average age of 25. Most respondents (79%) are undergraduate students. The average number of vehicles available for use is 1.73, and 96% of students have a driver's license. The average annual income is about $16,000, while 50% of students in the sample earn less than $10,000 per year. Income is based on coding the middle value of income categories to calculate the mean of household income. For instance, if the reported income is between $10,000 and $15,000, then the income is coded as $12,500 when calculating the sample mean. On average, 83% of students are full-time students, and 54% of them worked for profit. Overall, these numbers are reasonable.

The reported student daily trip frequency is, on average, more than 5 with a variance of 7.89; among them, 3 trips are made by driving with a variance of 6.35, and 2 trips are made by walking/bicycling, with a variance of 7.5. The mean and variances of the total trip frequency distribution are close enough to expect that Poisson regression will be appropriate; for automobile and walking/bicycling trips, over-dispersion indicates that Negative Binomial regression may be appropriate. The average number of daily trips for students is significantly higher (5% level) than a sample of the general population of Hampton Roads. Students made substantially more walk/bicycling trips than the sample of the general population, especially those students who live close to the campus—the Virginia NHTS add-on sample showed that 90% of the trips are by motorized vehicles. For automobile trips, the on-campus and near-campus students substantially made fewer trips than the sample of general population. Interestingly, students residing farther from the campus drive more than the regional population, partly because most of them (65%) are working while going to school.

Moreover, differences in travel behavior are found between on-campus students and off-campus students. On-campus students are often younger, unmarried, full-time, and most of them are undergraduate students, as expected. Compared with off-campus students, on-campus students have a higher daily trip rate, different mode choices, and different trip purposes. They tend to drive less and walk more. More specifically, on-campus students make 6.24 trips per day, compared with 5.34 trips (14.4% fewer total trips) for near-campus students, and 4.96 trips (20% less) for farther from campus students. However, on average the students in farther from campus region make nearly 2 more auto trips than near-campus students and near-campus students make 1 more auto trip than on-campus students. Consistent with our expectation, proximity to campus seems to be associated with both travel demand and mode choice.

6.2 Model Result

After checking correlation between independent variables, final models are presented in Table 9. Poisson and negative binomial models are estimated for both the total daily trips and trips by mode; the zero-inflated Poisson model is estimated for automobile trips and walk/bicycle trips. Note that students' family members at home may influence their travel decisions. The survey requested information about students' families, including whether the students are married and whether they live with family or roommates. These variables are considered for inclusion in the model specification. However, these variables are either highly correlated with students' proximity to campus, or did not show significant association with the dependent variable, i.e., trip frequency.

Thus, they are dropped from the final models. Furthermore, availability of bicycle lanes, sidewalks and proximity to transit stops are considered in the model specification. However, due to the predominantly auto-oriented regional design, the alternative mode facilities are limited, especially for off-campus students. Only a few bus stops are available in and around the campus. Therefore, these variables are dropped from further consideration in the model.

All models are statistically significant at the 0.05 level. The Poisson and Negative Binomial models for total trips provide very similar coefficients. The over-dispersion parameter (α) in the Negative Binomial model is close to zero, indicating that the simpler Poisson model may be acceptable (even if α is statistically significant). The automobile trip and walk/bicycle trip models have reasonable Pseduo-R^2 values. The Pseudo-R^2 for these models is higher for Poisson models indicating a better fit to the data. However, the Vuong test indicates that the zero-inflated Poisson model is more suitable compared with the standard Poisson regression model. Thus, it will be preferable to use the Poisson model for total daily trips and zero-inflated Poisson model for automobile and walk/bicycle trips.

Table 9. Trip frequency model results for Old Dominion University students

Independent variable	Total Trips				Auto Trips						Walk/bicycle Trips					
	Poisson Model		Negative Binomial Model		Poisson Model		Negative Binomial Model		Zero-inflated Poisson Model		Poisson Model		Negative Binomial Model		Zero-inflated Poisson Model	
	β	IRR	β	IRR	β	IRR	β	IRR	β	IRR	β	IRR	β	IRR	β	IRR
CONS.	1.536 (0.000)	-	1.535 (0.000)	-	-0.345 (0.000)	-	-0.376 (0.000)	-	0.651 (0.00)	-	1.756 (0.000)	-	1.459 (0.000)	-	1.546 (0.003)	-
WORK	0.098 (0.000)	1.100	0.101 (0.001)	1.106	0.211 (0.000)	1.235	0.230 (0.000)	1.256	0.123 (0.000)	1.132	Insig. (dropped)		Insig. (dropped)		Insig. (dropped)	
NEAR-CAMPUS	-0.125 (0.001)	0.882	-0.126 (0.006)	0.880	0.634 (0.000)	1.885	0.630 (0.000)	1.877	0.156 (0.064)	1.169	-0.246 (0.000)	0.782	-0.192 (0.082)	0.830	-0.189 (0.000)	0.827
FAR-CAMPUS	-0.180 (0.000)	0.836	-0.182 (0.000)	0.834	1.159 (0.000)	3.186	1.150 (0.000)	3.168	0.328 (0.000)	1.389	-1.088 (0.000)	0.337	-1.119 (0.000)	0.334	-0.487 (0.000)	0.614
UNDERGRAD	0.191 (0.000)	1.211	0.192 (0.000)	1.212	0.144 (0.001)	1.155	0.150 (0.002)	1.162	0.103 (0.009)	1.108	0.249 (0.000)	1.283	0.410 (0.000)	1.506	0.144 (0.064)	1.155
AGE	Insig. (dropped)				0.007 (0.000)	1.007	0.007 (0.007)	1.007	0.007 (0.000)	1.007	-0.039 (0.000)	0.962	-0.035 (0.000)	0.969	-0.010 (0.039)	0.90
NVEH					0.055 (0.000)	1.057	0.063 (0.001)	1.065	0.029 (0.040)	.03	-0.041 (0.023)	0.960	-0.047 (0.175)	0.955	-0.055 (0.004)	0.947
FULL-TIME	0.126 (0.001)	1.134	0.126 (0.002)	1.134	Insig. (dropped)						0.514 (0.000)	1.673	0.604 (0.000)	1.733	0.301 (0.004)	1..354
INCOME	Insig. (dropped)										-0.004 (0.006)	0.996	-0.005 (0.054)	0.942	Insig. (dropped)	
WEEKEND	-0.160 (0.000)	0.852	-0.159 (0.000)	0.853							-0.421 (0.000)	0.656	-0.448 (0.000)	0.628	-0.137 (0.004)	0.872

Summary statistics

Dependent variable	NTRIPS (Daily number of trips) (total number of obs.=1,468)		NTRIPS-AUTO (Total Number of obs.=1,468, Zero obs.=301)			NTRIPS-WALK/BICYCLE (Total Number of obs.=1,468, Zero obs.=751)		
Prob. > Chi^2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Pseudo-R^2	0.025	0.018	0.12	0.08	-	0.26	0.09	-
LR Chi^2	179.89	125.19	848.72	489.79	87.40	1,938.96	494.03	253.53
α	-	0.073	-	0.188	-	-	1.024	-
Prob. α	-	0.00	-	0.00	-	-	0.00	-
			Note: ZIP binary model Y = 0.63(0.0)-0.92*WORK (0.0)-0.86*NEAR-CAMPUS (0.0)-3.75*FAR- CAMPUS (0.0) (p-values in parentheses)			Note: ZIP binary model Y = -1.64(0.0)-0.64*FULL-TIME (0.0)+0.01*INCOME (0.0)+1.01*NEAR-CAMPUS (0.0)+2.53*FAR-CAMPUS (0.0) (p-values in parentheses)		

Note: P-value in parentheses; IRR means Incident Rate Ratios.

Table 9 also shows the Poisson models for zero and non-zero observations of automobile trips and walk/bicycle trips. Students who work and live near-campus or farther-from-campus are less likely to make zero automobile trips on their travel days; students who are part-time, live off-campus, and have higher income are more likely to make zero walk/bicycle trips on their travel days, as expected.

Incident Rate Ratios are exponentiations of parameters, e^{β}, and they facilitate the interpretation of coefficients, i.e., whether it is associated with an increase or decrease in the expected trip rate, when the value of the explanatory variable changes. As expected, the ratios show that undergraduate students, full-time students, and students who work are likely to make more daily trips; students who reside off-campus (including those who live near-campus or farther-from-campus) make fewer daily trips. Also, trip frequencies on weekends are lower than weekdays. Unlike the general population, total daily trips of students do not show significant association with the number of vehicles available to students or their income levels. The model can predict trip frequency as follows: if a full-time undergraduate student is also working, and resides near-campus, then on a weekday he or she is expected to make 5.29 trips (= $e^{(1.536 + 0.098 - 0.125 + 0.191 + 0.126 - 0.16)}$), based on the Poisson regression model for total trips.

Some coefficients show different signs in automobile trips and walk/bicycle trips models. Especially for living conditions, compared with on-campus students, near-campus students make nearly 90% more driving trips, while farther-from-campus students make four times as many driving trips, on average. The relationship is opposite for walk/bicycle trips. Compared with on-campus students, near-campus students make 20% (=1-0.8) less walk/bicycle trips and farther from campus students make 70% less walk/bicycle trips, holding other variables constant. In general, compared with on-campus students, near-campus and farther-from-campus students walk less and drive more, but with a net consequence of relatively fewer total trips. Figure 25 shows there are clear tendency of how students travel, based on the different ring area around campus.

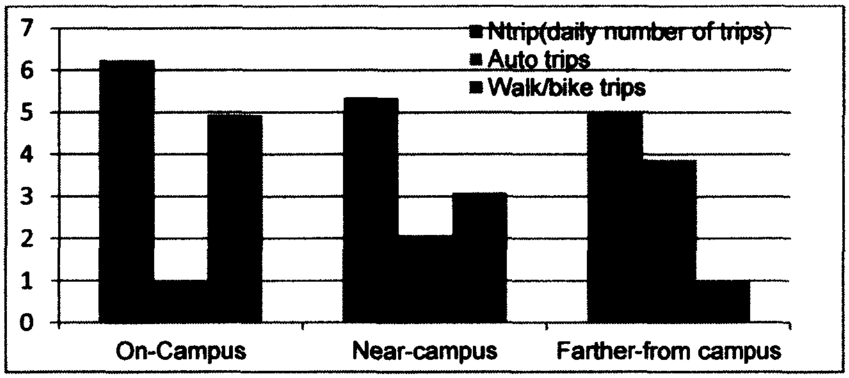


Figure 25. University students' trip characteristics by residential locations

6.3 Discussion

Given the complicated associations based on the model results, the directions between associations are summarized in Table 10.

Table 10. Associations found in the trip frequency analysis

	Variable	NTRIPS model	NTRIPS-AUTO model	NTRIPS-WALK/BICYCLE model
Personal Characteristics	NVEH	N/A	+	-
	INCOME	N/A	N/A	N/A
	AGE	N/A	+	-
	WORK	+	+	N/A
Living location	FAR-CAMPUS	-	+	-
	NEAR-CAMPUS	-	+	-
Academic condition	UNDERGRAD	+	+	+
	FULL-TIME	+	N/A	+
Travel day	WEEKEND	-	N/A	-

The results show that more active undergraduate students drive more, walk/bicycle more, and ultimately make significantly more trips than other students. Work for profit shows a significant association with total trips and automobile trips, but not with walk/bicycle trips. Also, age and number of vehicles show different signs in automobile trips and walk/bicycle trips, i.e., younger students walk/bicycle more while older students make more automobile trips. However, these associations cancel each other and do not show up in the total trips model, i.e., the age variable is not statistically significant. Vehicles available to students are associated with higher automobile trips and lower walk/bicycle trips, as expected. However, this variable did not show a statistically significant association (5% level) with total trip frequency. Full-time students make more walk/bicycle trips compared with part-time students. On weekends, students make fewer total trips, but this may due to less walking/bicycling trips, not necessarily mean they drive less.

The model results confirm there is a ring of mobility around university campus as a special trip generator. Notably, the proximity of the students' residence to campus is strongly associated with their travel. Compared with on-campus students, near-campus students and farther-from-campus students made fewer daily trips. If breaking down trips by transportation mode, the percent of walking trips drops from 79% for on-campus students to 20% for far-from-campus students, while the percent of driving trips increases from 17% for on-campus students to 78% for far-from campus students. Large university campuses are major trip generators and can impact the regional traffic. The findings about student travel can also help design practical strategies to improve the traffic conditions in and around the university campus by establishing satellite communities near the campus, e.g., providing better on-campus or near-campus student villages, encouraging traditional neighborhood developments within walking or bicycling distance from university campus (where feasible and appropriate), creating a pedestrian and bicycle friendly design on and near campus, adding public facilities in surrounding communities, and connecting regional transit corridors with university campuses.

7. TRAVELER INFORMATION AND TRAVEL DECISION CHANGES

The rapid development of information technologies applications in transportation has provided customers with more diverse and dynamic information. Advanced traveler information systems (ATIS), part of intelligent transportation systems, are playing a key role in this regard. Nowadays, a variety of technologies, including the internet, telephone services, television/radio broadcasts, dynamic message signs, and in-vehicle/on-board devices are available to provide pretrip and en-route information to help travelers make more informed decisions (change route, mode, departure time or cancel trip). It is generally believed that providing travelers with relevant information on travel options has the potential to change their behaviors in ways that are beneficial to the efficient use of the transport system (Chorus et al., 2006b). Thus, it is important to understand how travel information is used and whether travelers are willing to make travel decision changes based on the traveler information they receive.

Most of the current literature only concentrates on the non-spatial outcomes of travel decisions as well as traveler information delivery mechanisms. However, the association of various socio-economic and contextual factors may vary across space. For example, the usage of dynamic travel information may vary substantially across locales even for people with the same income level. It is desirable to ask: where are the parts of a region where people with higher/lower income or longer travel time are more sensitive to information acquisition and travel decision changes? Such a question cannot be fully answered directly by standard parameter estimation models —called a global regression model, since the estimated parameters are fixed and can be understood as a spatial average in global models. Although unrestricted models can be estimated for various spatial sub-classifications and compared with a global or pooled model, this is often cumbersome and rarely done—the problem of fixed coefficient still exists within the sub-classifications and the definition of the boundary of those sub-classifications will influence the estimation of the coefficients, referred to as the (undesirable) "boundary effect."

Besides the spatial heterogeneity issue mentioned above, notably the information technology innovations which have developed rapidly recently, represented by wider internet access, there is no comparison study to explore whether there is change in terms of how travelers respond to this new tendency. Meanwhile, as a special young subgroup of the population, university students have greater access to information technology, especially when they are on campus. They may be likely to be pre-disposed to media usage for planning their travel and especially to using new online technologies well into the future (Son et al., 2011). Therefore, university students' information access is likely to be an important factor influencing their travel information acquisition. Given that the market segment of university students is known to be more technology-savvy than the general population, it would be interesting to compare the ATIS acquisition behavior between general population and also compare how they respond to ATIS based on the traveler information they received.

To this end, this chapter attempts to understand the travelers' information acquisition and their travel decision adjustment based on the information received. Comparison between university students and general population is drawn to show their differences in preference. The spatial heterogeneity issue is captured by using GWR again when analyzing larger metropolitan areas and the coefficients are mapped to show the spatial pattern of the associations. The insights gained from this study will serve as the basis for developing specific policy guidelines to encourage travel decision changes in response to traveler information received and to reduce travel uncertainty.

7.1 Data Description

The surveys which cover both ATIS usage and activity-based travel behavior are not abundant. Thus, the data used in the chapter from multiple sources. The information for general population is from the activity-based travel survey dataset in the Greater Triangle Travel Study conducted in 2006. The information for college students is from the activity-based travel survey for university students (USTS) conducted at four universities of Virginia in 2009.

The information sources in the surveys include television, Internet, commercial radio, telephone or Traveler Information Hotline (511 in Virginia), Traveler Information Radio (TIR, in North Carolina) or Highway Advisory Radio (HAR in Virginia), and Variable Message Signs (VMS). Note that traveler information includes general pre-trip travel information such as commercial radio traffic reports, and television broadcasts of travel information, as well as en-route information available to travelers in the area such as the updates of traffic conditions and incident and travel advisories. The content of travel information available to travelers in the area is mostly qualitative traffic reports of congestion/delays, and real-time details of traffic incidents. Changes in travel decisions include changing departure travel time, mode, route and cancelling the trip.

7.1.1 The General Population Statistics

The descriptive statistics for the general population in the Research Triangle are shown in Figure 26 and Table 11. The survey only investigated whether the traveler adjusted their travel decisions based on the information received for those who acquired the traveler information. Those who did not acquire the traveler information are excluded for the decision changes questions. The majority of respondents (51%) reported that they acquired travel information at least once a week, which means that 49% seek travel information less than once a week. 78% of information seeker, that is about 40% of all respondents, reported that they changed travel plans based on the traveler information received.

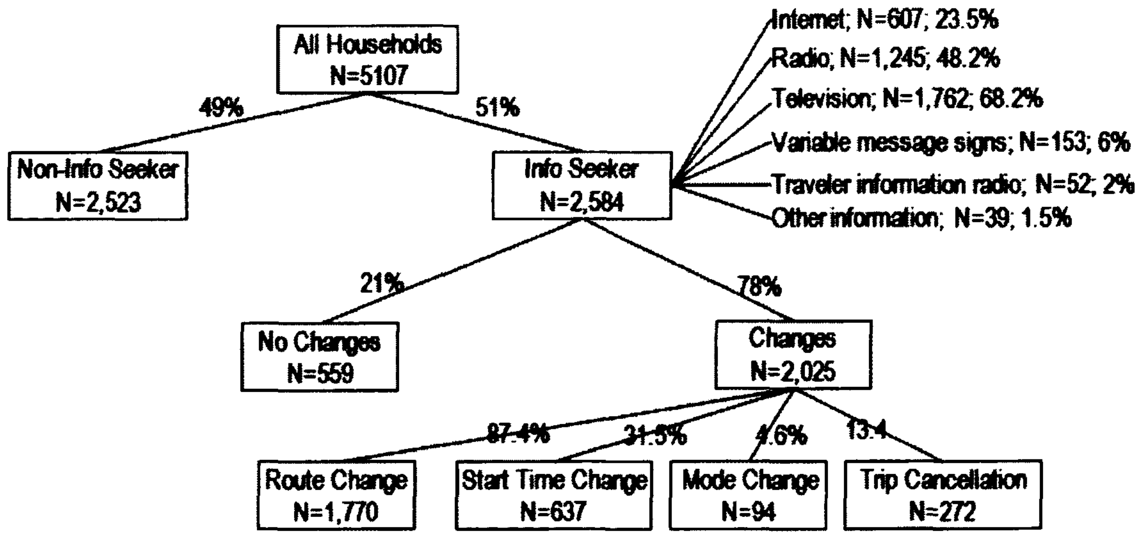


Figure 26. Traveler information acquisition for the general population

Table 11. Descriptive statistics for variables (Research Triangle, NC)

Variable		N	Mean	Std. Dev.	Min	Max
Info	Acquisition of traveler info or not (1=yes, 0=no)	5107	0.51	0.50	0	1
Change	Change travel plan or not (1=yes, 0=no)	5107	0.40	0.49	0	1
Na	Number of info sources accessed (based on info=1)	2584	1.49	0.73	1	5
Work	Work related travel time (minutes)	5107	26.04	43.08	0	673
Nonwork	Non-work related travel time (minutes)	5107	60.74	62.24	0	750
Dummy_w	Travel for work related purposes (1=yes, 0=no)	5107	0.43	0.50	0	1
Dummy_nw	Travel for nonwork related purposes (1=yes, 0=no)	5107	0.17	0.28	0	1
Finfo	Frequency of info acquisition (weekly)	5107	1.96	2.25	0	5
Live	Length lived at this address (year)	5107	6.68	3.51	0.5	10
Income	Household income (categories)	5107	6.289	3.09	0.75	10
Age	Age of respondent (household head, years)	5012	51.17	15.19	8	98
Hhveh	Household vehicle number	5107	2.01	1.02	0	8
INTERNET	Acquire info from internet or not	2584	0.23	0.42	0	1
RADIO	Acquire info from radio or not	2584	0.48	0.50	0	1
TEL	Acquire info from television or not	2584	0.68	0.47	0	1
VMS	Acquire info from variable message signs or not	2584	0.06	0.24	0	1
TIR	Acquire info from trans. info radio or not	2584	0.02	0.14	0	1
OTHER	Acquire info from other sources or not	2584	0.02	0.12	0	1

Note: The travel time used in the analysis is that reported by the head of the household.

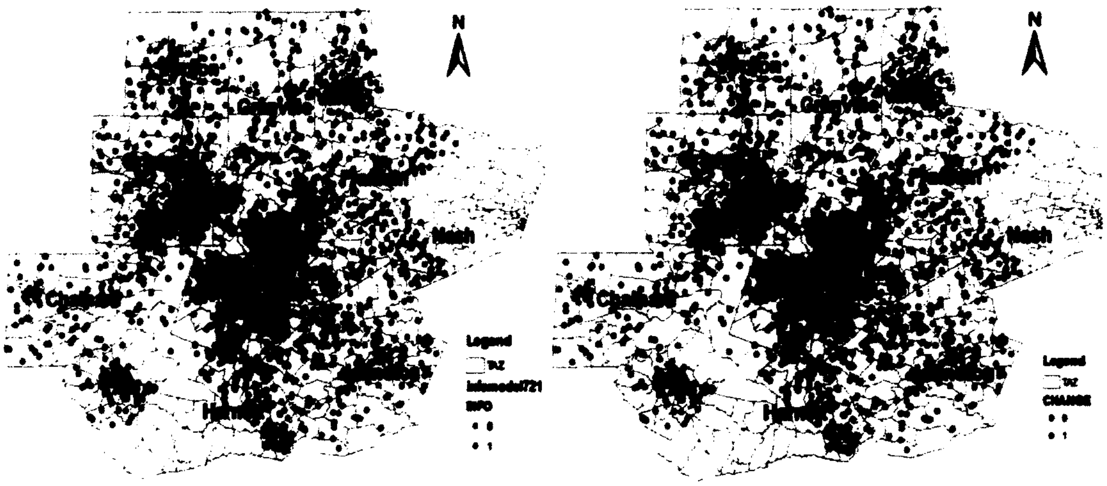
Several variables in the dataset are categorical variables, e.g. household income, the length lived at current address and the frequency of traveler information acquisition. These variables are converted to continuous variables to save computational burden, complexity of interpretation, and loss in degrees of freedom. For this purpose, the mean method (Rossi and Conan-Guez, 2002) is used to recode interval data. That is, the mean of each interval is used to represent the category. For household income, 0.75 = income is less than \$15,000, 2 = income is between \$15,000 and \$24,999, 3 = income is between

$25,000 and $34,999, 4 = income is between $35,000 and $49,999, 6.25 = income is between $50,000 and $74,999, 8.75 = income is between $75,000 and $99,999 and 10 = income is greater than $10,0000. There are 285 households who did not answer this question. These data are replaced by the mean of household income which is $62,890.

The length a person lived at their current address is recoded as 0. 5 = <1 year, 1.5 =1 to 2 years, 3.5 = 2 to 5 years, and 7.5 = 5 to 10 years, 10 = longer than 10 years. The average length a respondent had lived at their current address is 6.68 years. The average number of household vehicles is 2.0. Noting that there are some outliers with the variable age, 95 persons did not answer this question; the average age of the remaining respondents is 51 years.

To capture the various information sources used, the number of information sources accessed variable (NA) is created, by counting the sources used to seek travel information; where 0 = 0 information source used, 1 = 1 information source used… 5 = 5+ information sources used. The frequency of traffic information use (FINFO) is coded as 0 = never, 1 = at least once a week, 3 = 2-4 times per week, and 5 = 5+ times per week. The average number of information sources accessed is 0.76. Since NA and FINFO are only valid for people who were willing to use traveler information, they are only used as correlates in the Change model.

The variable 'INFO' and 'Change' are binary. The spatial distributions of the variables 'INFO' and 'Change' are shown in Figure 27. From the graphs, they are slightly different, but generally, people who access traveler information and those who adjust their travel decision are distributed all around the study region—strong spatial clustering cannot be observed.



a) Information acquisition b) Travel decision change (1=yes, 0=no)

Figure 27. Spatial distribution of two dependent variables

In terms of travel time, trips with a "work" or "work related" purpose and all return trips from a work place to home are defined as work-related travel. All the other trips are considered non-work related trips. The average non-work related travel time is longer than work related travel time (61 minutes vs. 26 minutes per day). Considering that differences in information usage/travel decision adaption may exist between travelers who traveled and those who did not make work or non-work related trips, two dummy variables are created. Among them, 'Dummy_w' captures whether work related trips are reported; 'Dummy_nw' represents whether non work related trips are reported. Nearly 43% of respondents did not make work related trips on the survey day and 17% of the respondents did not make non-work related trips on the survey day.

7.1.2 University Students Statistics

The descriptive statistics of sampled university students' personal characteristics in four Virginia universities are shown in Table 12.

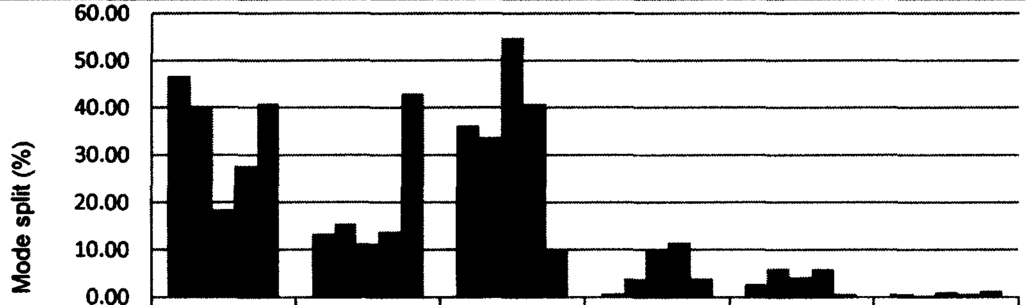
Table 12. Sample characteristics for university student surveys

Source: (Son et al., 2011)

		Urban		Suburban	
		ODU Ave. (SD)	VCU Ave. (SD)	UVA Ave. (SD)	VT Ave. (SD)
Sample size		962	661	996	1039
Gender (%)	Male	38	32	38	48
	Female	62	68	62	52
Degree (%)	Undergraduate	74	52	61	58
	Graduate	26	48	39	42
Residence (%)	On-campus	21	12	36	27
	Off-campus	79	88	64	73
Enrollment (%)	Full-time	80	83	96	96
	Part-time	20	17	4	4
Age (years old)		25.3 (8.3)	25.9 (7.8)	23.1 (6.6)	23.5 (6.3)
Annual income ($1000)		20.3 (25.8)	19.5 (25.2)	15.6 (22.4)	13.5 (17.7)
Commute distance (miles) 1) (university/work)		12.1 (14.2)	9.8 (16.3)	4.3 (14.6)	4.8 (14.9)
Commute duration (minutes) 2) (university/work)		23.6 (21.1)	19.4 (17.6)	15.1 (19.2)	14.6 (25.5)
Vehicle ownership (%)		91 (29)	90 (30)	72 (45)	82 (39)
Living year round (%)		82 (38)	75 (43)	39 (49)	43 (50)

NOTES: The maximum distance is truncated at 120 miles;
The maximum duration is truncated at 180 minutes.

University students are a young and low income group, as expected. The average age of survey respondents ranges from 23 to 26 years old across universities, and their average incomes are distributed from $13,500 to $20,300. Differences exist between urban and suburban campuses in terms of their personal characteristics and the daily traffic conditions. The income and vehicle ownership of students from urban campuses are higher than those from suburban campus, and they are also more likely to live in the town year-round compared with their suburban peers. Also, urban campuses have more part-time students and more students living off-campus. The students from urban campuses report longer average commute (to university/work) distances. Furthermore, the comparison shows different preference on how to travel (shown in Figure 28).



	drive alone	share ride	walk	bus	bicyle	other/missing
ODU	46.65	13.25	36.10	0.71	2.61	0.68
VCU	40.40	15.46	33.73	3.95	6.01	0.44
UVA	18.47	11.21	54.76	10.24	4.24	1.08
VT	27.63	13.76	40.76	11.40	5.81	0.63
VA Add-on V2	40.87	43.07	10.10	3.92	0.73	1.31

Figure 28. Mode split of university students and Virginia's general population

Note: Appropriate weights are applied when calculating the statistics for NHTS Virginia add-on data.

The two urban campuses (ODU and VCU) have higher drive alone trips—more than 40% of trips are single occupant vehicle trips. The percentages for shared-ride trips are similar among all universities (between 11% and 16%). Walking accounts for a large proportion of the mode split, and this percentage is higher in the two suburban campuses, where more than 40% of the trips are by walking. This difference could be due to a host of reasons that include different student population composition; different campus traffic management strategies, e.g. parking restrictions, walkability; and accessibility/proximity of campus buildings.

Table 13 shows the information acquisition and travel decision changes by different university student respondents. Similar to the survey for the general population in the Research Triangle, a large portion of the sampled student population (42%) did not acquire travel information. However, most of students (from 75% to 86%) who acquired traveler information reported that they changed their travel decisions based on the information they received.

Table 13. Information acquisition and travel decision changes

Source: (Son et al., 2011)

	Urban		Suburban	
	ODU	VCU	UVA	VT
Weekly traveler information acquisition	N=919 [2)]	N=652 [2)]	N=957 [2)]	N=935 [2)]
Never	42%	52%	69%	68%
At least once a week	33%	29%	19%	23%
2-4 times a week	13%	11%	7%	6%
5+ times a week	12%	8%	5%	3%
Traveler information sources [1)]	N=533	N=314	N=302	N=303
The Internet	74%	88%	95%	97%
Commercial radio	56%	46%	14%	11%
Television	51%	46%	19%	22%
Variable Message Signs (VMS)	37%	20%	10%	14%
Highway Advisory Radio (HAR)	24%	6%	2%	4%
Traveler Information Hotline (511)	10%	4%	4%	5%
Other	5%	5%	6%	8%
Travel decision change	N=533	N=314	N=302	N=303
Yes	86%	81%	78%	75%
No	14%	19%	22%	25%
Changes in travel decision [1)]	N=459	N=253	N=236	N=225
Route change	86%	77%	51%	51%
Departure time change	69%	65%	63%	68%
Mode change	7%	16%	46%	35%
Trip Cancellation	28%	22%	22%	22%

If comparing students from different regions, the students from urban campuses are more likely to acquire travel information than those from suburban campuses. This may be due to greater amounts of traffic congestion in urban areas and better availability of travel information provided by various sources. Figure 29 shows the different traveler information sources accessed by students from different universities. Noting that there are differences in the daily travel context across students from urban and suburban campuses (longer vs. shorter average commute distances and durations), behavioral responses to ATIS also vary across such campuses. Besides the Internet, other media such as radio, television, and VMS are still frequently used by students in urban campuses, whereas the Internet is almost the single dominant source in suburban campuses.

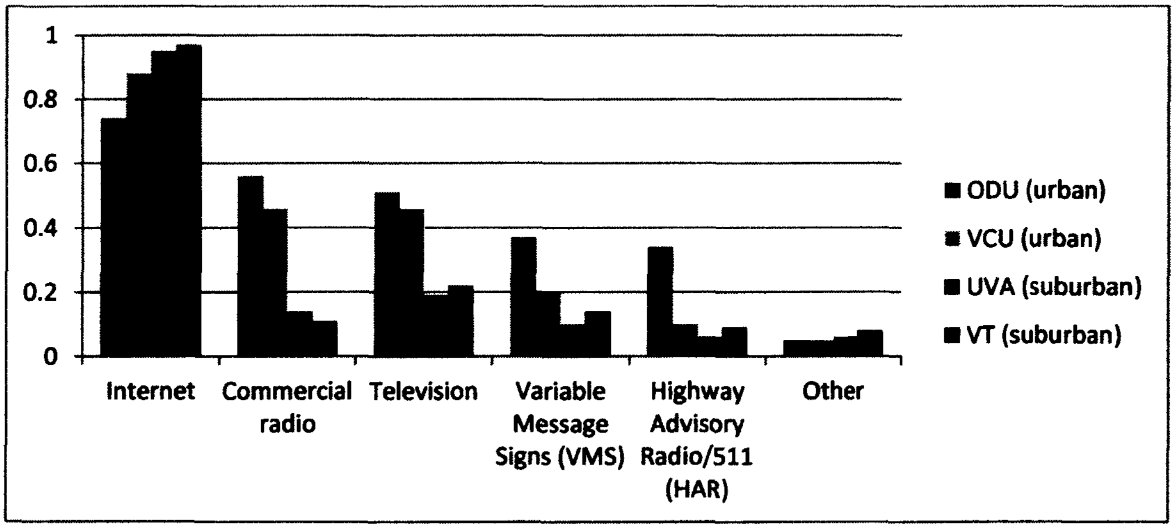


Figure 29. Traveler information sources accessed by students

Only a small portion of the students who acquired travel information did not change their travel decisions (14%-25%). Figure 30 shows how the sampled students responded to the information they received. Although students from urban campuses take the lead in all the ways of change, the difference between urban campus students and suburban campus student is more distinct when the changes are routes and departure time adjustments other than altering their travel modes or cancelling trips.

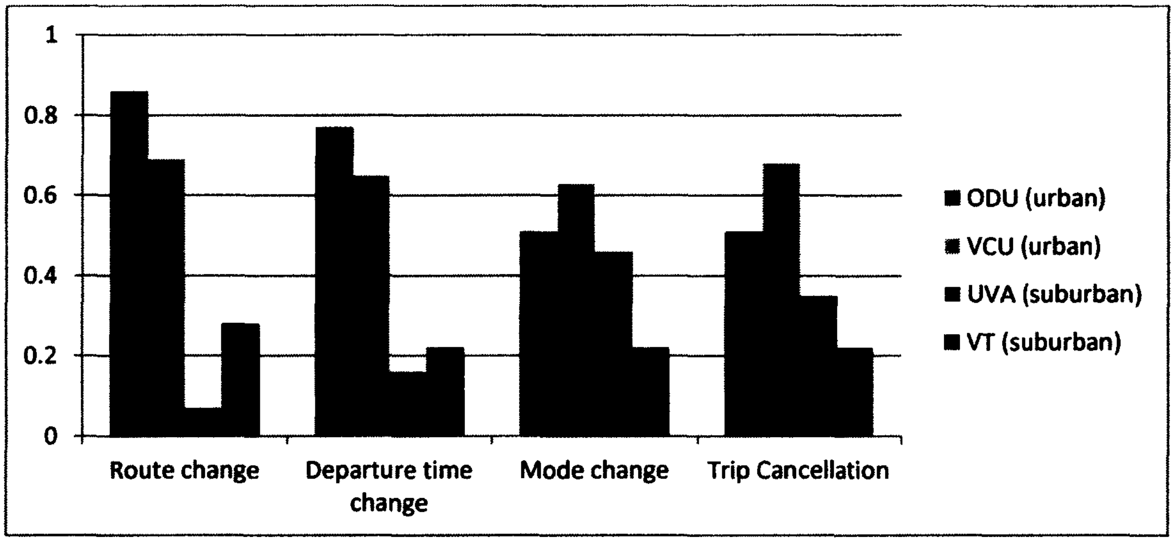


Figure 30. How students change their travel decisions

7.1.3 Comparison between the General Population and University Students

Table 14 summarizes the comparison on the personal characteristics and information acquisition behavior between the general population and the university students. Notably, the survey of the Research Triangle was conducted earlier in 2006, three years before the USTS was conducted. Considering substantial changes have been brought forward by emerging tele-communication technologies during past few years, it is understandable that the percentage of traveler information sources accessed by travelers has changed.

Table 14. Information acquisition and travel decision changes

Variable	The general population	University Student		
		Overall	Urban	Suburban
	N=5107	N=3463	N=1571	N=1892
INFO	51%	42%	54%	32%
CHANGE	0.40 (78% of 51%)	0.34 (80% of 42%)	0.45 (84% of 54%)	0.24 (76% Of 32%)
FINFO	1.96	2.06	2.21	1.93
INCOME	6.289	1.70	2.00	1.50
AGE	51.17	24.30	25.54	23.30
HHVEH	2.01			
INTERNET	23%	89%	79%	96%
RADIO	48%	30%	52%	12%
TEL	68%	33%	49%	21%
VMS	6%	20%	31%	13%
TIR/HAR(511)	2%	15% (6%)	25% (8%)	8% (5%)
OTHER	2%	6%	5%	7%
C_Route	87%	65%	83%	51%
C_DPTime	32%	66%	68%	66%
C_MODE	5%	27%	10%	39%
C_CAN	13%	24%	26%	22%

Overall, the average percent of student respondents who acquired traveler information is slightly lower than the general population (42% vs 51%). However, the percent of student respondents who adjusted their travel decisions based on information received are almost evenly matched with the general population. Considering that the Greater Triangle area is a metropolitan area, both the information acquisition and travel adjustment behavior of the general population in this region are closer to student respondents from an urban campus. In addition, substantial differences exist between university students and the general population in terms of the sources of information they acquired and how they adjust their travel decisions. Figure 31 and Figure 32 show these discrepancies.

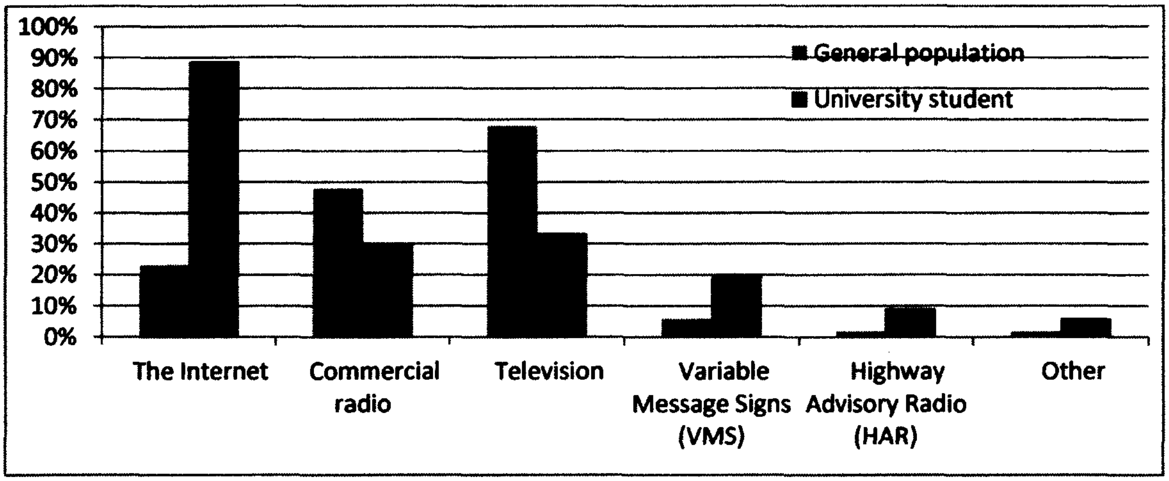


Figure 31. Comparison of what ATIS sources accessed

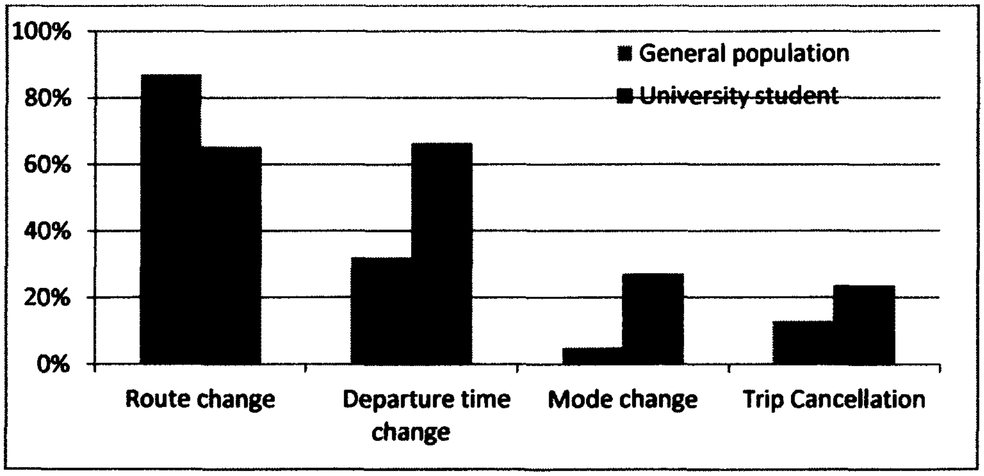


Figure 32. Comparison of how to change travel decision

The general population depends on more traditional information sources, such as commercial radio broadcasts and television. The percents of the general population who accessed information via television and commercial radio are 68% and 48%, respectively. These numbers for students are only 33% and 30%, respectively. Also, Internet, as an increasingly effective way of disseminating traffic information, is more widely used by university students compared with the general population. Besides the technology innovation in the past few years, this may also be due to different Internet coverage level, i.e. university students are 100% covered by the internet and they get used to using Internet in their daily lives. In addition, university students tend to use more en-route information than the general population. Specifically, about 20% and 15% student respondents checked VMS and highway transportation information radio, respectively. These numbers for the general population are much less, i.e. only 6% and 2%, respectively.

Although the percent of reported changes based on the information received is almost identical for university students and the general population (78% vs. 80%), the difference is distinct in terms of how they changed their travel decisions. Most of the general population (close to 90%) made a route change, and only about 30% of the general population changed their departure time. This is not the case for university students, i.e. less (about 66%) student respondents reported they changed their routes, though this number is higher (83%) for students from urban campuses. However, a higher percent of students reported they changed the departure time (65%). Interestingly, few respondents from the general population changed trip model (less than 5%). The percent for university students is 27%; for suburban campus students, it is as high as 39%. The percent for urban campus students is relatively low since the Hampton Roads Area has very limited transit service. Similarly, only 13% of respondents from the general population cancel their trips while the number for student respondents is almost two times.

These differences somehow reflect the fact that students have greater use of alternative modes and their daily schedule is very different from the general population. The students have different time-to-travel compared with the general population, i.e., they participate in more activities during the mid-day and in the evening after 6:00 pm (Khattak et al., 2011). Therefore it may be easier for students to adjust their departure time since the congestion level during mid-day and evening would be less severe compared with peak hours.

7.2 Model Framework

7.2.1 Conceptual Framework

A conceptual framework is established to address factors contributing to information acquisition as well as factors associated with travel decision changes. The framework, as shown in Figure 33, emphasizes the conditional link between information access/acquisition and travel decision change. The framework acknowledges there is a two-stage decision making process: 1) information access/acquisition; and 2) travel decision-making change. The latter stage is conditioned by the first stage, and those two stages are in turn influenced by socio-demographics and urban traffic conditions. Specifically, the conditionality between the two stages can be conceptualized as: before travelers adjust their travel decisions, they need to: a) own and be able to acquire information; b) perceive an adverse transportation condition, e.g., uncertainty, which provides a motivation to acquire and use dynamic information; and c) use this information to make travel decision changes.

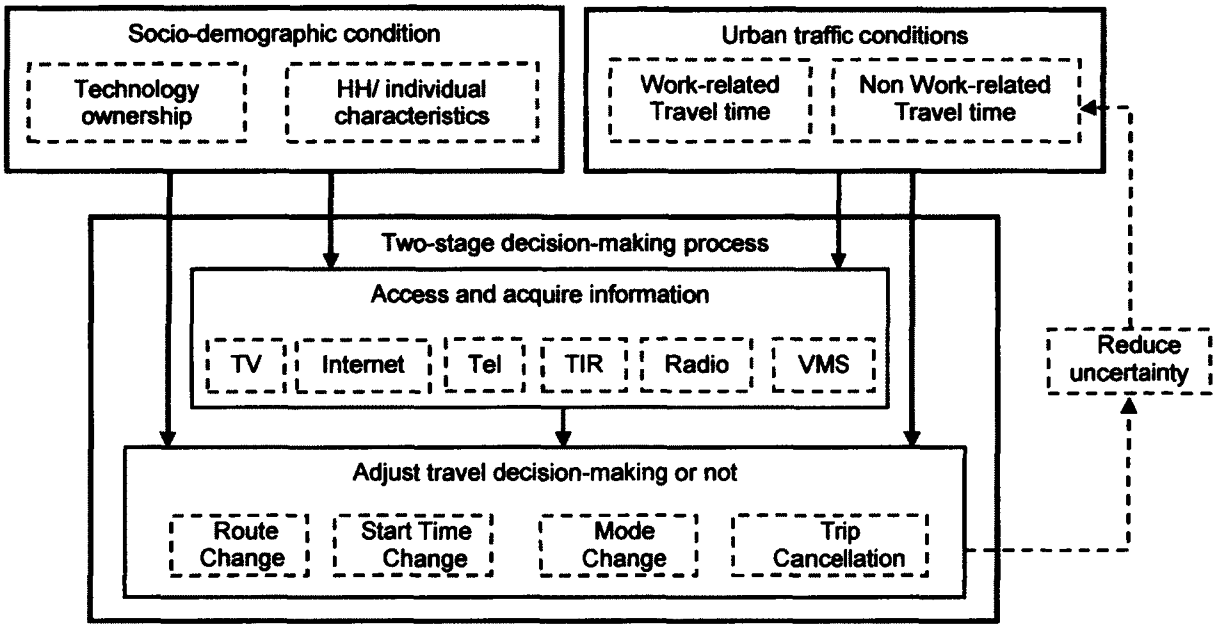
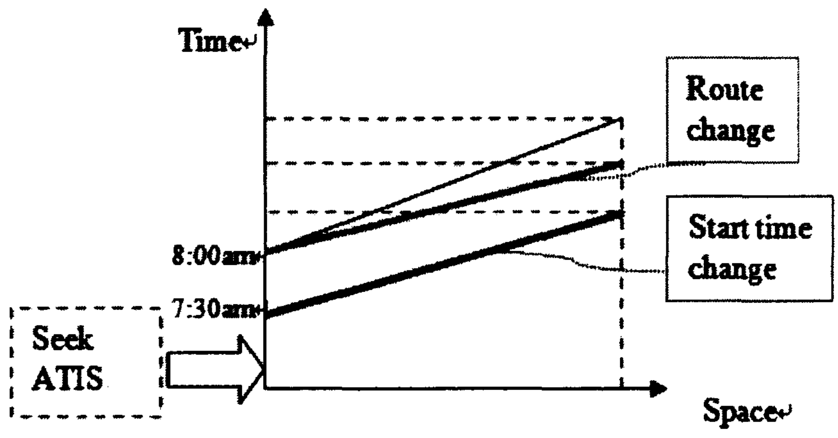


Figure 33. Conceptual framework showing information acquisition and travel change

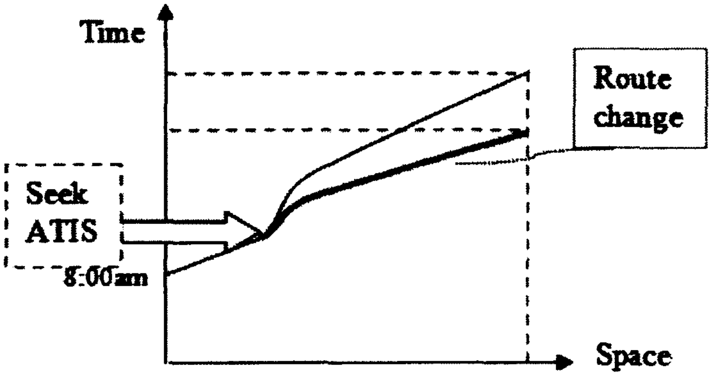
Socio-demographic factors used in this analysis include individual characteristics and household attributes such as income and technology ownership. Urban traffic condition is presented using personal travel time. This is because for individual travelers, their judgment for the urban and traffic conditions are most likely to come from their own experiences on their intended routes, especially for the routes they routinely take. These experiences can be proximately measured by self-reported travel time. We expect differences between work-related travel and non-work related travel time.

Hypotheses tested in this study are whether travelers are more likely to use travel information if they have longer work related travel time, longer non-work related travel time, higher household income, and more household vehicles. Also, younger respondents are expected to be more likely to use travel information. The possibility of changing travel plans would be higher if more information sources are accessed (including internet, radio, television, variable message signs, traveler information radio), the frequency of information acquisition is higher, and work related travel time and non-work related travel time is longer. More years living at their current address may represent higher familiarity with the surroundings, which can possibly counteract information usage. However, in cases that the household members know there would be substantial uncertainty caused by traffic congestion or incidents, they may still prefer to seek travel information. Therefore, impact of residential tenure is expected to be mixed.

The need to access and acquire information and adjust decisions exists before and during the trip. Figure 34 shows the logical thread of the hypotheses.



a) Pre-trip information usage



b) Pre-trip information usage

Figure 34. Pre-trip information usage vs. en-route information usage

Figures 34 (a) and (b), respectively, illustrate two types of travel information acquisition: pre-trip and en-route. Both types of information acquisition are largely aimed to save travel time, but they are different in terms of when travel information is accessed: in the pre-trip stage (Figure 34 (a)), a traveler acquires travel information before leaving and may find out about the presence of congestion on the intended route; they can decide to avoid the congestion by rescheduling departure time or changing route or mode. In the en-route stage (Figure 34 (b)), traveler acquires travel information during the trip and may adjust the route to avoid possible delay during the remaining journey. The X-axis in the figures represents the spatial movement of the traveler, while the Y-axis shows a traveler's location in time. The figures show changes in travel time when the traveler acquires the information and makes the change. Generally, information access and acquisition costs time and resources. However, acquiring travel information can inform individuals if the estimated travel time exceeds the desired travel time, giving travelers impetus to reschedule departure times or take alternative routes. Travelers have an incentive to seek travel information in order to reduce travel time uncertainty and to obtain certain benefits, e.g., travel time savings and knowledge of unexpected conditions.

7.2.2 Sample Selection Model

In order to capture both of these stages and explore the difference between factors that are associated with whether or not respondents seek travel information and travel change, a two-stage decision process is constructed. Since both stages are binary choices, i.e., seek information and change behavior, two probit models can be estimated. Clearly, behavioral changes are conditional on information acquisition, and this conditionality is captured appropriately in the sample selection model.

The sample selection model handles a fundamental question: Given that the respondents use transportation information, will they change their trip plan or not. Only if the respondents answered yes to the first question—whether or not he or she used transportation information—is that respondent included in the sample for the second question—change travel plan or not. Therefore, when info = 1, the respondent uses information at least once per week; while info = 0 indicates that the respondent never used travel information. When info = 0, change travel plan or not is not observed.

In the model, y is the dependent variable (change or not), which is observed only if a criterion, $z = 1$ is met (travel information is acquired). Specifically, if a traveler accesses/acquires travel information, then the change of traveler decision can be observed. The dependent variable y is related to independent variables $\boldsymbol{x}$ with the error term ε. z^* is unobserved and it can be estimated by independent variables $\boldsymbol{v}$, with the error term, $\boldsymbol{u}$.

$y = \beta'x + \varepsilon$ (change travel plan, binary probit) (Equation 10)

$z^* = \alpha'v + u$ (access/acquire info, binary probit) (Equation 11)

Here z^* is not observed and has an observed counterpart z, which is determined by:

$$z = 1 \text{ if } z^* > 0 \text{ and } z = 0 \text{ if } z^* \leq 0 \qquad \text{(Equation 12)}$$

$\boldsymbol{x}$ and $\boldsymbol{v}$ are two sets of explanatory variables that could contain either the same or different variables. In this research, work-related, non work-related travel time, age, etc. are common variables in both $\boldsymbol{x}$ and $\boldsymbol{v}$, but variables that capture information sources (e.g., radio and internet) are only included in $\boldsymbol{x}$. The model reports an index, ρ, to represent the correlation between the unobserved variables in the two equations. A statistically significant estimate for ρ indicates that modeling the change/adjustment decision and the access/acquisition decision simultaneously is superior to modeling them separately.

7.3 Model Result

7.3.1 Global Logistic Models

The traditional logistic models for information acquisition and behavior adaption are estimated separately by the maximum-likelihood algorithm and the results are presented in Table 15. The first model has traveler information acquisition as the dependent variable. The second model has change travel decision or not as the dependent variable. Both of these models are statistically significant overall.

Table 15. Global logistic model results

	INFO Model			CHANGE Model		
Independent Variable	β	T-stat.	Marg.	β	T-stat.	Marg.
Intercept	-0.133	-0.948	-	-3.337	-13.739*	-
WORK	0.004	4.028*	0.001	0.000	0.112	0.000
Dummy_W	-0.221	-2.983*	-0.055	-0.224	-1.922	-0.047
NWORK	0.002	3.813*	0.001	0.001	1.259	0.000
Dummy_NW	-0.010	-0.116	-0.003	-0.106	-0.760	-0.022
NA	-	-	-	1.700	20.014*	0.361
LIVED	0.003	0.295	0.001	0.054	3.571*	0.011
INCOME	0.053	4.932*	0.013	0.126	7.080*	0.027
HHVEH	0.112	3.441*	0.028	0.076	1.442	0.016
FINFO	-		-	0.532	20.701*	0.113
AGE	-0.011	-4.614*	-0.003	-0.021	-5.543*	-0.004
Summary Statistics						
Number of obs.	5107			5107		
Local sample size	1489			2526		
Log-likelihood	-3428.818			-1547.805		
AIC (Akaike Information Criterion)	6875.635			3117.609		

Note: * means significant at the 0.05 level.

Marg. means marginal effects of the variable at its mean value, with other variables controlled.

In the information acquisition model, in line with the expectation, the possibility of information acquisition is higher when the travel time is longer. Marginal effects show that one hour increase in work related travel time or non-work related travel time is associated with 6% higher possibility of travel information acquisition when the work related travel time is equal to its mean value. The negative coefficients of the dummy variable of work related time indicates that if the work related travel time is not available, these households are less likely to use traveler information. Marginal effects also show that one additional vehicle in the household is associated with a 3% higher possibility of travel information usage for those households with average number of vehicles. If the respondent is ten years younger, then information acquisition possibility increases by 3% for those travelers with the average age (51 years). A ten thousand-dollar increase in household income would enhance the information usage possibility by 1.3% for those households with average income. The length of living at current address shows no significant association with information acquisition.

In the travel decision adaption model, as expected, people's inclination of travel decision changes is positively associated with more information sources that he or she can access, accessing travel information more frequently, younger age, and higher income. The marginal effects show that one additional attempt to acquire travel information per week is associated with higher change probability by 11.3% at the mean of the information acquisition frequency. A respondent that is 10 years younger will have 4% higher likelihood of travel decision change compared with the travelers of average age (51 years). The length of living at the address shows positive significant association with the likelihood of travel decision changes, which means on average, the longer the traveler lived at the address, the more likely they change their travel decisions. All travel time related variables, including work related and non-work related travel time, are not statistically significant in predicting travel decision change, while they are significant in information acquisition models. It seems that travel time is more closely associated with the acquisition of travel information rather than with changes of travel decisions for the Triangle area respondents.

The traditional logistic regression models explain the determinants of probability of travel information usage and travel decision changes from a global average point of view separately. For an information acquisition model, these analyses would be fine if the change or not change travel time data are completely at random. However, the decision to use information or not is made by the individual household. Thus, those who did not use information constitute a self-selected sample and not a random sample. It is likely some of the household with low income or no vehicle choose not to use travel information and this would account for their reluctance to change their travel plan. So, this model is likely to overestimate the chances of travel plan changes in the population. Therefore, there is need to account for the portion of sample on the non-using information households. For travel decision change model, the model could be underestimated due to sample selection error. The solution to this quandary is to use the Heckman selection model.

7.3.2 Sample Selection Models

The General Population Model

Three models are listed in Table 16. The first two are probit models with sample selection. In order to capture differences in decision changes, two models with different dependent variables are estimated. One has a dependent variable of change or not, considering all changes such as time, mode, route or cancelling trip. The second model considers route change only. The third is a probit model without sample selection.

Table 16. Probit model with sample selection of travel behavior changes

Dependent	Change (time, mode, route) or cancel trip Model 1		Change route Model 2		Change (time, mode, route) or cancel trip Model 3	
	β	Marg.	β	Marg.	β	Marg.
Finfo	0.0770***	0.0182	0.0593***	0.0170	0.3234***	0.1131
Work	-0.0234	-0.0055	-0.0397	-0.0114	0.0281	0.0098
Nonwrk	-0.0138	-0.0032	-0.0277	-0.0080	0.0496**	0.0173
Dummy_w	-0.0602	-0.0143	-0.0546	-0.0157	-0.1718***	-0.0596
Dummy_nw	-0.0892	-0.0217	-0.1178*	-0.0349	-0.0609	-0.0211
INTERNET	0.3294***	0.0680	0.1405***	0.0386	1.2910***	0.4813
RADIO	0.2102***	0.0470	0.3525**	0.0936	0.9754***	0.3625
TEL	0.0319	0.0075	0.0185	0.0053	0.8667***	0.3134
VMS	-0.1446	-0.0366	-0.0383	-0.0112	-0.4615***	-0.1405
TIR	0.3194*	0.0636	0.3734**	0.0909	0.8602***	0.3316
Lived	0.0270***	0.0064	0.0219***	0.0063	0.0391***	0.0137
Age	-0.0047**	-0.0011	-0.0056***	-0.0016	-0.0112***	-0.0039
Constant	0.8977***		0.8114***		-1.5006***	
Dependent	Info acquisition		Info acquisition			
Lived	-0.0006	-0.0003	-0.0010	-0.0004		
Work	0.1310***	0.0523	0.1345***	0.0537		
Nonwrk	0.0726***	0.0290	0.0716***	0.0286		
Dummy_w	-0.1221***	-0.0487	-0.1096**	-0.0437		
Dummy_nw	-0.0023	-0.0009	-0.0089	-0.0036		
Income	0.0440***	0.0176	0.0433***	0.0173		
Age	-0.0064***	-0.0025	-0.0066***	-0.0026		
Hhveh	0.0734***	0.0293	0.0880***	0.0351		
Constant	-0.1590*		-0.1770**			
ρ	-.938 Prob > χ^2=0.000***		-.963 Prob > χ^2=0.000***			
	Number of obs. = 5107 Censored obs.= 2523 Uncensored obs. = 2584 Wald Chi-square= 82.3 Log likelihood = -4668.151		Number of obs = 5107 Censored obs = 2523 Uncensored obs = 2584 Wald Chi-square = 112.58 Log likelihood = -4869.41		Number of obs = 5107 Pseudo R^2 = 0.5499 LR Chi-square = 3771.78 Log likelihood = -1543.83	

Note: * p<0.10; ** p<0.05; *** p<0.01.

Used average value of 51.17 to replace the missing data for Age variable; The unit for travel time is hour. The marginal effect listed in change model is for the probability of a positive outcome, Pr (change=1) at the mean of each independent variables; The marginal effect listed in info model is for the probability of the dependent variable of the selection model being observed, Pr(info=1), at the mean of each independent variable.

Chi-Square tests show that all models are statistically significant. To answer which model is suitable for information access/acquisition and decision change process, the results in Model 1 are compared with Model 3 (the same model as Model 1 but without sample selection). In Model 1, with sample selection, ρ can represent the correlation of the residuals in the selection and outcome equations, which is statistically significant. This indicates that the sample selection structure is preferable to the two separate probit models. Comparison also finds that all the predictors have a lower magnitude in Model 1. For instance, in Model 3, the frequency of travel information acquisition has a large coefficient (0.32), whereas in the sample selectivity model (Model 1), the coefficient decreases to 0.077. This is reasonable because the model without sample selection assumes that all the travelers without information usage would not change their travel decisions. This could cause underestimation of the probability of decision changes in the population.

The constant term is positive in Model 1 and Model 2 (with sample selection), but it is negative in Model 3 (without sample selection). This indicates that travelers generally are likely to adjust their travel decisions given that travel information is accessed and acquired. However, the model without sample selection overlooks this point by providing constant with wrong direction. Therefore, the model with sample selection is methodologically reasonable and sophisticated. Also it provides a more nuanced interpretation.

In the information usage model, in line with our expectation, the possibility of information acquisition is higher when the travel time is longer. The effect size of work related travel time is larger than that of non work related travel time. One hour increase in work related travel time is associated with higher possibility of travel information acquisition by 5%, while the same increase in non work related travel time increases the information acquisition possibility by only 3%.

Consistent with the earlier expectations, factors associated with higher information acquisition possibility are higher number of household vehicles, more household income, and younger head of the household. Marginal effects show that one additional vehicle in household is associated with a 3% higher possibility of travel information usage. If the respondent is ten years younger, then information acquisition possibility increases by 2.5%. A ten thousand-dollar increase in household income would enhance the information usage possibility by 1.8%.

In the change travel decision model, in line with our earlier expectations, the changes are more likely if the household acquires travel information more frequently. One additional attempt to acquire travel information per week would increase the change possibility by 1.8%. A respondent that is 10 years younger will have 1% higher likelihood of travel plan change. The length of living at the address, while showing no association with information acquisition, is significantly associated with the likelihood of travel plan change. Ten more years of living at the same address would be associated with higher likelihood of changes by 6.4%.

Inconsistent with the expectations, all coefficients of travel time variables, including work related and non-work related travel time, are not statistically significant in predicting travel plan change, while most of them are significant in information acquisition models. It seems for the Triangle area respondents, travel time is more closely associated with the acquisition of travel information other than being associated with travel decision changes.

Among all information technologies, the Internet is associated with the highest propensity to change travel decisions, followed by radio. This study further found that television, VMS and highway traveler information radio are not statistically significantly associated with travel decision changes. This is interesting as statistics show that although television is the most widely used source to acquire travel information, it does not show an effective impact on travel decision change. The internet has a different result from television: while only 23% of respondents acquire travel information from the internet, it imposes the greatest influence on travel plan changes. The marginal effects show that the internet acquisition acts more strongly on the probability of changing travel plans than on the probability of changing routes (7% vs. 4% at the mean respectively). Radio and TIR provide the strongest effect on route changes, which is consistent with the literature. This may be due to more efficient information for en-route trips provided by radio. VMS is not statistically significantly associated with travel decision changes. This may be partially because of the low levels of VMS deployments in the Triangle area of North Carolina.

The findings reaffirm an advantage of sample selection model. That is, by estimating a two stage model, it can easily capture the different influences of factors on the two stages separately. More specifically, it answered the question whether travel time is a factor which influences the second stage of travel decision adjustment directly or backhandedly through affecting the first stage of information acquisition.

An interesting finding is that the length of living at the current address is insignificantly associated with the information usage but significantly and positively associated with travel decision changes. Longer residential tenure can represent higher familiarity with the surroundings, e.g., more knowledge of alternate routes, which may counteract information access/acquisition. For instance, a long-time resident may be more adept at making decisions when encountering unexpected congestion, reducing their need to seek travel information.

Student model

The sample selection model for university student is also estimated to compare it with the general population model (shown in Table 17). Similar variables are used in student models for the purpose of comparison. To capture the special environment of university campus, a dummy variable is added to differentiate an urban campus from a university town. Also, commute duration in the student model denotes how long it will take to drive between a residence and campus, similar to work-related travel time in the model of the general population. Two binary probit sample selection models are presented with/without statistically insignificant (at 5% level) variables. Chi-Square tests show that both models are statistically significant.

Table 17. Probit model with sample selection of student travel behavior changes

Source: (Son et al., 2011)

Dependent variables	Models / Independent variables	Model 1 (with all variables) β	Model 1 (with all variables) Marg.	Model 2 (removed insig. variables) β	Model 2 (removed insig. variables) Marg.
Travel plan change (1=yes, 0=no)	Commute Duration (hr)	-0.0048	-0.0074	-	-
	Acquisition Frequency	0.2613**	0.0674	0.2541***	0.0698
	Info Source: Internet	0.3436**	0.0990	0.2801**	0.0599
	Info Source: ComRadio	0.1685	0.0246	-	-
	Info Source: TV	-0.0066	-0.0018	-	-
	Info Source: VMS	0.4404**	0.1007	0.5255***	0.1311
	Info Source: HAR	0.1827	0.0439	-	-
	Info Source: 511	0.2348	0.0545	-	-
	Off-Campus x Part-time	0.3747***	0.0808	0.3207**	0.0958
	Urban Campus	0.2151	0.0338	-	-
	Constant	-0.608	-	0.2243	-
Travel information acquisition (1=yes, 0=no)	Commute Duration (hr)	0.1677***	0.0640	0.1626***	0.0641
	Off-Campus x Part-time	0.1198*	0.0498	0.1204*	0.0498
	Urban Campus	0.5421***	0.2076	0.5434***	0.2076
	Constant	-0.5292***	-	-0.5283***	-

Summary Statistics

	Model 1	Model 2
P	0.2259	-0.4575
LR χ2 of ρ	0.01	4.19
Prob.> χ2	0.9425	0.0407
N (observations)	3398	
N (censored observations)	1981	
N (uncensored observations)	1417	
Log-likelihood at convergence	-2858.552	-2862.0839
Comparison Log-likelihood	-2858.554	-2864.1785
Wald χ2	80.73	45.09
Prob.> χ2	0.0000	0.0000
ρ^2 (rho-squared) for separate models	0.04/0.07	0.04/0.06

NOTES: * p<0.10; ** p<0.05; *** p<0.01.

The marginal effects listed in the model are for the probability of a positive outcome, Pr (travel plan change or information acquisition=1) at the mean of each independent variables. Observations with missing values for any of the variables are excluded.

Consistent with the general population model and expectations, commuting time is positively associated with higher possibility of information acquisition instead of travel decision changes. The marginal effects show that one hour increase in commute time (to university) has a statistically significant association with a higher possibility of travel information acquisition by 6.4%. The frequency of receiving information is also positively associated with changing intended travel plans, similar to the general population model. One additional attempt to acquire travel information per week is associated with 6.7% higher chances of travel decision adjustment, controlling for other variables.

Different groups of university students show different possibilities of traveler information acquisition. Off-campus and part-time students are more likely to acquire travel information, and meanwhile, they have higher propensity to change their travel decision compared to other students. Their possibility to access traveler information is 5.0% greater than other subgroups, with all the other variables controlled. The indicator variable for campus type is also statistically significant, indicating that students in urban campuses are 20% more likely to acquire travel information compared with rural campuses. This implies that information seeking behavior is context-dependent, which may depend on different level of traffic congestion and availability of travel information. However, the campus location indicator does not show a statistical significance in the change model, implying that travel decision change behavior is not different across the campuses and may be transferable to other contexts.

For information sources, only Internet and VMS are statistically significant in the change model. The Internet consistently shows its strong influence on travelers' decision adjustment. However, different from the general population model, commercial radio and TIR no longer shows significant association with travel decision changes. Instead, VMS shows statistically significant association with travel decision changes for university student respondents in Virginia. This difference in terms of how various traveler information technologies impact travel decision change may be due to various factors, e.g. different level of information availability, traffic context and the preference of people.

7.3.3 Local Models

Although the sample selection model can compensate for the error brought by ignoring data censoring, the associations shown by sample selection model are still globally fixed. It cannot answer the question of whether the potential spatial heterogeneity exists. Geographical Weighted Logistic Regression (GWLR) model is then estimated to capture the possible spatial heterogeneity of both the information acquisition and travel decision changes for the general population of the Research Triangle region. The GWLR is calibrated using the GWR 3 package, and the calibration took 24 hours with more than 5000 samples, much longer than estimation of conventional logistic model.

The local estimation describes the situation at an individual location level. As there are 5107 households (regression locations) in the dataset, 5107 local models are estimated for INFO and CHANGE variables. Based on the local parameter estimates, a set of parameter surfaces are generated to reveal the spatial variations of these independent variables. An Inverse Distance Weighted (IDW) interpolation algorithm is used to assign values to unknown points based on the 5107 known household parameters, thus a continuous coefficient surface covering the whole region is generated. IDW assumes that each measured point has a local influence that diminishes with distance, with higher weights given for locations closer to the prediction location than those locations farther away.

Table 18 shows the parameter summary for GWLR model. It provides information about the extent of a parameter's spatial variance.

Table 18. Parameter summary for GWLR models (N=5107)

	Variable	Min.	Lwr Quartile	Upr Quartile	Maximum	Upr-Lwr	2*Std. Errors	(Upr-Lwr)>= 2*Std. Errors
INFO Model	Intrcept	-1.282	-0.443	0.397	0.839	0.840	0.28	TRUE
	Work	-0.002	0.001	0.004	0.013	0.003	0.002	TRUE
	Dummy_w	-0.717	-0.396	-0.138	0.170	0.258	0.148	TRUE
	Nonwork	-0.001	0.001	0.003	0.005	0.002	0.002	TRUE
	Dummy_nw	-0.487	-0.092	0.139	0.417	0.231	0.174	TRUE
	LIVED	-0.056	-0.003	0.039	0.058	0.042	0.018	TRUE
	INCOM	-0.067	0.009	0.052	0.153	0.043	0.022	TRUE
	HHVEH	-0.229	0.101	0.168	0.409	0.067	0.064	TRUE
	AGE1	-0.030	-0.014	-0.008	0.008	0.007	0.004	TRUE
CHANGE model	Intercept	-4.092	-3.545	-3.194	-2.804	0.352	0.486	FALSE
	Work	-0.003	-0.002	0.001	0.004	0.002	0.002	FALSE
	Dummy_w	-0.542	-0.255	-0.127	0.126	0.129	0.234	FALSE
	Nonwork	0.000	0.000	0.001	0.004	0.001	0.002	FALSE
	Dummy_nw	-0.560	-0.448	-0.020	0.724	0.429	0.28	TRUE
	NA	1.261	1.589	1.782	2.222	0.193	0.17	TRUE
	LIVED	-0.009	0.037	0.080	0.124	0.043	0.03	TRUE
	INCOM	0.048	0.100	0.132	0.160	0.032	0.036	FALSE
	HHVEH	-0.066	0.120	0.169	0.195	0.049	0.106	FALSE
	FINFO	0.388	0.531	0.583	0.623	0.052	0.052	TRUE
	AGE1	-0.032	-0.023	-0.017	-0.006	0.006	0.008	FALSE

Theoretically, all parameters can vary in space when GWLR is used (Fotheringham et al., 2002). Therefore it is important to determine if the spatial variance is significant enough to be captured by using the more complicated GWLR model. If the spatial variation is modest, it will be reasonable to use the global model known as stationarity in space. To decide whether the spatial variance is significant, the difference between the lower quartile and upper quartile of a parameter is compared with the standard error. If the difference is larger than two standard errors, then the parameter is non-stationary in space, indicating that spatial heterogeneity is statistically significant and misspecification exists in the global model (Fotheringham et al., 2002). From the results shown in the table, almost all explanatory variables show significant spatial variance in the INFO model. However, this is not the case for the CHANGE model.

Figure 35 demonstrates an example of how the coefficients distribute on a 3D map for better understanding of the local models. Income in the CHANGE model is shown in this figure. The x-y dimension shows the geographical shape of study area. The altitude represents the income coefficient (β) in the CHANGE model. Results of both models (local and global) are shown on the same map for easy comparison. Since income in the global model has only one value (β=0.126), it produces a flat surface with the altitude of 0.053 in the space. However, the coefficient of income in local model produces a continuously changing surface (β changes between 0.05 and 0.16), which gives a hilly landscape. Although β changes, the average value of β is still equal to 0.126, which is equal to the result of global model. Figures 36 and 37 present the generated parameter surfaces for key variables with cell size of 1 km by 1 km.

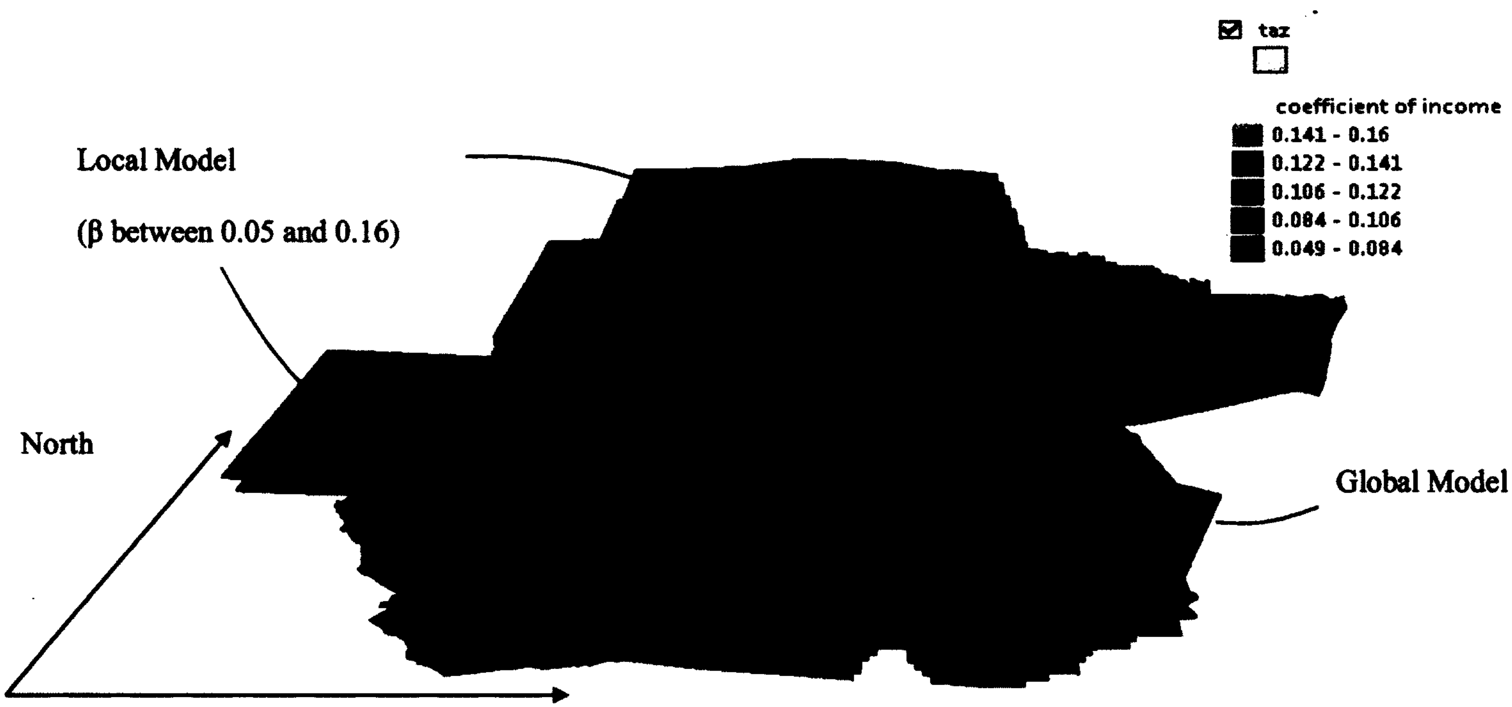
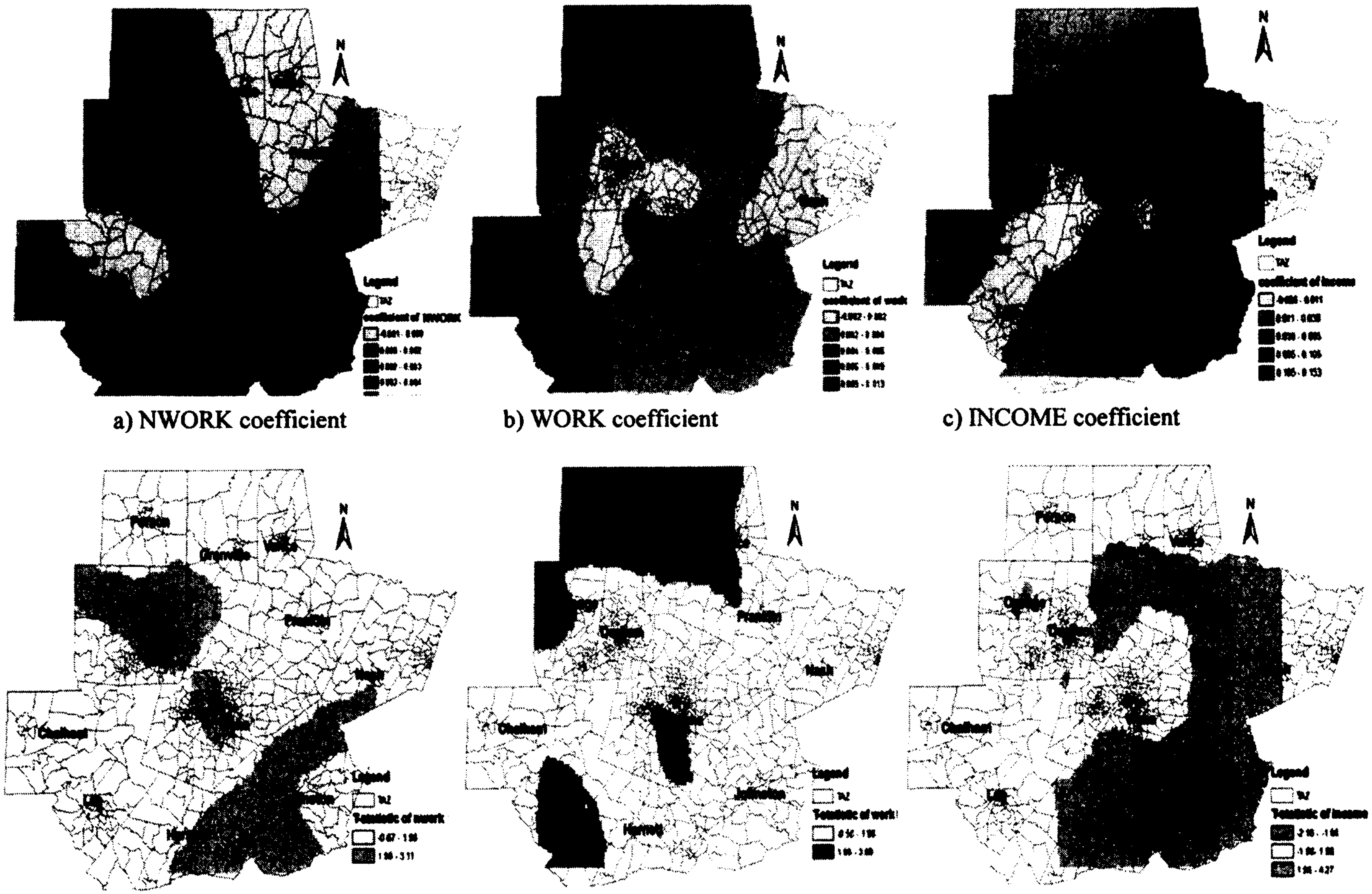


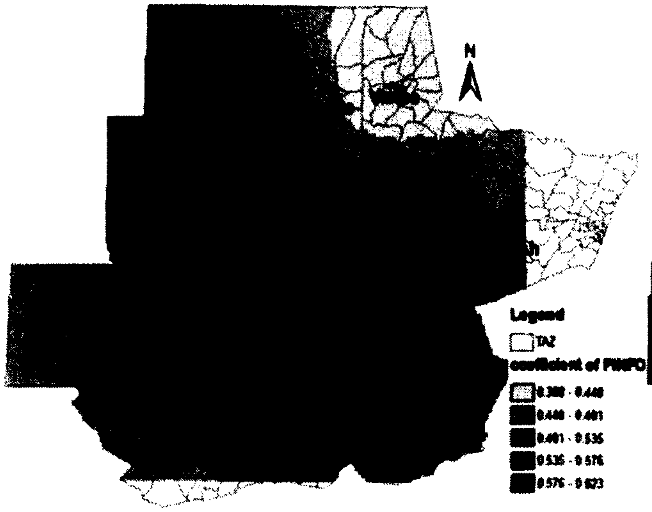
Figure 35. Global vs. local change model (variable: income)



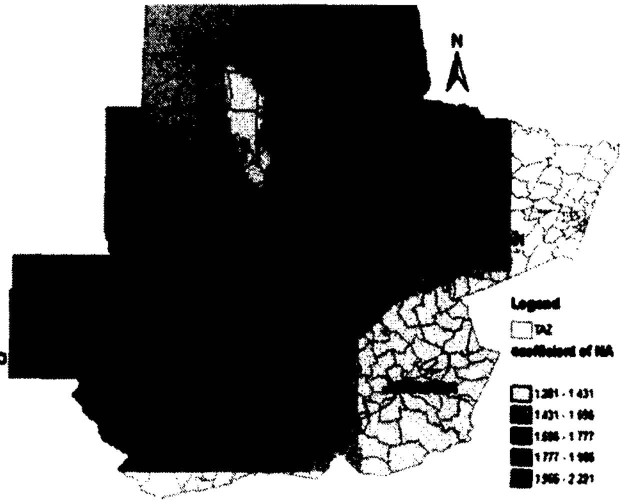
a) NWORK coefficient b) WORK coefficient c) INCOME coefficient

d) NWORK t-statistic e) WORK t-statistic f) INCOME t-statistic

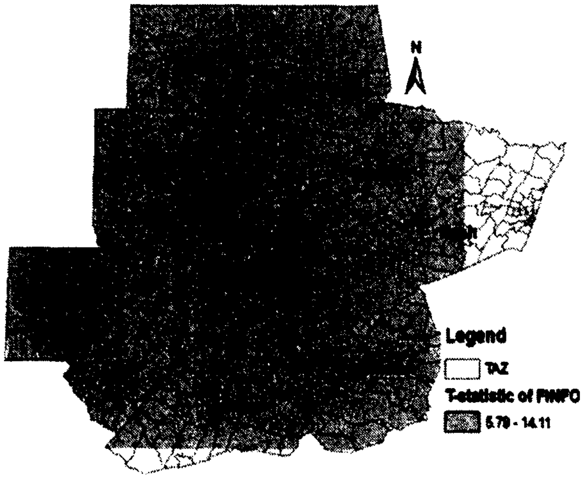
Figure 36. Local model results and t-statistic for INFO model



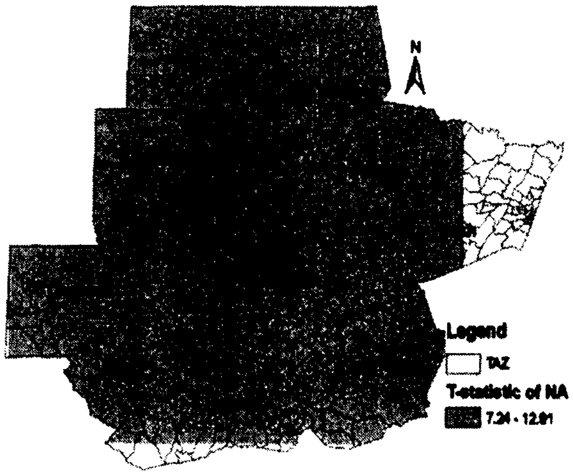
a) FINFO coefficient



b) NA coefficient



c) FINFO t-statistic



d) NA t-statistic

Figure 37. Local model results and t-statistic for CHANGE model

For the travel information acquisition model (Figure 36), although both work related and non-work related travel times are significant in the global model, these associations are not necessarily significant throughout the region. Moreover, non-work related travel time shows substantially different spatial distribution from that of work related travel time. For instance, the coefficient of work related travel time is only statistically significant in Person and Granville counties (Figure 36e); non-work related travel time is statistically significant in Orange, Durham and parts of Wake, Harnett, and Johnson counties (Figure 36d). It seems in urban areas, non-work related travel time tends to be significantly associated with information acquisition; but in suburban areas, work related travel time show significant association with information acquisition. For the coefficient of income, it has greater magnitude with information acquisition in the southeastern areas especially in Johnston County, Nash and Harnett County (Figure 36c). Although global coefficient shows 0.053 on average, while in these areas, it can be as high as 0.1, which is twice of the average level. In most other parts of the study area, the relationship between income and information acquisition is very weak, as shown by the pseudo t-statistic.

For the decision change model, Figure 37 shows the spatial distribution of coefficient and t-statistic of FINFO and NA. This shows that in those areas with darker colors, additional higher frequency of information acquisition or more traveler information resources available to a person are associated with higher possibility of travel decision changes. Overall, relationship between travel decision changes with the frequency of information usage and number of traveler information resources is positive. However, compared with suburban areas, frequency of information usage has higher marginal effects on probability of travel decision changes in urban areas, while number of traveler information resources has less marginal effects on travel decision adaption in urban areas.

7.3.4 Comparison between Global and Local Models

Goodness of fit measures for global vs. local models are summarized in Table 19. By comparing the AICs for the global and local models, it is clear that the local model has a smaller AIC than the global model (6722 vs. 6876 in the information acquisition model, 3105 vs. 3117 in the decision change model). Improvements in the AIC that are less than 3 in value could easily arise as a result of sampling error (Fotheringham et al., 2002), while here the difference between the global and local models is greater than 3, which indicates that the local models are statistically better than the global model. Furthermore, MAD (Mean Absolute Deviation) and RMSE (Root Mean Square Error) are calculated to compare the accuracy of the two models. The results show that GWR model has lower MAD and RMSE, which indicates that GWR provides better predictions compared with the global model.

Table 19. Goodness of fit measures for global vs. local models (N=5107)

	INFO model		CHANGE model	
	Global	Local	Global	Local
AIC (Akaike Information Criterion)	6876	6722	3117	3105
MAD (Mean Absolute Deviation)	0.488	0.126	0.202	0.065
RMSE (Root Mean Square Error)	0.505	0.154	0.329	0.114

7.4 Key Findings

This chapter examines how the general population and university students are different in terms of their behavior with regard to traveler information acquisition and travel decision adjustment in response to the traveler information received. A comprehensive regional dataset of the Research Triangle area (2006) and a behavior survey targeting university students in Virginia (2009) are compared with each other.

The comparison show that a large portion of the general population and the student population acquires traveler information; the percent of students from urban campuses is slightly higher than the general population, while the student from suburban campuses is much lower. Higher information acquisition percentages are observed on students from urban campuses (48% to 58%), compared to students from suburban campuses (31% to 32%). Furthermore, a large portion of students who access information (75% to 86%, depending on the university) is likely to change their decisions based on the information received. This percent is almost evenly matched with the general population.

Students and the general population have very different preferences over the information source they acquired and how they changed their travel decision based on information received. Internet is information source accessed by students mostly, while the general population depends on more traditional information sources, such as commercial radio broadcasts and television.s radio. Also, en-route information sources such as VMS and highway transportation information radio are more frequently used by students than by the general population. It is also clear that the Internet has emerged as a key source of information access for younger, more technology savvy students.

Consideration of investments in these technologies, along with concomitant investments in data collection and data fusion are needed, to allow easier access and plan readjustment. When it comes to travel decision adjustments, fewer students change their travel routes (this percent for students from suburban campus are even lower), but a higher percent of students change their departure time, mode or cancel the trip. These differences are partially coming from the innovation of information technology that emerged recently, considering the surveys were investigated in different year. Also, the information distribution service level is different by regions. More importantly, it may relate to different travel preferences among students and the general population, e.g. a student has a relatively flexible schedule.

In addition, a customized information delivery strategy for ATIS is desirable. Students in urban areas, where more route options exist, are likely to shift their routes requiring dynamic route information. Furthermore, students seem to alter their mode of travel relatively frequently, especially at suburban campuses, pointing to the delivery of multimodal information on and around campuses. Students need information that enables them to make mode choice decisions and information on public transit, bicycle and pedestrian facilities should be offered widely (e.g., on-campus and off-campus through multiple media) in order to facilitate readjustments of mode choices. This can result in a more dynamic readjustment of demand, potentially benefiting transportation network performance.

A key issue is whether appropriate model specification is used. Using rigorous statistical regression models, factors associated with traveler information acquisition behavior and propensity to change travel plans are explored. Three different models are presented, including traditional logistic model, the Probit model with/without sample selection and Geographical Weighted Regression model. Both of the previous two models are global models, where associations hold fixed across the study region. GWR is a local model, which can capture the spatial variance of the associations and show how the coefficients vary across space. However, the GWR model cannot deal with censored data, which is the case here given survey design. The sample selection model is a two-step modeling approach which can deal with censored data, but it cannot capture spatial variance.

A critical finding is that unobserved spatial heterogeneity exists in traveler information acquisition and travel decision adaption. This means that the associations between traveler information acquisition, travel decision changes and associated variables such as travel time and household income, varies significantly over the study area. These potentially useful insights, it should be noted, would have remained uncovered by the global models that assumed away heterogeneity. The results presented in this chapter support the application of GWR as an appropriate tool for providing insights into the spatial distribution of parameter estimates. The results highlight the importance of modeling local relationships when considering traveler decisions.

8. CONCLUSIONS

This chapter summarizes the findings in the dissertation, draws conclusions, discusses the policy implications, and points out future research directions.

8.1 Contributions

8.1.1 Methodology

From the perspective of the researchers and the policy makers, analytical techniques are a means to an end, since researchers always want to understand events and processes thoroughly by developing rigorous methodologies while policy makers want to benefit their strategic and tactical implementation from these methodologies.

Accuracy of geo-imputation

A key purpose of doing residential buffer analysis is to understand the associations of travel behavior, especially walking, bicycling, and public transit use, with micro-environment factors around residences. In order to explore associations empirically, researchers estimate statistically-based regression models using trips (walking, bicycling, and public transit) as dependent variables and the built environment/accessibility in buffers as correlates. Geo-imputation is presented in this dissertation to assign synthetic point based location to residences which can be used together with dynamic network based buffer to capture built environment more accurately. However, the synthetic assignment of residential locations also introduces potential systematic errors in the relationships modeled when smaller buffer sizes are used. Smaller buffer sizes of 0.25 are considered more appropriate for explaining walking in a neighborhood and for accessing public transit. The somewhat surprising result is that for larger buffer sizes (0.75 miles), the systematic errors are not substantial and the results indicate that synthetic assignment may be able to reproduce the relationship that may exist in reality for larger buffer sizes. Overall, geo-imputation is a more accurate method to assign location to residences compared with using centroid of a zone commonly used in current transportation literature. Geo-imputation can be used to replace exact geo-codes for transportation research purposes, if it is to be used in urban area or larger buffer sizes are to be used.

Capturing built environment

This study develops a unique database from behavioral data combined with a variety of spatial data. It applies GIS network analysis to capture built environmental characteristics in a dynamic network buffer around sampled residences. This method is more accurate since it can dynamically self-adapt its buffer size according to the accessibility and layout structure of the network around residences. More importantly, all the land use or public facilities covered by the network based buffer are actually accessible to the residences. However, this is the main drawback of the usual method of fixed buffer analysis.

University campus as a special trip generator

Large university campuses are major trip generators and can impact the regional traffic. The university-based travel demand model presented in this study can help improve the accuracy of regional models, especially in regions with larger student populations. Travel patterns for special generators (such as universities) are different from standard land uses and have received little attention in the conventional four-step travel demand forecasting literature. More broadly, the methodologies developed based on university student trip generation models set good examples to develop demand models for other sub-populations or special generators such as military bases and large hospitals. The findings about student travel can also help design practical strategies to improve the traffic conditions in and around the university campus by establishing satellite communities near the campus, e.g., providing better on-campus or near-campus student villages, encouraging traditional neighborhood developments within walking or bicycling distance from university campus (where feasible and appropriate), creating a pedestrian and bicycle friendly design on and near campus, adding public facilities in surrounding communities, and connecting regional transit corridors with university campuses.

Spatial heterogeneity

Timely contribution has been made by this study to understand there is spatial heterogeneity in models of traveler behavior and how they respond to traveler information. Instead of using the standard global models, which tend to compromise spatial heterogeneity in favor of average estimates and simplicities, attention in this dissertation focuses on capturing the spatial variations of links between trip-making and built environment, as well as associations between contextual factors with traveler information acquisition and travel decision changes.

Given the computing burden of estimating local models, the key question is whether or not the gains of simplicity and statistical efficiency of global models offset the losses from overlooking spatial heterogeneity. This study concludes that it is necessary to account for spatial heterogeneity in travel demand models and traveler information delivery mechanisms models, especially in metropolitan areas with substantial variations in geographical environments. The local models, e.g. GWR (Geographically Weighted Regression) and its relatives, GWPR and GWLR provide a more accurate representation of capturing the unobserved spatial heterogeneity in such models.

The direct use of the method is to map both the coefficient magnitudes and their statistical significances on a continuous surface, which provides an intuitive way to visualize spatial patterns of coefficients. This is distinct from commonly used (global) regression models which only reflect average associations; subsequently, it is not necessary to map global model results since it does not allow coefficients to change across space. Therefore, specific guidelines can be drawn based on these spatial patterns. Accounting for spatial variations in associations between built environment and trip making can help identify areas where focusing land use policies can have the highest impacts. Also, mapping the t-statistic can help to identify where certain policy have effective impact, given the factor is significant in that region.

The broader purpose of this method is to demonstrate that taking advantage of state-of-the-art developments of GIS technologies is important since valuable insights can be obtained by incorporating spatial information in travel behavior models.

8.1.2 Model Findings

Built environment and trip-making

For improved planning, understanding links between the built environment and travel is of great interest to researchers as well as practitioners. The mean and the variance of motorized trip frequency and their associated socio-demographic factors are found to be not identical across the Hampton Roads metropolitan area. Importantly, the association between household auto trips and built environment variables varies significantly over the Hampton Roads Area. This suggests that simply applying a pooled model without considering spatial heterogeneity can be misleading. Finding proper levels of spatial clusters, e.g. using neighborhood databases within a regional survey, should be considered carefully when exploring the associations of built environmental variables with travel demand.

The uniqueness of university campus

Rather than tinkering with traditional trip generation models, the study develops a deeper understanding of various factors associated with students travel behavior. Notably, traditional travel demand model considers trip rates with socio-economic information such as household size, automobile ownership, and household income, which does not appropriately represent the travel patterns of university students. The expected differences due to distance from campus and related factors are highlighted in the university students' travel demand model. The ring of mobility is confirmed by using spatial analysis. These differences in different ring areas around university campuses are the likely consequence of their different socio-demographics, accessibility to the built environment of the campus and the types of activities available on and near campus. Linkages are quantified between university students' residential location, demographics, and their travel patterns.

Interestingly, due to the concentration of campus buildings, university campuses usually represent more livable environments that are higher density, mixed use, and alternative mode friendly, offering greater access to activities in close proximity. Despite the fact that some campuses under investigation are located in an automobile dominant urban area, walking/bicycling still has a relatively larger share for students residing on-campus and near-campus. These findings are consistent with the findings from mixed-use traditional neighborhood developments that have shown less driving and more walking (Khattak and Rodriguez, 2005).

The evolvement of personal mobility is also confirmed. Shown in Figure 38, on average, university students made more non-motorized trips and less motorized trips compared with the general population. This difference reflects a tendency that as travelers age, they walk/bike less and become more dependent on car travel than they were in young age. It relates to their different life stages and the associated changed built environment.

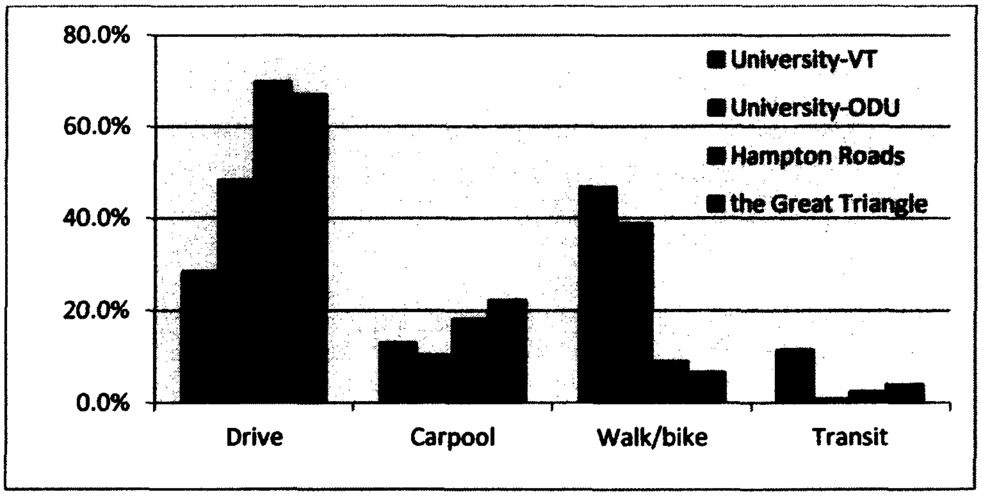


Figure 38. Mode split of university students and the general population

Travel decision changes based on traveler information

Overall, the contribution of this part is focused on the role of travel information access and acquisition in influencing travelers' propensity to "change course." Three issues relevant to traveler information delivery mechanisms have been highlighted in this dissertation.

Firstly, this research contributes by comprehensively exploring and quantifying travel information access and their propensity to change travel decisions in response to the received information. There is clear evidence that exposure to travel information is related to the higher likelihood of adjusting planned travel. Also, travel information acquisition is positively related to longer travel times or commuting duration. The uncertainty in longer trips may motivate respondents to acquire information. However, longer trip times do not seem to play a critical role in the change decisions. Furthermore, travellers are generally reluctant to change their routine travel decisions, partly due to behavioral inertia. However, this reluctance may not be due to their inherent resistance to change, but to the lack of travel information. This implies positive news for information service providers—information users are inclined to adjust their travel decisions and derive associated benefits.

Secondly, different modelling structures are used, highlighting different mechanisms during the process of traveller information delivery. The two-stage travel decision process sheds light on the two-stage links between information technology acquisition and travel decision processes, which is a lightly researched area in the literature. The local model GWR highlighted that there is spatial variations in both information acquisition and travel decision change process. Based on the spatial variance shown by local models, different resources may be focused on improving traveler information acquisition options in suburban areas by providing diverse traveler information. While in urban areas, resources may be invested to encourage seeking traveler information more frequently, by providing high quality real-time travel information on specific routes, e.g., congestion levels, incident locations and durations, delay and detour information. Finally, policies can be formulated that set targets for higher travel information access and use by underserved populations.

Thirdly, college students are used as comparison peer with the general population. Students as a young and lower income sub-population have a high proclivity to use alternative modes, nearly ubiquitous access to electronic media, and potentially high levels of use in the future, representing a younger segment of technology savvy early adopters of information technologies. The most used ATIS technologies are different for students compared with the general population, i.e. Internet and VMS for students and commercial radio for the general population. However, this may also be partially due to the emergence of new communication technologies given different survey years; therefore, this should be interpreted with caution. Also, transferability of information acquisition models is context-dependent; however, greater transferability of travel decision change is found given substantial differences in decision change behavior are not found across campuses.

8.2 Limitations and Future Research

8.2.1 Geo-imputation

This study examines the issue of whether geo-imputation can be used to replace actual geo-coded location for research purposes. A limited set of accessibility variables are used, i.e., only roadway length and number of transit stops in the buffer area are analyzed. Also, more research is needed to evaluate factors related to higher error levels of geo-imputation. For instance, if a residence is randomly assigned to a cul-de-sac (dangle point in roadway network) of a neighborhood, while it is located at the entrance of that neighborhood, then the calculated roadway length in its buffer area can be substantially lower compared with the actual situation. This means that synthetic assignments may not be able to reproduce the (average) relationship that may exist in reality for smaller buffer sizes. It will be interesting to explore whether using built environment accessibility variables obtained from geo-imputation can improve behavioral models, e.g., trip generation models.

Methods can be used to increase the accuracy of geo-imputation further, e.g., by adding more constraints such as residential density and other demographic and land use constraints. However, the feasibility of such analysis will depend on availability of demographic and GIS data at detailed geographic levels, e.g., at the level of parcel. Another idea that can be explored is to use the socio-demographic information in the NHTS samples to assign them randomly within a specific area or neighborhood, e.g., higher income samples will be assigned to higher income neighborhoods and vice versa. Again, this will require detailed spatial information about demographics. To conduct more detailed analysis of the NHTS data, non-residential destination locations should be analyzed also. However, it is difficult to synthetically assign non-residential destinations and do substantive analysis of activity participation without knowing the actual destinations.

8.2.2 Model Issue

A key issue is whether appropriate model specification is used. The study is limited by use of cross-sectional data from universities in a single state.

Also, in ATIS usage and travel decision change model, simultaneity may exist between information acquisition and decision changes, which can cause the problem of endogeneity. However, this question cannot be answered by using a censored dataset, which is the case with this study. Moreover, the sample selection model and GWR are used to stress different modeling aspects, but they are mainly applied separately in this research. Future studies can incorporate sample selection into the logistic GWR model for investigating the conditional interactions between information usage and behavioral changes, which can make the best of both methods.

8.2.3 Transferability Issues

This dissertation uses the Virginia Add-on and the university surveys in Virginia, which are both implemented at a time of economic recession, higher unemployment levels, and higher gasoline prices. These macroscopic factors might influence travelers' daily travel, and also their response to traveler information. Therefore the conclusion from this time period may be not able to transfer to other time directly.

The transferability of the findings from the GWR model to other urban regions is somewhat limited. The reason is that the spatial relationships that GWR captures is only valid for the targeted study region, which is to say, that any GWR model is based on the spatial characteristics of this particular region and possibly cannot be transferred to any other region. Also, only residential locations are used for weights calculation in GWR since most of socio-economic factors used in models are household characteristics. It may be interesting to compare local models estimated based on other relevant locations, such as job location.

8.3 Closing Remarks

In general, this dissertation represents an attempt to combine spatial pattern analysis with the study of travel behavior and response to traveler information systematically. It is unique in several ways:

It has contributed by first demonstrating how to create new types of data through innovative spatial analysis with GIS support, e.g. geo-imputation and network based buffers. They can give more accurate representation of behavior or spatial contextual factors in space. Transportation researchers or related professionals can use geo-imputation to create synthetic geographical units which can help to avoid privacy issues while ensuring that the results are within the limit of acceptable accuracy. Also, a network based buffer is useful as it integrates network analysis with proximity analysis. This can enable researchers to capture the land use within certain driving distance accurately, instead of using Euclidean distance, which is simplified and not accurate to the real situation.

In addition, from the modeling perspective, spatial heterogeneity is indeed important to account for in transportation models. It also suggests that subsequent policy inferences drawn from global models may be poorly suited to many local settings. Therefore, the spatial patterns of coefficients drawn from the model can help develop specific policy guidelines to achieve certain goal, e.g. encourage travel decision changes in response to travel uncertainty. For instance, using the spatial distribution of ATIS access, the population can be segmented according to their demographics, attitudes, and their travel context. Private information providers can use this information to target their potential customers more precisely. It can also help transportation managers such as the Departments of Transportation identify areas that are underserved by ATIS and locations where travelers are more or less likely to adjust their decisions. Moreover, this method can be also used to improve travel demand modeling. For instance, current travel demand model involves using cross-tabulation table for different group of population. While based on local models, the unified cross-tabulation table for the study region can be break down or customized for TAZs, which provides better accuracy.

Moreover, it is extremely useful and important to bring the data process and spatial modeling together. The reason is spatial modeling, as a data-hungry method, requires location information, which is usually not provided by current travel behavior data. The data process presented in this study can then be used effectively to compensate for this regret without compromising the reasonable accuracy. These two perspectives together come into being a robust methodology.

Besides the methodological contributions, by offering travel behavior and response to traveler information analysis on different target group of population, this dissertation is able to give more explicit and more meaningful attention to understand their differences. Moreover, young university students are selected as a case study since they represent one of the most active segments of our population. Meanwhile, it illustrates that travel decision is inherently complicated, and the special group in the population deserves specific research efforts and policy implementation. It will help us to prepare for the future in a more sustainable and savvy manner with emerging traveler information systems.

LITERATURE CITED

Abdel-Aty, M. A., Kitamura, R. & Jovanis, P. P. 1997. Using Stated Preference Data for Studying the Effect of Advanced Traffic Information on Drivers' Route Choice. *Transportation Research Part C: Emerging Technologies,* 5, 39-50.

Adler, J. L. & Blue, V. J. 1998. Toward the Design of Intelligent Traveler Information Systems. *Transportation Research Part C: Emerging Technologies,* 6, 157-172.

Ali, K., Partridge, M. D. & Olfert, M. R. 2007. Can Geographically Weighted Regressions Improve Regional Analysis and Policy Making? *International Regional Science Review,* 30, 300-329.

Bandura, A. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory,* Englewood Cliffs, N.J., Prentice-Hall, Inc.

Bhat, C. & Zhao, H. 2002. The Spatial Analysis of Activity Stop Generation. *Transportation Research Part B,* 36, 557-575.

Bhat, C. R. & Guo, J. 2004. A Mixed Spatially Correlated Logit Model: Formulation and Application to Residential Choice Modeling. *Transportation Research Part B: Methodological,* 38, 147-168.

Boarnet, M. & Crane, R. 2001. *Travel by Design: the Influence of Urban Design on Travel,* New York, Oxford University Press.

Boarnet, M. G. & Sarmiento, S. 1998. Can Land-use Policy Really Affect Travel Behaviour? A Study of the Link between Non-work Travel and Land-use Characteristics. *Urban Studies,* 35, 1155-1169.

Brownson, R. C., Hoehner, C. M., Day, K., Forsyth, A. & Sallis, J. F. 2009. Measuring the Built Environment for Physical Activity: State of the Science. *American Journal of Preventive Medicine,* 36, S99-S123.e12.

Buliung, R. & Remmel, T. 2008. Open Source, Spatial Analysis, and Activity-travel Behaviour Research: Capabilities of the Aspace Package. *Journal of Geographical Systems,* 10, 191-216.

Buliung, R. N., ,P. S. Kanaroglou 2004. A GIS Toolkit for Exploring Geographies of Household Activity/travel Behavior. *Journal of Transport Geography,* 14, 35-51.

Buliung, R. N. & Kanaroglou, P. S. 2007. Activity-Travel Behaviour Research: Conceptual Issues, State of the Art, and Emerging Perspectives on Behavioural Analysis and Simulation Modelling. *Transport Reviews,* 27, 151-187.

Buliung , R. N., P. S. Kanaroglou 2006. Urban Form and Household Activity-Travel Behavior. *Growth and Change,* 37, 172-199.

Cambridge Systematics, I., Fienberg, S. E. & Love, T. M. T. P. 2009. Disclosure Avoidance Techniques to Improve ACS Data Availability for Transportation Planners. American Association of State Highway and Transportation Officials.

Cao, X. 2009. Disentangling the Influence of Neighborhood Type and Self-selection on Driving Behavior: an Application of Sample Selection Model. *Transportation,* 36, 207-222.

Chatman, D. G. 2005. *How the Built Environment Influences Non-Work Travel: Theoretical and Empirical Essays.* Dissertation for the degree of Doctor of Philosophy, University of California at Los Angeles.

Choo, S. & Mokhtarian, P. L. 2007. Telecommunications and Travel Demand and Supply: Aggregate Structural Equation Models for the US. *Transportation Research Part A: Policy and Practice,* 41, 4-18.

Chorus, C. G., Arentze, T. A. & Timmermans, H. J. P. 2007a. Information Impact on Quality of Multimodal Travel Choices: Conceptualizations and Empirical Analyses. *Transportation,* 34, 625-645.

Chorus, C. G., Molin, E. J. E., Arentze, T. A., Hoogendoorn, S. P., Timmermans, H. J. P. & Van Wee, B. 2007b. Validation of a Multimodal Travel Simulator with Travel Information Provision. *Transportation Research Part C,* 15, 191-207.

Chorus, C. G., Molin, E. J. E. & Van Wee, B. 2006a. Use and Effects of Advanced Traveller Information Services (ATIS): A Review of the Literature. *Transport Reviews: A Transnational Transdisciplinary Journal,* 26, 127 - 149.

Chorus, C. G., Molin, E. J. E. & Wee, B. V. 2006b. Travel Information as an Instrument to Change Car-drivers' Travel Choices: a Literature Review. *European Journal of Transport and Infrastructure Research,* 6, pp. 335-364.

Chow, L.-F., Zhao, F., Liu, X., Li, M.-T. & Ubaka, I. 2006. Transit Ridership Model Based on Geographically Weighted Regression. *Transportation Research Record,* 1972, 105-114.

Chow, W. M. 1976. A Study of Traffic Performance Models under Incident Conditions. *Transportation Research Record,* 567, 31-36.

Clark, S. D. 2007. Estimating Local Car Ownership Models. *Journal of Transport Geography,* 15, 184-197.

Cliff, A. & Ord, J. K. 1981. *Spatial Processes, Models and Applications*, Pion London.

Crane, R. 2000. The Influence of Urban Form on Travel: An Interpretive Review. *Journal of Planning Literature,* 15, 3-23.

Crane, R. & Crepeau, R. 1998. Does Neighborhood Design Influence Travel?: A Behavioral Analysis of Travel Diary and GIS Data. *Transportation Research Part D: Transport and Environment,* 3, 225-238.

Dia, H. & Panwai, S. 2010. Evaluation of Discrete Choice and Neural Network Approaches for Modelling Driver Compliance with Traffic Information. *Transportmetrica,* 6, 249 - 270.

Dill, J. 2003. Measuring Network Connectivity for Bicycling and Walking. *Paper presented at Joint Congress of ACSP-AESOP.* Leuven, Belgium.

Du, H. B., Corinne Mulley 2006. Relationship Between Transport Accessibility and Land Value: Local Model Approach with Geographically Weighted Regression. *Transportation Research Record,* 1977, 197-205.

Englisher, L. S., Juster, R. D., Bregman, S., Koses, D. G. & Wilson, A. P. 1996. User Perceptions of SmarTraveler Advanced Traveler Information System: Findings from Second-year Evaluation. *Transportation Research Record,* 63-69.

Ewing, R. & Cervero, R. 2001. Travel and the Built Environment: A Synthesis. *Transportation Research Record: Journal of the Transportation Research Board,* 1780, 87-114.

Ewing, R. & Cervero, R. 2010. Travel and the Built Environment: A Meta-Analysis. *Journal of the American Planning Association,* 76, 265-294.

Fan, Y. 2007. *The Built Environment, Activity Space, and Time Allocation: An activity-based framework for modeling the land use and travel connection.* Ph.D. dissertation, University of North Carolina at Chapel Hill.

Fan, Y. & Khattak, A. 2008a. Urban Form, Individual Spatial Footprints, and Travel: Examination of Space-Use Behavior. *Transportation Research Record: Journal of the Transportation Research Board,* 2082, 98-106.

Fan, Y. & Khattak, A. J. 2008b. Urban form, individual spatial footprints, and travel examination of space-use behavior. *Transportation Research Record*, 98-106.

Fan, Y. & Khattak, A. J. 2009. Does urban form matter in solo and joint activity engagement? *Landscape and Urban Planning,* 92, 199-209.

Fienberg, S. E., Steele, R. J. & Makov, U. Statistical Notions of Data Disclosure Avoidance and Their Relationship to Traditional Statistical Methodology: Data Swapping and Loglinear Models.

Fotheringham, A. S., Brunsdon, C. & Charlton, M. 2002. *Geographically Weighted Regression, the analysis of spatially varying relationships*, John Wiley & Sons, Ltd.

Fotheringham, A. S. & Wegener, M. 2000. *Spatial Models and GIS: New potential and New models,* London, Taylor & Francis.

Fotheringham, A. S. & Wong, D. W. S. 1991. The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. *Environment and Planning A,* 23, 1025-1044.

Fotheringham, S. A., Brundson, C. & Charlton, M. 2000. *Quantitative Geography: Perspectives on Spatial Data Analysis*, London: Sage.

Garrido, R. A. & Mahmassani, H. S. 2000. Forecasting Freight Transportation Demand with the Space-time Multinomial Probit Model. *Transportation Research Part B: Methodological,* 34, 403-418.

Gehlke, C. E. & Biehl, H. 1934. Certain Effects of Grouping upon the Size of the Correlation Coefficient in Census Tract Material. *Journal of the American Statistical Association, Supplement,* 29, 169-170.

Griffith, D. A. 1983. The Boundary Value Problem in Spatial Statistical Analysis. *Journal of Regional Science,* 23, 377-387.

Griffith, D. A. & Layne, L. J. 1999. *A Casebook for Spatial Statistical Data Analysis: A Compilation of Analyses of Different Thematic Data Sets,* New York, Oxford University Press.

Hadayeghi, A., Shalaby, A. S. & Persaud, B. N. 2003. Macrolevel Accident Prediction Models for Evaluating Safety of Urban Transportation Systems. *Transportation Research Record*, 87-95.

Handy, S. 2005. Critical Assessment of the Literature on the Relationships Among Transportation, Land Use, and Physical Activity

Handy, S., Cao, X. & Mokhtarian, P. 2005. Correlation or Causality between the Built Environment and Travel Behavior? Evidence from Northern California. *Transportation Research Part D: Transport and Environment,* 10, 427-444.

Hay, G., Kypri, K., Whigham, P. & Langley, J. 2009. Potential Biases due to Geocoding Error in Spatial Analyses of Official Data. *Health & Place,* 15, 562-567.

Henry, K. & Boscoe, F. 2008. Estimating the Accuracy of Geographical Imputation. *International Journal of Health Geographics,* 7, 3.

Hewko, J., Smoyer, K. E. & Tomic, M. J. H. 2002. Measuring Neighbourhood Spatial Accessibility to Urban Amenities: Does Aggregation Error Matter? *Environment and Planning A,* 34, 1185-1206.

Hillsman, E. & Rhoda, R. 1978. Errors in Measuring Distances from Populations to Services Centers. *Annals of Regional Science,* 12, 74-88.

Joh, K. 2009. *Unraveling the Complexity of Land Use and Travel Behavior Relationships: A Four-Part Quantitative Case Study of the South Bay Area of Los Angeles.* Dissertation.

Jones, S. G., Ashby, A. J., Momin, S. R. & Naidoo, A. 2010. Spatial Implications Associated with Using Euclidean Distance Measurements and Geographic Centroid Imputation in Health Care Research. *Health Services Research,* 45, 316-327.

Khattak, A., Polydoropoulou, A. & Ben-Akiva, M. 1996. Modeling Revealed and Stated Pretrip Travel Response to Advanced Traveler Information Systems. *Transportation Research Record,* 1537, 46-54.

Khattak, A. & Rodriguez, D. 2005. Travel Behavior in Neo-traditional Neighborhood Developments: A Case Study in USA. *Transportation Research Part A: Policy and Practice,* 39, 481-500.

Khattak, A., Targa, F. & Yim, Y. 2003a. Investigation of Traveler Information and Related Travel Behavior in the San Francisco Bay Area, Working Paper: UCB-ITS-PWP-2003-06. California Partners for Advanced Transit and Highways (PATH), Accessed Nov. 2011; http://repositories.cdlib.org/its/path/papers/UCB-ITS-PWP-2003-06.

Khattak, A., Wang, X., Son, S. & Agnello, P. 2011. University Student Travel in Virginia: Is it different from the General Population? . *Transportation Research Record: Journal of the Transportation Research Board, Washington, D.C.,* 2255, 137-145.

Khattak, A. J., Pan, X., Williams, B., Rouphail, N. & Fan, Y. 2008. Traveler Information Delivery Mechanisms: Impact on Consumer Behavior. *Transportation Research Record,* 2069, 77-84.

Khattak, A. J. & Wang, X. 2011. Virginia National Household Travel Survey: Geo-Coordinate Imputation. Virginia Department of Transportation, Transportation & Mobility Planning Division.

Khattak, A. J., Yim, Y. & Prokopy, L. S. 2003b. Willingness to Pay for Travel Information. *Transportation Research Part C: Emerging Technologies,* 11, 137-159.

Kim, H. & Kwan, M. 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.

Kitamura, R., Laidet, L. & Mar, F. 1994. The Evaluation Study of a Low-cost, Telephone-based Traveler Information System. PATH Research Report.

Kligerman, M., Sallis, J. F., Ryan, S., Frank, L. D. & Nader, P. R. 2007. Association of Neighborhood Design and Recreation Environment Variables with Physical Activity and Body Mass Index in Adolescents *American Journal of Health Promotion,* 21, 274–277.

Kockelman, K. 1997. Travel Behavior as Function of Accessibility, Land Use Mixing, and Land Use Balance: Evidence from San Francisco Bay Area. *Transportation Research Record: Journal of the Transportation Research Board,* 1607, 116-125.

Krenzke, T., Li, J., Freedman, M., Judkins, D., Hubble, D., Roisman, R. & Larsen, M. 2011. NCHRP Project 08-79 Final Report: Producing Transportation Data Products from the American Community Survey That Comply With Disclosure Rules. Transportation Research Board.

Krizek, K. J. 2003. Operationalizing Neighborhood Accessibility for Land Use-Travel Behavior Research and Regional Modeling. *Journal of Planning Education and Research,* 22, 270-287.

Kwigizile, V. 2007. *Incorporating Spatial Characteristics in Travel Demand Models.* Ph.D. dissertation, University of Nevada.

Kwigizile, V. & Teng, H. 2009. Comparison of methods for defining geographical connectivity for variables of trip generation models. *Journal Of Transportation Engineering,* 135, 454-466.

Lappin, J. & Bottom, J. 2001. Understanding and Predicting Traveler Response to Information: A Literature Review. Washington DC: Federal Highway Administration.

Levinson, D. 2003. The Value of Advanced Traveler Information Systems for Route Choice. *Transportation Research Part C: Emerging Technologies,* 11, 75-87.

Lloyd, C. L. 2007. *Local Models for Spatial Analysis*, Boca Raton, FL: CRC Press, Taylor & Francis Group.

Luo, L., Mclafferty, S. & Wang, F. 2010. Analyzing Spatial Aggregation Error in Statistical Models of Late-stage Cancer Risk: a Monte Carlo Simulation Approach. *International Journal of Health Geographics,* 9.

Mamun, M. S., Yin, Y. & Srinivasan, S. 2010. Trip Generation Characteristics of Special Generators. University of Florida.

Mannering, F., Kim, S.-G., Barfield, W. & Ng, L. 1994. Statistical Analysis of Commuters' Route, Mode, and Departure Time Flexibility. *Transportation Research Part C: Emerging Technologies,* 2, 35-47.

Mcnally, M. & Kulkarni, A. 1997. Assessment of Influence of Land Use-Transportation System on Travel Behavior. *Transportation Research Record: Journal of the Transportation Research Board,* 1607, 105-115.

Miller, H. J. 1999. Potential Contributions of Spatial Analysis to Geographic Information Systems for Transportation (GIS-T). *Geographical Analysis,* 31, 373-399.

Miller, H. J. 2003. What about People in Geographic Information Science. *Computers, Environment and Urban Systems,* 27, 447-453.

Miller, H. J. 2005. A Measurement Theory for Time Geography. *Geographical Analysis,* 37, 17-45.

Miller, H. J. & Shaw, S.-L. 2001. *Geographic Information Systems for Transportation: Principles and Applications*, Oxford University Press, USA

Moudon, A., Hess, P., Snyder, M. & Stanilov, K. 1997. Effects of Site Design on Pedestrian Travel in Mixed-Use, Medium-Density Environments. *Transportation Research Record: Journal of the Transportation Research Board,* 1578, 48-55.

O'sullivan, S. & Morrall, J. 1996. Walking Distances to and from Light Rail Transit Stations. *Transportation Research Record,* 1538, 19–26.

Oliver, L. N., Schuurman, N. & Hall, A. W. 2007. Comparing Circular and Network Buffers to Examine the Influence of Land Use on Walking for Leisure and Errands. *International Journal of Health Geographics,* 6.

Openshaw, S. & Taylor, P. 1979. A Million or so Correlation Coefficients: Three Experiments on the Modifiable Areal Unit Problem. *Statistical Methods in the Spatial Science. Edited by: Wrigley N. London,* 127-144.

Páez, A. & Scott, D. 2005. Spatial Statistics for Urban Analysis: A Review of Techniques with Examples. *GeoJournal,* 61, 53-67.

Páez, A., Uchida, T. & Miyamoto, K. 2002. A General Framework for Estimation and Inference of Geographically Weighted Regression Models: 1. Location-specific Kernel Bandwidths and a Test for Locational Heterogeneity. *Environment and Planning A,* 34, 733-754.

Parthasarathi, P., Hochmair, H. & Levinson, D. 2011. Network Structure and Spatial Structure. *TRB 90th Annual Meeting CD-ROM.* Washington, DC.

Parthasarathi, P. K. 2011. *Network Structure and Travel.* Doctor of Philosophy, UNIVERSITY OF MINNESOTA.

Peirce, S. & Lappin, J. 2003. *Acquisition of Traveler Information and Its Effects on Travel Choices: Evidence from a Seattle-Area Travel Diary Survey* [Online]. Available: http://www.itsdocs.fhwa.dot.gov/JPODOCS/REPTS_TE/13813.html, [Accessed 1 Nov. 2008] [Accessed].

Petrella, M. & Lappin, J. 2004. Comparative Analysis of Customer Response to Online Traffic Information in Two Cities: Los Angeles, California, and Seattle, Washington. *Transportation Research Record: Journal of the Transportation Research Board,* 1886, 10-17.

Polak, J. & Jones, P. 1993. Acquisition of Pre-trip Information: a Stated Preference Approach. *Transportation,* 20, 179-198.

Polydoropoulou, A., Ben-Akiva, M., Khattak, A. & Lauprete, G. 1996. Modeling Revealed and Stated En-Route Travel Response to Advanced Traveler Information Systems. *Transportation Research Record,* 1537, 38-45.

Rossi, F. & Conan-Guez, B. 2002. Multi-layer Perceptron on Interval Data. *Classification, Clustering, and Data Analysis,* 427–436.

Rushton, G., Armstrong, M. P., Gittler, J., Greene, B. R., Pavlik, C. E., West, M. M. & Zimmerman, D. L. 2006. Geocoding in Cancer Research: A Review. *American Journal of Preventive Medicine,* 30, S16-S24.

Schönfelder, S. & Axhausen, K. W. 2003. Activity Spaces: Measures of Social Exclusion? *Transport Policy,* 10, 273-286.

Schwanen, T., F. Dieleman, and M. Dijst 2001. Travel Behaviour in Dutch Monocentric and Policentric Urban Systems. *Journal of Transport Geography,* 9, 173-186.

Shawn Turner, A. H., Gordon Shunk 1997. Bicycle and Pedestrian Travel Demand Forecasting: Literature Review.

Shay, E. & Khattak, A. 2005. Automobile Ownership and Use in Neotraditional and Conventional Neighborhoods. *Transportation Research Record: Journal of the Transportation Research Board,* 1902, 18-25.

Son, S., Khattak, A. & Chen, J.-Y. 2011. A Comparative Analysis of University Students' Acquisition and Use of Travel Information. *Forthcoming in Transportation Research Record.*

Son, S., Khattak, A., Wang, X. & Chen, J.-Y. 2012. Transforming the Telephone-Based National Household Travel Survey to the Internet: Application to University Students. *Accepted in Transportation Research Record: Journal of the Transportation Research Board.*

Srinivasan, S. 2001. Quantifying Spatial Characteristics for Travel Behavior Models. *Transportation Research Record: Journal of the Transportation Research Board,* 1777, 1-15.

Srinivasan, S. 2002. Quantifying Spatial Characteristics of Cities. *Urban Studies,* 39, 2005-2028.

Stead, D. 2001. Relationships between Land Use, Socioeconomic Factors, and Travel Patterns in Britain. *Environment and Planning B: Planning and Design,* 28, 499-528.

Sultana, S. 2002. Job/Housing Imbalance and Commuting Time in the Atlanta Metropolitan Area: Exploration of Causes of Longer Commuting Time. *Urban Geography,* 23, 728-749.

Taylor, M. a. P., Woolley, J. E. & Zito, R. 2000. Integration of the Global Positioning System and Geographical Information Systems for Traffic Congestion Studies. *Transportation Research Part C: Emerging Technologies,* 8, 257-285.

Thill, J. & Kim, M. 2005. Trip Making, Induced Travel Demand, and Accessibility. *Journal of Geographical Systems,* 7, 229-248.

Toledo, T. & Beinhaker, R. 2006. Evaluation of the Potential Benefits of Advanced Traveler Information Systems. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations,* 10, 173 - 183.

Toppen, A., Jung, S., Shah, V. & Wunderlich, K. 2004. Toward a Strategy for Cost-Effective Deployment of Advanced Traveler Information Systems. *Transportation Research Record,* 1899, 27-34.

Van Wee, B. 2002. Land Use and Transport: Research and Policy Challenges. *Journal of Transport Geography,* 10, 259-271.

Veneri, P. 2010. Urban Polycentricity and the Costs of Commuting: Evidence from Italian Metropolitan Areas. *Growth and Change,* 41, 403-429.

Vovsha, P., Bradley, M. A. & Bowman, J. L. 2004. Activity-based travel forecasting models in the United States: Progress since 1995 and Prospects for the Future *presentation at the EIRASS Conference on Progress in Activity-Based Analysis.* Vaeshartelt Castle, Maastricht, The Netherlands.

Wang, X., Khattak, A. & Fan, Y. 2009. The Role of Dynamic Information in Supporting Travel Behavior Changes: A Two-stage Travel Decision Process. *Transportation Research Record,* 2138, 85-93.

Xie, F. & Levinson, D. 2007. Measuring the Structure of Road Networks. *Geographical Analysis,* 39, 336-356.

Yim, Y., Khattak, A. & Raw, J. 2002. Traveler Response to New Dynamic Information Sources: Analyzing Corridor and Areawide Behavioral Surveys. *Transportation Research Record: Journal of the Transportation Research Board,* 1803, 66-75.

Zayatz, L., Lucero, J., Massell, P. & Ramanayake, A. 2010. Disclosure Avoidance for Census 2010 and American Community Survey Five-year Tabular Data Products. *Section on Survey Research Methods – JSM 2010.*

Zhao, F., N. Park 2004. Using Geographically Weighted Regression Models to Estimate Annual Average Daily Traffic. *Transportation Research Record,* 1879, 99-107.

Zimmerman, D. L. 2008. Estimating the Intensity of a Spatial Point Process from Locations Coarsened by Incomplete Geocoding. *Biometrics,* 64, 262-270.

Zinszer, K., Jauvin, C., Verma, A., Bedard, L., Allard, R., Schwartzman, K., De Montigny, L., Charland, K. & Buckeridge, D. L. 2010. Residential Address Errors in Public Health Surveillance Data: A Description and Analysis of the Impact on Geocoding. *Spatial and Spatio-temporal Epidemiology,* 1, 163-168.

Zito, P., Amato, G., Amoroso, S. & Berrittella, M. 2011. The Effect of Advanced Traveller Information Systems on Public Transport Demand and Its Uncertainty. *Transportmetrica,* 7, 31 - 43.

VITA

Xin Wang is a native of Nanchong City, Sichuan province, China. She was born in 1979. She received her Bachelor's Degree in Engineering in 2000 from the College of Traffic and Transportation, Southwest Jiaotong University, Sichuan Province in China, and she received her Master's Degree in Engineering in 2003 from the Transportation College of Southeast University, Jiangsu Province, China.

Xin worked as a traffic engineer for the Jiangsu Institute of Urban Planning and Design from 2003 till 2007. That same year, August of 2007, she entered the Department of Civil and Environmental Engineering of Old Dominion University to pursue her Doctor of Philosophy degree. She published eight papers in refereed journals during the course of her Ph.D.