SPATIAL DYNAMICS OF DISAGGREGATED URBAN COMMUTING PATTERNS

by

WOO SUK JANG

(Under the Direction of XIAOBAI YAO)

ABSTRACT

Due to the rapid change of urban sprawl or suburbanization, commuting patterns have become more complex. Though commuting only accounts for about 25 percent of weekday trips, it is a major contribution to traffic congestion as well as a major driver of highway construction. Thus, the investigation of commuting patterns plays an important role in our understanding of various aspects of spatial separation between home and work. The overall aim of this research is to investigate the dynamics of the commuting patterns of racial/ethnic groups of workers in urban space over time, and to develop a new measurement of racial/ethnic segregation in the commuting space. The primary data used in this dissertation are from the Census Transportation Planning Package (CTPP). This decennial data series release data aggregated at a variety of geographic scales, for a number of different years. Two major problems occur for the spatiotemporal analysis of commuting patterns. First, trip data are not always available at the same spatial levels of census units (e.g., traffic analysis zone in 1990 and census tract 2000). Second, commuting trip data are not available in subgroups by demographic or socio-economic factors. To deal with these problems, this dissertation has the following three main objectives: (1) to develop approaches to aligning commuting flow data in different
geographic units to a common spatial framework; (2) to investigate spatiotemporal analysis of disaggregated commuting patterns by different population subgroups such as those categorized by race/ethnicity, gender, and income; (3) to develop a new measurement of racial/ethnic segregation based on trip data for the purpose of analyzing segregation as it occurs in spaces of travel to and from work. To achieve these objectives, the dissertation research proposed two new concepts, namely the flow line interpolation and segregation in the commuting space, and developed models for them. By applying the models, this dissertation took Atlanta as a case study area, analyzed the ethnically-divided urban commuting patterns, and tracked changes of the patterns over the last two decades. Furthermore, it explored Atlanta’s racial segregation in the commuting space. Through application of the methods developed in this dissertation, researchers can achieve a more in-depth understanding of the spatial dynamics of disaggregated urban commuting patterns between jobs and housing in any U.S. metropolitan area. The methods developed in this dissertation can also be applied to areas of study other than racial segregation, e.g., use of trip analysis of the ageing baby-boomer population to determine access to quality-of-life related amenities and resources, or emergency preparedness.

INDEX WORDS: Dasymetric mapping, Flow line interpolation, Spatiotemporal analysis, Urban commuting, CTPP, Segregation
SPATIAL DYNAMICS OF DISAGGREGATED URBAN COMMUTING PATTERNS

by

WOO SUK JANG
B.S., Minnesota State University, Mankato, 2002
M.S., Minnesota State University, Mankato, 2006

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2012
SPATIAL DYNAMICS OF DISAGGREGATED URBAN COMMUTING PATTERNS

by

WOO SUK JANG

Major Professor: Xiaobai Yao
Committee: Lan Mu
Marguerite Madden
Steven R. Holloway

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2012
DEDICATION

To my parents, Taeyoung Jang and Jungae Kim
ACKNOWLEDGEMENTS

During my Ph.D. program in the Department of Geography at the University of Georgia (UGA), I have benefited from the instruction, encouragement, and support of many people who have. First of all, my most sincere thanks go to Dr. Xiaobai Yao, my major professor, whose thoughtful comments, kind encouragement, and willing assistance helped bring the dissertation to a successful conclusion. Second, Dr. Lan Mu, Dr. Marguerite Madden, and Dr. Steve Holloway provided me with invaluable feedback and assistance over the course of this research as members of my dissertation committee. Third, I would like to express my sincere gratitude to colleagues, friends, and the staff of the Department of Geography at the University of Georgia. They have treated me like family, which made me feel at home during my doctoral study. Last but not least, special thanks are due to God for his endless guidance. Since emigrating from South Korea in 1998, He has been with me on every step of my journey towards the completion of my doctoral study here in the United States.
# TABLE OF CONTENTS

Page

<table>
<thead>
<tr>
<th>ACKNOWLEDGEMENTS</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>x</td>
</tr>
</tbody>
</table>

CHAPTER

1 INTRODUCTION .............................................................................................................1
   Research Background ..........................................................................................1
   Research Objectives .........................................................................................6
   Significance of Study .........................................................................................7
   Dissertation Organization ...............................................................................8
   References .........................................................................................................10

2 INTERPOLATING SPATIAL INTERACTION DATA .........................................................14
   Abstract ............................................................................................................15
   Introduction .......................................................................................................16
   Prior Related Work .........................................................................................18
   Research Design ...............................................................................................22
   Model Evaluation and Case Study .....................................................................30
   Conclusions .......................................................................................................35
   References .........................................................................................................37
LIST OF ACRONYMS ........................................................................................................108
LIST OF TABLES

Table 2-1: Interpolated commuting trip for census tract ..................................................27
Table 2-2: Interpolation accuracy analysis summary .............................................................33
Table 3-1: Changing ethnoracial compositions of workers in metropolitan Atlanta 1990 - 2008.........................................................................................................................51
Table 3-2: Commuting trips by race/ethnicity .................................................................57
Table 4-1: Racial compositions of workers in four counties in Atlanta, 2000 ..............88
Table 4-2: Isolation indices of home, work, and JTW ......................................................89
Table 4-3: Interaction indices of home, work, and JTW ..................................................90
Table 4-4: Sensitivity analysis of isolation index based on tract and small place........91
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Two examples of zoning schemes and flows – TAZs and census tracts ..........24</td>
</tr>
<tr>
<td>2-2</td>
<td>Illustration of proportions of units in the source zoning scheme overlapping</td>
</tr>
<tr>
<td></td>
<td>units in the target zoning schemes ..................................................26</td>
</tr>
<tr>
<td>2-3</td>
<td>Fulton county in Georgia .....................................................................31</td>
</tr>
<tr>
<td>3-1</td>
<td>Metropolitan Atlanta Area in Georgia ..................................................51</td>
</tr>
<tr>
<td>3-2</td>
<td>Population density at residence and workplace in 2000 ..........................59</td>
</tr>
<tr>
<td>3-3</td>
<td>Commuting patterns by race/ethnicity in the three time-periods .............60</td>
</tr>
<tr>
<td>3-4</td>
<td>Changes in commuting trips by ethnicity in Metropolitan Atlanta over the</td>
</tr>
<tr>
<td></td>
<td>time period 1990-2008 ........................................................................62</td>
</tr>
<tr>
<td>3-5</td>
<td>Average commuting distance by ethnicity ..............................................64</td>
</tr>
<tr>
<td>4-1</td>
<td>Four counties in Metro Atlanta, Georgia, 2000 .....................................82</td>
</tr>
<tr>
<td>4-2</td>
<td>Commuting space: a reference space in trip A from home to workplace (left)</td>
</tr>
<tr>
<td></td>
<td>and a control space with associated trip A (right) ..............................84</td>
</tr>
<tr>
<td>4-3</td>
<td>Extended flow matrix (7 * 7 example) ..................................................85</td>
</tr>
<tr>
<td>4-4</td>
<td>Trip isolation by race/ethnicity .........................................................92</td>
</tr>
<tr>
<td>4-5</td>
<td>Black trip isolation with residence and workplace .............................93</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION

RESEARCH BACKGROUND

Commuting time for American workers has increased over time (U.S. Census Bureau 2009). According to the Census, average travel time from home to work increased from 22.4 minutes in 1990 to 25.5 minutes in 2000, and declined only slightly to 25.1 minutes in 2009. Commuting has been a major contributor to traffic congestion and highway construction, even though it only accounts for about 25 percent of weekday trips (Jang and Yao 2011). Making use of an increasing body of commuting data, urban geographers and GIScientists have investigated commuting patterns in different metropolitan areas in order to understand various aspects of spatial separation between home and work. Previous research has typically focused upon excessive commuting (Horner and Murray 2002, Yang 2008), accessibility (Wang 2000), urban public transit (Pirie 1979, Sinha 2003, Yao 2007), mobility (Sanchez, Shen and Peng 2004), and spatial mismatches between jobs and housing for various population subgroups (Blumenberg 2004, Covington 2009). One of the most influential drivers of change in commuting patterns is the rapid growth of suburbanization. In particular, suburbanization has greatly contributed to the limitations faced by minority groups with respect to residential choices and job opportunities in the suburbs (Kain 1968). For instance, low-income people are often trapped in inner-city areas, leading to high levels of competition over a scarcity of
what are mostly low-wage jobs (Wilson 1997). Assessing commuting patterns of spatial mismatch helps urban geographers and transportation planners understand the wide differences in commuting behavior between different groups of people.

Previous studies of urban commuting have typically used the Census Transportation Planning Package (CTPP), which serves as the major data source for studies of spatial mismatches between jobs and housing. However, these studies were unable to undertake a spatiotemporal trajectory analysis of commuting patterns due to data inconsistency in the decennial CTPP data series and technical barriers in aligning commuting data on inconsistent spatial zonations. Furthermore, few studies to date managed to undertake detailed analysis of the commuting patterns of population sub-groups such as those by categories of gender, income level, or race/ethnicity. The combination of the two above issues resulted in absence of research in tracking spatio-temporal changes of ethnically-divided commuting patterns.

Albeit a great source of commuting information, the CTPP data present two major problems to those attempting to use it for analysis of the spatial dynamics of commuting over time at a disaggregated scale. First, the geographical units for which aggregated data are provided are not consistent over time as spatial structures of cities change with time. For example, the boundaries of Traffic Analysis Zones (TAZs) are different in 1990 and 2000. The same is true with other units such as Census Tracts. This creates problems when users attempt to analyze the changing commuting patterns over time. Some previous studies adopted interpolation techniques in an attempt to deal with this problem (Horner 2007, Li, Corcoran and Burke 2010). However, these studies applied the standard single-unit interpolation technique, in which each value was associated with one
spatial unit only. This approach is not consistent with the nature trip data which need to deal with a pair of spatial zones. Second, trip data are aggregated at various levels of spatial granularity for all working population in respective zones. For instance, there are no such data as white trips, black trips, or Latino trips. Some research investigates trips by subgroups (e.g., gender and occupation) for the purpose of determining and analyzing heterogeneous commuting patterns (Kim et al. 2012, O'Kelly and Lee 2005a, Sang, O’Kelly and Kwan 2010). Their approaches also aim at investigating commuting patterns based on a one-time snapshot. The synergy of the above two problems markedly limits any close-up examination of the changing commuting patterns of different population groups in a city over time.

When looking at spatiotemporal patterns in certain metropolitan areas, the inconsistency of geographic units will lead to the modifiable area unit problem (MAUP), which is a potential source of error that undermines the ability of spatial studies to utilize aggregate data sources (Unwin 1996). Making data comparable across time would obviously improve the quality of spatial-temporal analysis of commuting patterns. Spatial interpolation has been widely used to improve the spatial granularity of data, or to mediate between inconsistent zoning schemes of spatial data (Lam 1983). The objective of spatial interpolations is estimating data values (e.g., population) from one set of spatial units to another set of spatial units in the same area. Some researchers broadly classified spatial interpolations into simple interpolation (Flowerdew and Green 1992, Goodchild, Anselin and Deichmann 1993, Goodchild and Lam 1980, Lam 1983) and intelligent interpolation (Eicher and Brewer 2001, Jang and Yao 2011, Langford 2003). The basic assumption in simple interpolation is that data are uniformly distributed within each
source zone, while the intelligent interpolation techniques are freed from the assumption by using additional information, e.g., land cover or road density, to discern the heterogeneous spatial structure of source zones. Many studies have shown that simple interpolation provides less accurate estimates than intelligent interpolation (Eicher and Brewer 2001, Kim and Yao 2010, Langford 2003). Traditional interpolation, however, only deals with point or area analysis, and is thus difficult to apply to flow data. To tackle the aforementioned issues, it is necessary to develop a new method of interpolation which is capable of dealing with a pair of spatial zones when dealing with trip data (Jang and Yao 2011).

Much of the analysis of commuting patterns in the past has dealt only with homogeneous characteristics such as the total commuting flow of a given metropolitan area (Horner and Murray 2002, Ma and Banister 2006, Peng 1997, Sultana 2002). For example, the journey-to-work (JTW) data are only available for total workers in 1990, majority (white), and minority (non-white) in 2000. This aggregated approach is appropriate for predicting overall flow patterns in a metropolitan area to help transportation planners reconstruct or maintain highways in a certain zone in order to alleviate traffic. However, in reality, JTW varies markedly by categories such as gender, income-level, race and ethnicity. To deal with this problem, some researchers have developed a disaggregated approach which aims to extract the data of subgroups, rather than simply analyze the whole, thus making it possible to analyze the heterogeneous commuting patterns of individual workers (Kim et al. 2012, O'Kelly and Lee 2005a, Sang et al. 2010). For instance, data about work trips of low-income Latino workers would be of great value to research on spatial mismatch between their jobs and housing locations.
This is quite difficult to accomplish with the JTW data in its current form. Thus, future study to reveal commuting patterns disaggregated by ethnicity or race will allow a more realistic representation of job-housing relations.

Disaggregated data on commuting trips can also provide a more realistic evaluation of spatial separation (segregation) with regard to spaces of residence and employment, and the space between them. Much of the segregation index analysis has dealt only with the degree of segregation in spaces of residence and employment (Aslund and Skans 2010, Horner and Marion 2009, Leonard 1987, Massey and Denton 1987, Massey and Denton 1988, Morgan 1983). Massey and Denton (1988) identified five different dimensions of segregation, including evenness (the similarity of majority-minority ratios across neighborhoods), exposure (potential contact between two groups), concentration (minority group is restricted to certain areas), centralization (degree to which a group lives near the center of an urban area), and clustering (minority groups living close to one another). Two commonly-used indices are dissimilarity (evenness) and exposure. Many studies have shown that neither dissimilarity nor exposure indices are spatial, which means they are insensitive to the spatial distribution of sub-areas. To overcome this problem, numerous studies have attempted to adding the spatial component through measurements of adjacency (Morrill 1991), a boundary length element (Wong 1993), sub-area shape and size (Wong 1993), and distance decay (Morgan 1983). However, none of the prior studies considered evaluating segregation in the commuting space (coupled space of residence and workplace). In reality, it is possible that some minority groups have severe residential segregation but less employment segregation, or vice versa.
RESEARCH OBJECTIVES

The overall aim of this research is to investigate the dynamics of commuting patterns of racial/ethnic groups of workers in any urban space over time, and to measure ethnoracial segregation in the commuting space. In particular, racially dichotomous studies of spatial mismatch make important contributions to the understanding of commuting patterns in metropolitan areas, but typically limit their analysis to whites and blacks. As the two most recent censuses (2000 and 2010) indicated that Latinos have become the largest ethnic minority group as well as the fastest growing population in the U.S., research focusing exclusively on blacks and whites may misrepresent current dynamics and problems of urban commuting. Thus, this dissertation pays particular attention to the Latino population in comparison with other groups. To fill the gaps identified in previous research, the objectives of this dissertation are:

1. To develop approaches to interpolating commuting flow data between inconsistent spatial units (e.g., census tracts vs. traffic analysis zones),
2. To investigate spatiotemporal analysis of disaggregated commuting patterns by different population subgroups (e.g., ethnicity, gender, and income), with particular focus on the Latino population
3. To develop a new measure of segregation based on trip data and evaluate segregation of trips by ethnicity in the commuting space.
SIGNIFICANCE OF STUDY

This study is significant in several ways. First, the study is among the first efforts to develop approaches to interpolating flow data. Traditional areal interpolation methods translate values in which each is associated with one spatial unit. These traditional methods are unable to deal with flow data, as each of the latter is associated with a pair of origin and destination zones. Thus, this study takes an important step in matching and analyzing commuting datasets that are currently incomparable due to different zoning schemes. For instance, census tracts have different spatial partitions in 1990 and 2000. By applying interpolation of flow data, time series flow data can be produced in an identical spatial zoning scheme. Spatial analysis can then be performed to investigate the spatial-temporal changes of house-job relations in the study area. Furthermore, with the launch of the American Community Survey (ACS) program, there is an increasing availability of more frequent census data at coarser levels of spatial granularity, e.g., at the county level. Part of the ACS data includes trip-related information. This method may potentially maximize the use of such annually available data by interpolating it into data at higher levels of spatial granularity.

Secondly, the study is also significant for its contribution to our understanding of travel patterns by ethnicity over space and time. The primary principle of a two-step analysis is to interpolate trip data in different years to a common spatial zoning scheme, and to disaggregate trip data by categorical demographic or socio-economic factors. This two-step analysis of commuting patterns will shed light on our understanding of the changing job-housing dynamics in U.S. metro areas. By applying the two-step analysis, analysis of disaggregated work trips will give a more realistic representation of the
correlations between jobs and housing for the Latino population. Transportation planners can examine the specific areas in which flows are significantly different and then make decisions on how to adjust network infrastructures in the study area.

Thirdly, this research creates a new measurement of segregation based on trip data to assess the degree of segregation by ethnicity in the commuting space. Unlike the traditional segregation indices which deal with residential or employment space separately, this index will observe spaces where commuting patterns exist between residence and workplace. The commuting segregation index will give us a realistic evaluation of spatial separation where jobs and housing are ethnically divided. For instance, the commuting segregation index will allow us to determine whether some subgroups face segregation in job locations even though they are not residentially segregated. Thus, the new commuting segregation index will help to provide a more comprehensive perspective of segregation of both residence and workplace, as well as the spaces between them.

DISSEPTION ORGANIZATION

This dissertation consists of three research papers, with an introduction and a conclusion. Chapter 1 begins with a literature review of spatial interpolation/interaction and segregation related to commuting data, and discusses their limitations. Research objectives and the significance of study are explained, and finally, the organization of the dissertation is described.

Chapter 2 focuses on interpolating commuting data. This research develops an intelligent flow line interpolation method, which can help lessen the zoning aspect of the
modified area unit problem (MAUP) by generating flow data at various spatial aggregation levels. Three models are presented: areal-weighted, intelligent, and gravity-type flow line interpolation, and accuracy assessment by root mean square error (RMSE) is applied.

Chapter 3 examines commuting patterns by subgroups, e.g., race, gender, and income. Disaggregating commuting data by subgroups in this way makes possible the analysis of unique commuting patterns of various groups through space and time, a level of refinement which aggregated commuting flow data does not allow. Spatial changes in the commuting patterns of Latino workers are of particular interest in this study.

Chapter 4 develops a new measure of segregation based on trip data in order to evaluate the segregation of trips by ethnicity. The measurement principal is that the commuting flows between residence and workplace are treated as occurring in spaces in which segregation may occur. The index for trip data will give a more realistic evaluation of segregation beyond either residential or employment space, and help us understand how the commuting patterns differ by ethnicity over space and time.

Finally, Chapter 5 summarizes the three methodologies of flow interpolation, the two-step analysis of commuting patterns, and the commuting segregation index applied in chapters 2 through 4. It also summarizes the major findings of this dissertation, and describes some further research work, e.g., a close-up evaluation and dynamic visualization of commuting trips, and some potential applications of methods developed and used in this dissertation.
REFERENCES


U.S. Census Bureau. 2009.


CHAPTER 2

INTERPOLATING SPATIAL INTERACTION DATA

---

ABSTRACT

Spatial interpolation has been widely used to improve the spatial granularity of data, or to mediate between inconsistent zoning schemes of spatial data. Traditional areal interpolation methods translate values of source zones to those of target zones. These methods have difficulty in dealing with the flow data, as each instance of which is associated with a pair of zones. This study develops a new concept, flow line interpolation, to fill the above mentioned gap. We also develop a first flow line interpolation method to estimate commuting flow data between spatial units in a target zoning scheme based on such data in a source zoning scheme. Three models (i.e., areal-weighted, intelligent, and gravity-type flow line interpolation) are presented. To test the estimation accuracy and the application potential of these models, a case study of Fulton County in Georgia is conducted. The results reveal that both the areal-weighted and the intelligent models are very promising flow line interpolation methods. Keywords: spatial interpolation, spatial interaction, commuting flow, urban transportation
INTRODUCTION

Due to the fast growing process of suburbanization or urban sprawl, travel patterns become more complex with the increasingly multifaceted ways in which people interact with each other and with the environments. Trip data play an important role in our understanding of the spatial interactions in an urban area. Some of the trips are discretionary trips such as those made for shopping, recreational, and family or personal activities, while commuting trips refer to the journeys to work which are more stable in terms of frequency and routes. Even though commuting trips only account for about 25 percent of weekday trips, they make major contribution to traffic congestion and highway construction due to the concentrated travel time during the day. According to the U.S. Census (2008), most American workers spend more time commuting to work than in the past, with those driving private vehicles making up 75.5 percent of workers commuting to their workplaces. Average travel time to work increased from 22.4 minutes in 1990 to 25.5 minutes in 2008. Moreover, the number of people who start a commuting trip very early (5 am to 6 am) to avoid traffic jams on highways that typically occur during peak commuting hours (6 am – 10 am) has risen significantly from 6.4 percent in 1990 to 7.6 percent in 2000, and up to 8.6 percent in 2008.

The Census Transportation Planning Package (CTPP) data, published by the Department of Transportation (DOT), are special tabulations from the U.S. census data designed for transportation planners. It is published as a decennial dataset, along with the census. This data series has been the most popular and authoritative data source for studies of commuting patterns in metropolitan areas in the United States. This decennial data series releases data aggregated by geographical units such as census statistical units.
(e.g. census tracts) and Traffic Analysis Zones (TAZs). However, as the zoning schemes of census statistical units and the TAZs often change over time, the spatial units of the same data series become inconsistent. For example, TAZ boundaries shifted from 1990 to 2000. This creates problems for time-series studies of commuting patterns. To make the problem even more complicated, CTPP trip data are not always available at detailed levels of spatial aggregation. For some places, the county level is the smallest unit of analysis. This creates significant data barrier when researchers need commuting data at finer spatial granularity. Furthermore, the above discussed insufficiency and inconsistency with flow data may also expose studies to the modifiable area unit problem (MAUP). This problem was particularly cautioned for spatial studies that utilize aggregate data sources (Unwin 1996). Spatial interpolation has been widely used to improve the spatial granularity of data, or to mediate between inconsistent zoning schemes of spatial data. However, traditional areal interpolation methods only deal with point or areal data. These methods are not directly applicable to flow (line) data in which each basic element for interpolation involves a pair of points or areas. Our research aims to propose the concept of flow-data interpolation and to develop some first models of such.

This paper organized as follows. The next section reviews relevant prior research on spatial interpolation and spatial interaction. Section 3 defines the concept of flow line interpolation and presents some first models for it. Several software tools were developed for convenient application of the methods. Section 4 evaluates the models with a case study of commuting trips in Atlanta metropolitan area. The section also compares and
interprets the performances of each of the proposed models. The article concludes with summaries and discussions of future research avenues in the final section.

PRIOR RELATED WORK

Spatial interpolation and spatial interaction (SI) modeling are two research fields that are closely related to the topic of this study. Spatial interpolation translates data values from one zoning scheme to another (Lam 1983). Spatial interaction modeling concerns direct or abstract movements between origins and destinations. In urban transportation research, SI models have been widely used to estimate trips between home and workplace locations (Fotheringham and O'Kelly 1989, Horner 2002).

Spatial Interpolation

Spatial interpolation (SI) estimates the values of zones in a target zoning scheme according to those in a source zoning scheme. Some popular SI techniques include areal-weighted interpolation (Flowerdew and Green 1992, Sadahiro 1999), stochastic interpolation (Negreiros et al. 2010, Sadahiro 1999), and artificial intelligence methods (Jarvis, Stuart and Cooper 2003). According to whether additional informative data are used to assist the interpolation, SI techniques can be further categorized as simple interpolation (e.g., areal-weighted interpolation) and intelligent interpolation (e.g., dasymetric method) (Kim and Yao 2010, Mennis 2003). For a long time, simple interpolation methods such as point and areal-weighted interpolation methods have been the commonly used interpolation techniques (Burrough 1998, Flowerdew and Green 1992). Point interpolation derives data at new points based on a set of existing points with
known data. The areal-weighted interpolation method is the simplest and arguably most popular interpolation method for zone-based data (Flowerdew and Green 1992, Goodchild et al. 1993, Lam 1983). This interpolation method assumes that data are uniformly distributed in the source zones. It then applies a fraction of the data from the source zone to the target zone, in proportion to the intersection area between the target zone and the source zone. One of the major criticisms of current interpolation method focuses on their oversimplified assumption of uniform distribution of data within source zones. Further research efforts have been made to present alternative interpolation approaches that are free of such an assumption. A notable example is Tobler’s Pycnophylactic interpolation (Tobler 1979).

More recently, a new line of interpolation methods, collectively known as intelligent interpolation (e.g. Kim and Yao 2010), appear to be able to significantly improve estimation accuracy. These methods make use of additional (or ancillary) information that shed lights on the spatial structure of the data at issue. The best known of such interpolation is that uses dasymetric method or its variations. Other notable studies include Flowerdew and Green (1992)’s employment of EM algorithm that makes use of ancillary data to improve the accuracy of interpolation. Some examples of ancillary data are land use / land cover (Fisher and Langford 1995, Kim and Yao 2010, Langford 2003, Mennis 2003, Wu and Murray 2005) and street network (Reibel and Bufalino 2005). Rather than using arbitrarily defined administrative or statistical boundaries, the intelligent interpolation methods use ancillary datasets such as land use/land cover information to delineate areas of homogeneous values, thus creating areas that are much more realistic and accurate (Eicher and Brewer 2001, Fisher and Langford...
Many researchers suggest that dasymetric method or more generally intelligent interpolation can produce significantly more accurate results (Fisher and Langford 1995, Langford 2003).

Spatial interpolation is often used as a solution for some problems associated with the well-known modifiable unit problem (MAUP). Opensaw (1984) demonstrates that because spatial zoning systems of geographic data can be arbitrarily designed and are modifiable, the results of the same type of analysis vary among different spatial zoning systems. This issue is referred to as the modifiable area unit problem (MAUP). The MAUP can be an outcome of either scale or zoning effects (Amrhein 1995, Opensaw 1984). Scale effect refers to the impact of the level of spatial aggregation. Geographic data that span across different scales or spatial resolutions may give rise to inconsistent results of analysis. The zoning effect is related to inconsistent findings resulting from differences in the spatial partitioning of zones, even though the scale of the zones remains constant. Spatial interpolation can be useful in dealing with both scale and zoning effects by interpolating data to identical spatial partitioning of zones (or zoning schemes).

Spatial Interaction Modeling

Spatial interaction refers to physical or abstract flows between origins and destinations. Spatial interaction models had been applied in various studies of migration, transportation planning, retail location modeling, and others (Fotheringham and O'Kelly (1989). Some best known spatial interaction models include the gravity model (Fotheringham and O'Kelly 1989), entropy maximization (O'Kelly 2010, Wilson 1967, Wilson 1971), bi-proportional adjustment (de Mesnard 2002), random utility models
(Fotheringham 1986), and the log-liner models (Willekens 1983). Fotheringham, Brunsdon et al (2000) categorized the models into four distinct stages. The first stage is the social physics, during which period the Newtonian gravity model, Lowry model (Lowry 1964), and extensions of the Lowry model were developed. The second stage is the statistical mechanics, during which the entropy model (Wilson 1967, Wilson 1971) appeared. The third stage is called the aspatial information processing during which discrete choice model (McFadden 1974) was proposed and well accepted. The final stage is termed spatial information processing, in which the representative models include the nested logit and competing destination model (Fotheringham et al. 2000, Roy and Thill 2003). They further argue that the first three types did not specifically deal with spatial issues by simply adopting the concepts from physics and statistics.

The major point of concern in spatial interaction modeling is the unit of information with regards to aggregation level. According to that, there are the aggregate versus disaggregate spatial interaction models. The first two types of the above mentioned stages are primarily aggregate models where data aggregated at certain spatial zones are modeled. The latter two types of stages deal with data at individual levels and therefore individual choices behaviors are modeled. These latest modeling efforts consider the likelihood of an alternative (spatial interaction) option in the true spatial choice set and incorporate variables regarding individual decisions, behavior, and spatial awareness. Many previous studies also highlighted the unique advantage of disaggregated modeling with regards to realistic representations of reality and the ability to model individual’s behaviors, among other characteristics (Horner and Murray 2002, O'Kelly and Lee 2005a).
However, as this study is about the redistribution of flows in different zoning schemes, the problem is by nature about data aggregated in zones. Therefore, we will review one of the most popular aggregated spatial interaction models as it is highly related to this study. The gravity model is the earliest and most widely applied approach to modeling spatial interactions. In the gravity model, the quantity of interaction (flow) is proportional to the propulsion power of the origin and the attraction power of the destination, while it is inversely proportional to the impedance (e.g., distance or a function of distance) between the origin and the destination. Thus, commuting flows of shorter distance anticipate higher magnitudes than those of longer distance. Usually the trip distance is the network distance between the centroid of the origin zone and that of the destination zone. For trips that start and end in the same zone, the distance is thus modeled as 0. These trips are called intrazonal trips. Traditional gravity models neglect intrazonal trips by assuming zero values for the intrazonal interactions. However, many researchers recognize the importance of including intrazonal interactions and suggested solutions. Frost et al. (1998) suggest that this problem could be solved by maintaining non-zero values for the actual minimum distance of each zone. To derive reasonable trip distance values for the intrazonal flows, O’Kelly and Lee (2005) proposed a method of estimating them (see Equations 9 and 10 for details) which has been widely adopted.

**RESEARCH DESIGN**

The purpose of this research is to present a conceptual discussion for the interpolation of flow lines and to develop first models for it. Each of the models has some unique considerations and assumptions to be evaluated and compared.
Definition. Flow line interpolation: a process of converting spatial flows between a pair of spatial units, namely the origin zone and the destination zone, from one zoning scheme to another.

For illustration purpose, we will use commuting trips as examples of spatial interaction data in discussions hereafter. Figure 2-1(a) shows flows in two example zoning schemes, one is the traffic analysis zone (TAZ) and the other zoning scheme is census tract. Spatial interactions can be represented either graphically or in a matrix form. Figure 2-1(b) shows commuting trips in a matrix form under each zoning scheme. In the matrix representation, the cell value at row \( i \) and column \( j \) refers to the magnitude of flow (e.g. number of trips) from zone \( i \) to zone \( j \). Figure 2-1(b) shows that between the five TAZs, there exist \( 5 \times 4 \) interzonal flows and \( 5 \) intrazonal flows. Similarly, in the census tract zonation, a total of \( 3 \times 2 \) interzonal flow and \( 3 \) intrazonal flows exist between the three zones. In this example, the flow volumes (numbers of trips) between pairs of TAZs are known and those between census tracts are unknown. We will illustrate how to use the proposed flow line interpolation methods to estimate flow volumes between tracts.
Model – 1. Areal-weighted Flow Line Interpolation

This proposed interpolation method is based on an assumption that for a known volume of flows between \( i \) and \( j \) (\( T_{ij} \)), their origins are uniformly distributed in zone \( i \) and destinations uniformly distributed in zone \( j \). As the assumption is comparable to that of the areal weighting interpolation for areal data, we name it accordingly. With this assumption, Equation (1) presents the model to calculate the flow amount between spatial units \( i \) and \( j \) in the target zoning scheme based on flow information in a source zoning scheme. The estimated amount is the summation of every flow volume in the source zonation weighted by the flow’s proportions of origin and destination areas falling in the spatial units \( i \) and \( j \) respectively.

\[
T^t_{ij} = \sum_{k=1}^{m} \sum_{h=1}^{n} [p_k \times p_h \times T^s_{kh}] 
\]

(1)
In the equations, $T_{ij}^{t}$ is the flow volume from unit $i$ to unit $j$ in the target zoning scheme, and $T_{kh}^{s}$ is the flow volume from $k^{th}$ zone to $h^{th}$ zone in the source zoning scheme. Variables $m$ and $n$ are the total numbers of origins and destination zones respectively in the source zoning scheme. $p_k$ is the proportion of the intersection area between zone $k$ in the source zonation and zone $i$ in the target zonation in relation to the total area of zone $k$. Similarly, $p_h$ is the proportion of intersection area of $h$ and $j$ in relation to the total area of zone $h$. In Equations 2 and 3, these proportions are formulated. $A_{kli}^{snt}$ indicates the intersection area of trip origin zone $k$ in the source zonation and trip origin zone $i$ in the target zonation. $A_k^s$ is the area of zone $k$. $A_{hj}^{snt}$ and $A_h^s$ are defined similarly for the trip destination zones.

Let’s use the example in Figure 2-1 to explain the meaning of the equations. In Figure 2-1, the trips between census tracts are unknown. We can estimate them using the proposed flow line interpolation method. We can start with the estimation of the flow volume between zone A and zone C. As illustrated in Figure 2-2, zone A consists of portions of several zones in the TAZ zonation, and so does zone C. Therefore the flow volume from A to C can be estimated by finding contributions from portions of TAZ flows whose origin zone intersects zone A and its destination zone intersects zone C. In computation, all flows are considered and their contribution is weighted by its portion of areas intersecting zone A or C. If a flow has an origin zone that does not overlap zone A or a destination zone that does not overlap zone C, the contribution is automatically 0 as 

\[
p_k = \frac{A_{kli}^{snt}}{A_k^s} \quad (2)
\]

\[
p_h = \frac{A_{hj}^{snt}}{A_h^s} \quad (3)
\]
the corresponding $p_k$ or $p_h$ would be 0. Figure 2-2 shows that TAZs E, F, and H intersect census tract A, while TAZs F, G, H, and I intersect zone C. The proportions of the overlapping areas in relation to the respective TAZs are computed and reported in Figure 2-2(b).

![Image](image1.png)

Figure 2-2. Illustration of proportions of units in the source zoning scheme overlapping units in the target zoning schemes

To implement the proposed method, several software tools have been developed in the study. Finding the intersection area can be done by topological analysis in a GIS environment. To automate the process, an ArcObjects script with Microsoft Visual Basic for Application (VBA) is developed to perform the tasks in ArcGIS. Another script is developed using VBA for MS Excel to automatically compile the calculated proportions from ArcGIS and to compute the contributions from each flow in the source zoning scheme. The results are stored in a txt file and then be imported into MS Access for final
flow data. In MS Access, Structure Query Language (SQL) scripting is used for multiple joining operations combining fields from multiple tables which contain necessary components for Equation 1. Final flow volumes are computed in MS Access following Equation 1. The result is shown in Table 2-1.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>4</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>B</td>
<td>6</td>
<td>85</td>
<td>56</td>
</tr>
<tr>
<td>C</td>
<td>8</td>
<td>12</td>
<td>17</td>
</tr>
</tbody>
</table>

Table 2-1. Interpolated commuting trip for census tract

*Model – 2: Intelligent Flow Line Interpolation*

The simple assumption of even distributions of trips in the origin and destination zones for the areal-weighted flow line interpolation can be criticized in similar ways as those to its area interpolation counterpart. Indeed, interactions (e.g. trips) may vary significantly within each zone in reality. In the case of trips, for instance, it is reasonable to assume that urban areas home higher traffic flows than forest and barren areas do. It makes perfect sense to distinguish between different types of land use and assigning trips accordingly. Thus, we propose to use ancillary information that can shed lights on the possible distribution of flows in this so-called intelligent flow line interpolation model. In the case of commuting trips, land use types or road densities are both good examples of useful ancillary information. In this illustration, we will use land use/land cover data in this discussion. First, based on classified remote sensing images, we further group land use/land cover types into built-up areas and non built-up area. Then equations (2) and (3)
are revised to replace the proportions based on total areas with the proportions based on built-up areas specifically. The revised equations are shown in Equations (4) and (5).

\[
\begin{align*}
\rho_k &= \frac{B_{k_i}^{s\cap t}}{B_k^s} \quad (4) \\
\rho_h &= \frac{B_{h_j}^{s\cap t}}{B_h^s} \quad (5)
\end{align*}
\]

in which \(B_{k_i}^{s\cap t}\) indicates the built-up area location within the intersection between trip origin zone \(k\) in the source zonation and trip origin zone \(i\) in the target zonation, \(B_k^s\) is the total built-up area of zone \(k\). \(B_{h_j}^{s\cap t}\) and \(B_h^s\) are defined similarly for the trip destination zones. Thus \(\rho_k\) is the proportion of built-up area located in the intersection of \(k\) and \(i\) in relation to the total built-up area of zone \(k\). Similarly, \(\rho_h\) is the proportion of built-up area in intersection of \(h\) and \(j\) in relation to the total built-up area of zone \(h\). The two proportions will then be applied in Equation (1) for interpolation. The remaining procedures will be the same as that discussed for Model-1.

**Model – 3: Gravity-type Flow Line Interpolation**

The third model draws on the classic gravity model for spatial interaction modeling. It has been almost universally accepted that distance plays a paramount role in spatial intersections. This model attempts to explicitly factor it in while the previous two models believe that the friction of distance is already embedded in the known trip magnitudes in the source zonation. First, we sum up the total trips originating and ending at each zone in the source zonation. Then by applying simple areal interpolation similar to what is discussed in Model-1, we can derive the total trips originating and ending at each zone in the target zonation. This basically provides the row sums and column sums
of the empty trip matrix for the target zoning scheme. Estimating the empty cells of the matrix is exactly what the interpolation is about. The doubly-constraint gravity model is employed to estimate them. The model is expressed in equations (6) through (8).

Parameters of the gravity model are estimated with the complete matrix for the source zoning scheme and the distance matrix created in a GIS environment (TransCAD). The adoption of parameters is based on the reasonable assumption that the same type of friction of distance prevails in the study area regardless of zoning scheme.

\[ T_{ij} = A_i O_i B_j D_j \exp(-\beta c_{ij}) \quad (6) \]

Subject to:

\[ \sum_j T_{ij} = O_i \forall i \quad (7) \]

\[ \sum_i T_{ij} = D_j \forall j \quad (8) \]

where, \( i = \) origin, \( j = \) destination, \( A_i \) and \( B_j \) are balancing factors, \( O_i \) and \( D_j \) are forecasted trips produced and attracted.

It is important to note that the traditional doubly constrained gravity model neglects intrazonal trips (i.e., \( i = j \)) with a zero value. In reality, depending on the aggregation level and the spatial pattern of land uses, intrazonal trips can be very significant in most urban areas. Thus, the traditional neglect of intrazonal trips should be viewed with reservation. Instead, we adopt O’Kelly and Lee (2005)’s method to estimate the actual minimum distance for each intrazonal flow direction, as shown in equations (9) and (10). Then we modify the distance matrix derived in GIS by replacing the 0 intrazonal trip distances (the diagonal cells) with the estimated intrazonal distances.

\[ y_{ij} = \sqrt{\frac{R_i}{\pi}} \quad (9) \]
\[ y_{ij} = \frac{y_{i}^{\text{min}}}{y_{ij}} \times y_{i}^{\text{min}}, \]  
\( i = j, R_i = \text{area of } i, Y_i = \text{newly intrazonal distance of } i, Y_i^{\text{min}} = \text{minimum distance of row } i \)

MODEL EVALUATION AND CASE STUDY

In order to evaluate the validity and accuracy of the proposed methods, a case study of Fulton County in Georgia is conducted. The ground truth data are the corresponding CTPP data which are distributed by the U.S. department of transportation. Root Mean Square Error (RMSE) is utilized for accuracy assessment.

Study Area and Data

The study area is Fulton County, located in the middle of the Atlanta metropolitan area in Georgia. The county is the most populous area in the state of Georgia. As of the 2008 Census, the population was 1,014,932, up from 816,006 in 2000 and 648,951 in 1990. It has been experiencing significant growth over the past decades. The population increase was 20.5 percent from 1990 to 2000, and 24.4 percent from 2000 to 2008. The total employment in Fulton County was 315,336 in 1990 and 385,442 in 2000, an increase of 18.2%. Figure 2-3 shows a map of Fulton County. There are 167 census tracts and 492 small places listed by the Metropolitan Planning Organization (MPO).

The spatial boundaries of census tracts and small places as well as the commuting trips between each pair of the spatial units are obtained from the decennial census CTPP...
data products for the year of 2000. It is important to note that small places in the MPO are mostly portions of a census tract, indicating a higher level of spatial granularity for the flow interpolation. Census tract and small places in the MPO are the finest zoning schemes available in the 2000 CTPP. According to the CTPP data, there are totally 264,100 trips between census tracts and 262,021 trips between small places in Fulton County.

Figure 2-3. Fulton county in Georgia
**Accuracy Evaluation**

We use Root Mean Square Error (RMSE) to examine the accuracy of the interpolation results from each of the three interpolation models. The RMSE is commonly used measures of differences between estimated values and the corresponding observed values. It is the square root of the variance of the residuals between actual and expected values. Equation 11 is the mathematical definitions of the RMSE.

\[
\sqrt{\frac{\sum (X'' - X)^2}{N}} \tag{11}
\]

where \(X\) is the actual value, \(X'\) is the estimated value, and \(N\) is the number of flow directions.

To have a more comprehensive evaluation of the interpolation method, we tested it in both directions with regard to aggregation level of the spatial units. We call them disaggregating interpolation and aggregating interpolation respectively. Disaggregating interpolation refers to that from a coarser scheme (larger zones) to a finer zoning scheme (smaller zones). Aggregating interpolation is from a finer data to a coarser scheme.

Specific to the case study of Fulton County, the disaggregating interpolation is from the there are 167 census tracts to the 492 small places, and vice versa. We performed two independent interpolations, one aggregating and one disaggregating interpolation, using each of the three models. Summary of the estimation accuracies are shown in Table 2-2.
Table 2-2. Interpolation accuracy analysis summary

<table>
<thead>
<tr>
<th></th>
<th>Areal-weighted</th>
<th></th>
<th>Intelligent</th>
<th></th>
<th>Gravity-type</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Aggregating</td>
<td>Disaggregating</td>
<td>Aggregating</td>
<td>Disaggregating</td>
<td>Aggregating</td>
<td>Disaggregating</td>
</tr>
<tr>
<td>Actual total trips</td>
<td>264,100</td>
<td>262,021</td>
<td>264,100</td>
<td>262,021</td>
<td>264,100</td>
<td>262,021</td>
</tr>
<tr>
<td>Estimated total trips</td>
<td>261,815</td>
<td>264,021</td>
<td>261,813</td>
<td>264,021</td>
<td>264,100</td>
<td>262,021</td>
</tr>
<tr>
<td>RMSE (Overall)</td>
<td>1.1287</td>
<td>4.9960</td>
<td>1.2865</td>
<td>4.8867</td>
<td>21.4838</td>
<td>5.3748</td>
</tr>
<tr>
<td>Intrazonal flow</td>
<td>3.5739</td>
<td>52.7399</td>
<td>3.5730</td>
<td>51.6302</td>
<td>212.2508</td>
<td>87.5975</td>
</tr>
<tr>
<td>Interzonal flow</td>
<td>1.2508</td>
<td>4.3984</td>
<td>1.2571</td>
<td>4.2778</td>
<td>13.8936</td>
<td>3.6496</td>
</tr>
</tbody>
</table>

Discussion

From Table 2-2, it can be found that both the areal-weighted and the intelligent flow line interpolation deliver very accurate estimations. The best performing model is the intelligent model. This should give us no surprise as this method makes use of most information, including the data shared by all models and the additional ancillary land use/land cover data that shed lights on trip patterns. It is also observed that the second best performer, the areal-weighted flow line interpolation, has a very close performance. We think it is mostly due to the nature of the case study area. The area is a county located in downtown Atlanta and the most populous area in Georgia. This means that the urbanization has been fairly thorough in the study area. Thus the additional information about non built-up area only covers a small area. We anticipate that the gap between the performances of the two models may substantially increase as the study area moves towards the suburb, urban fringe, and ex-urban regions.

Another important and somewhat surprising finding is the poor performance of the gravity-type of flow line interpolation. We initially thought the consideration of friction of distance would contribute positively to estimation accuracy. But after careful examination of data and the model, we think it does make a lot of sense. There are two
major reasons for it. First, in order to use the spatial interaction model, the detailed trip distribution data were reduced to highly aggregated trip totals (row sums and column sums). Much of the valuable detailed trip distribution information is thus left unused. Secondly, the role of distance has already embedded in the raw data in the source zoning scheme. Only the distance differences between different intersection areas in the same source zones are left unconsidered in first two interpolation models. Because such differences are usually not too big comparing to trip length, the loss of accuracy from it is not as much as that could have been lost from the underuse of available information as discussed in the first point. Therefore, the interpolation results are much less accurate than the other two models that do make use of every piece of the source data.

The third interesting observation is that while both areal-weighted and intelligent interpolations produce good results, the disaggregating interpolations are consistently less accurate comparing to their aggregating counterparts. This might not be surprising because usually aggregation of detailed data gives more accurate result than the disaggregation of general data. Several sources of uncertainty may account for it. The most important one is that the disaggregating interpolation relies more on the model’s ability to accurately estimate the distribution pattern of origins (or destinations) within each source zone. Thus the areal-weighted disaggregating interpolation is much more vulnerable to the oversimplified assumption of uniform trip distribution within each spatial unit. This is also why we can observe the most significant advantage of the intelligent flow line interpolation over other models for the case of disaggregating interpolation. The final noteworthy point is that all interpolation models show much better accuracy for interzonal flows than the intrazonal flows. We think it may partly due
to the inaccuracies in the ground truth data due to spatial aggregation. Note that although total trips in a county should be the same regardless of zoning schemes, there is a difference of 2079 trips between the 2000 CTPP trip data in the two different zoning schemes. Other reasons of course also include the inability to know the real trip distances for intrazonal trips.

CONCLUSIONS

Traditional areal interpolation methods deal with point or areal data but not flow lines. This study coins the concept of flow line interpolation and also makes a first attempt to model the interpolation of flow data from one zoning scheme to another. Three models were developed and we find two of them, namely the areal-weighted and the intelligent flow line interpolation models, are very promising. The choice of the two models depends on the characteristics of the study area as well as the availability of ancillary information. An evaluation with case study performed on real data for a study area in the metropolitan Atlanta suggests that the developed models can produce very good interpolation results. It is also observed that interpolation accuracy can be largely influenced by interpolation direction. Aggregating interpolations (from finer source zoning scheme to coarser target zoning scheme) usually have excellent accuracy.

Flow line interpolation may have far-reaching application potentials. Discussed below are just a few examples. By applying the proposed flow line interpolation, time series flow data can be produced in any identical spatial zoning scheme. Thus further investigation of spatial-temporal changes of flows can be performed. It also helps to derive flow data at various spatial aggregation levels. For instance, since the launch of
American Community Survey (ACS) program, commuting data are released much more frequently but at much coarser levels of spatial granularity (e.g. county level). Flow line interpolation will allow researchers to convert such aggregated flows to data of finer spatial granularity.

As a first attempt, the study of flow line interpolation has ample room for further exploration. For instance, a future study needs to consider heterogeneous commuting patterns. In our method, commuting data only contains total commuting flows between zones, aggregating trips by different ethnic, social, and economics groups. This aggregated approach is appropriate for predicting overall flow patterns in a metropolitan area, but may mislead if heterogeneous commuting patterns (e.g., race, gender, and income) are of concern. Future research on disaggregated work trips with the interpolated commuting data will provide a more realistic representation of the correlations between jobs and housing with consideration of race, gender and income levels over time.
REFERENCES


Jarvis C H, Stuart N and Cooper W 2003 Infometric and statistical diagnostics to provide artificially-intelligent support for spatial analysis: the example of interpolation. *International Journal of Geographical Information Science*, 17: 495


*Journal of Geographical Systems, 1*: 323-346

Tobler W R 1979 Smooth Pycnophylactic Interpolation for Geographical Regions. 

*Journal of the American Statistical Association, 74*: 519-530


Wilson A G 1971 A family of spatial interaction models, and associated developments. 

*Environment and Planning, 3*: 1-32

CHAPTER 3

TRACKING ETHNICALLY DIVIDED COMMUTING PATTERNS OVER TIME
– METHOD AND A CASE STUDY OF ATLANTA

\[2\] Jang, W. and X. Yao. Submitted to *The Professional Geographer*, 05/10/2012.
ABSTRACT

This study is motivated by insufficiencies in two areas in the literature. First, some technical barriers have hindered investigations of changing job-housing patterns over time. Secondly, traditional dichotomous studies (black-white) of ethnically divided commuting patterns are insufficient to paint the big picture of such dynamics in a multi-ethnicity metropolitan area. This research fills the gap by presenting an approach to the spatio-temporal analysis of commuting patterns by ethnicity. A case study is performed to track changing commuting patterns for whites, blacks, and Latinos in Atlanta over the last two decades. The results shed light on our understanding of the changing job-housing dynamics, particularly that of Latinos. Keywords: urban commuting, job-housing relationship, spatio-temporal changes, race and ethnicity, Latinos.
INTRODUCTION

Job-housing relationships have received considerable attention in the urban geography literature. Abundant studies have accumulated to investigate job-housing relationship, particularly the imbalance between jobs and housing for different population groups in urban spaces. A noteworthy thread of research is the spatial mismatch hypothesis (SMH). Kain (1968) first used the term to describe the phenomenon of inner-city minority residents facing serious limitations in job opportunities in suburban areas. Since then, many studies found empirical support for SMH in various cities and social groups and provided theoretical discussions for the geographical and socioeconomic phenomenon (Wilson 1987; Houston 2005; Sultana 2005; Johnson 2006; Bohon, Stamps, and Atiles 2008; Fernandez 2008; Joassart-Marcelli 2009).

A central theme in the discourse is that the phenomenon is racially-or-ethnically (hereafter ethnically) divided. It is suggested that various socioeconomic factors disproportionately affect inner city minorities who have longer commuting distances between jobs and housing, thus ultimately discouraging them from job search in the suburbs and leading them to compete with other minority groups for low-paying inner-city jobs (Chung, Myers Jr, and Saunders 2001; Houston 2005; Gobillon, Selod, and Zenou 2007). In a discussion of the mechanisms of spatial mismatch for blacks, Gobillon, Selod, and Zenou (2007) suggested seven factors behind such phenomena. Four of these factors are from the job-seekers’ perspective, including relatively high commuting cost, less access to information for more distant job centers, lack of incentive to travel a longer distance and thus restrictive search activities in farther job markets.
It is noted, however, that most of the theoretical discussions and empirical evidences are made between whites as the majority and blacks as the minority. This dichotomous approach is problematic because population compositions in American cities have experienced dramatic changes over the past few decades. The two most recent censuses (in 2000 and 2010) indicate that the Latino has been the fastest growing population in the United States and has already become the largest ethnic minority group. In the South, for instance, Latino population increased an average of 341 percent in the South (Odem and Lacy 2009). The dramatic increase of Latino population has been influencing the cultural, economic, and political aspects of the South in significant ways (Smith and Winders 2008; Hume 2010; Winders 2011). Consequently, it is increasingly important to consider Latinos in the study of job-housing relationships in the context of this changing population makeup.

In prior studies, the methods of studying job-housing relationships range from simple measures of jobs-housing ratio to identifying spatial patterns of such relationships from urban commuting data. A gap in the literature is the lack of trajectory analysis of commuting pattern changes over time. This missing link may relate significantly to the methodological barriers discussed below. In the United States, the Census Transportation Planning package (CTPP) has served as the major data source for studies of job-housing relationships and particularly commuting patterns. Data availability and technical barriers, however, exist for comparative studies of time-series commuting patterns for different ethnic groups (O’Kelly and Lee 2005; Sang, O’Kelly, and Kwan 2010; Jang and Yao 2011). Because the CTPP is part of the decennial census, two major problems occur for the analysis at issue. First, trip data are not always available at the same spatial levels
of census units in different years, and even the same census level (e.g. census tracts) often alter their spatial boundaries over the ten-year period of time between two consecutive censuses. This inconsistency of underlying spatial units makes it impossible to assess directly the temporal changes of spatial patterns. Secondly, the commuting flow data are either available as total number of commuters in the respective pair of spatial zones or disaggregated by a limited number of categorical variables such as poverty status and mode of transportation. A problem occurs when researchers want to analyze trip patterns for different ethnic groups, which require disaggregated trip data by ethnic category. The synergy of the above two problems greatly obstructs any close-up examination of the changing commuting patterns by different population groups in a city.

This study addresses the aforementioned issues by developing an approach to analyzing the changing spatial patterns of commuting trips by different population subgroups. We also apply this approach in a case study of metropolitan Atlanta to investigate the changing patterns of commuting for whites, blacks, and Latinos during the past two decades. The study attempts to make a threefold contribution. First, it makes a methodological contribution by introducing a two-step approach to analyzing time-series commuting data. Secondly, it makes an empirical contribution by revealing changing commuting patterns in Atlanta for different ethnoracial groups. Finally, with the spotlight focusing on Latinos, the study attempts to shed light on our understanding of the job-housing relationships of the underexplored but fastest-growing ethnic group vis-a-vis other population groups.
PRIOR RELATED WORK

Urban Commuting Patterns of Latino American

A large body of literature has accumulated on urban commuting in the U.S. Prior studies investigated job accessibility (Wang 2000), urban public transit (Polzin 1999; Sinha 2003; Yao 2007), mobility (Sanchez, Shen, and Peng 2004), and spatial mismatch of jobs and housing for population subgroups (Blumenberg 2004; Covington 2009), to name a few examples. However, for reasons discussed above (the CTPP data issues), few previous studies developed a spatio-temporal perspective to study commuting patterns. Moreover, a majority of those studies focused on the differences between blacks and whites, while Latinos received insufficient research attention for the investigation of their commuting patterns. Thus, previous studies provided inadequate answers to many questions of Latinos’ job-housing dynamics, especially those about changes over time.

According to the U.S. Census in 2004 (http://www.census.gov/population/www/pop-profile/files/dynamic/RACEHO.pdf), the Latino population has surpassed the black population and became the largest ethnic group in the United States. The most important factor of dramatic growth of the Latino population is believed to be the 1986 Immigration Reform and Control Act (IRCA), which was enacted as part of congressional reform of the U.S. immigration law (Kossoudji and Cobb-Clark 1996). Although the IRCA limits undocumented immigration, it also granted amnesty to long-term illegal immigrants who entered the United States before 1982. The post-IRCA era provided greater mobility for Latinos who achieved legal status. Along with other factors such as anti-immigrant environments and economic recession in the traditional areas of largest Latino population (e.g. Texas and
California), the IRCA has motivated rapid Latino migration to the South which attracted Latinos with better economic opportunities (Furuseth and Smith 2006). Many scholars have described the impact of such rapid influx of Latinos on demographic (Furuseth and Smith 2006; Yarbrough 2010), socio-economic (Furuseth and Smith 2006; McClain et al. 2006; Smith 2006; Smith and Winders 2008), and political (Winders 2005; Winders 2006) aspects in the South. Smith (2006) argued that due to the huge influx of Latino migration in the South, black workers face more competition with Latinos for low-income jobs (e.g. poultry, service and construction).

Albeit without detailed spatial analysis of georeferenced trip data, a number of studies reported general findings about some characteristics of Latino commuting patterns. Latinos are found to be more suburbanized in housing and job locations than blacks (Raphael and Stoll 2002). However, for those low-skilled Latinos in central cities, they share similar job-housing experiences as that of blacks, including longer commuting times (Liu 2009) and higher levels of job-skill requirements (Stoll 2005). In a study of the employment accessibility of low-skilled Latino immigrants in three major metropolitan areas, Chicago, Los Angeles, and Washington D.C., Liu (2009) found that Latino inner-city residents had longer commute times than whites. Based on an analysis of Los Angeles’s (blacks, whites, and Latinos) and Atlanta’s year 2000 census (only blacks and whites data available), Stoll (2005) argued that compared with less-educated whites, less-educated Latinos and blacks suffer from so-called geographical skills mismatch as they tend to reside in high-skill job markets (e.g. central city) and thus poorer employment outcomes. The problem was found to be more severe for blacks but is significant enough for Latinos as well. As argued by Joassart-Marcelli (2009), Latino immigrants’
employment opportunities are more likely to be geographically constrained due to poorer accessibility to jobs and more dependency on the ethnic neighborhood networks. In addition to housing locations, poorer accessibility is also highly related to more restrictive transportation mode choices due to limited car ownership. Bohon et al (2008) examined the impact of limited transportation options for Latinos in Georgia and suggested that Latino employees were less likely to be driving alone to work than whites due to lower car ownership.

Commuting Trip Data: Aggregated vs. Disaggregated

Considerable scholarly work has addressed the important yet different roles of aggregated and disaggregated spatial interaction data in spatial analysis (Horner and Murray 2002). Two types of aggregation can be identified for commuting trip data, those aggregated spatially and those aggregated categorically by non-spatial variables such as socio-economic factors (e.g. race/ethnicity, income level, mode of transportation). Pirie (1979) argued that aggregated data ignore the spatial distribution of activities and lack necessary details, as compared with more disaggregated data. Generally, disaggregated trip data bear more details of spatial or non-spatial characteristics of activities.

Past studies utilized interpolation to disaggregate data. Interpolation is a process of estimating data values between spatial units in a target zoning scheme based on such data in a source zoning scheme. Interpolation can be further categorized as simple interpolation (e.g. areal-weighted) and intelligent interpolation (e.g. dasymetric method) (Flowerdew and Green 1992; Goodchild, Anselin, and Deichmann 1993; Sadahiro 1999; Mennis 2003; Kim and Yao 2010). The basic assumption in simple interpolation is that
data are uniformly distributed within each source zone, while the intelligent interpolation techniques are freed from the assumption by using additional information (e.g. land cover or road networks) to discern the heterogeneous spatial structure of source zones. Many studies showed that the simple interpolation was a weaker estimator than the intelligent interpolation (Mennis 2003; Kim and Yao 2010).

However, none of the traditional interpolation methods are able to deal with flow data, as each of the latter was associated with a pair of spatial zones. Research on interpolating flow data is still in its infancy. Although some previous studies adopted interpolation techniques to trip data to examine commuting patterns (Horner 2007; Li, Corcoran, and Burke 2010), those studies applied the traditional single-unit interpolation in which each value was associated with one spatial unit only. Jang and Yao (2011) recently developed a new line of interpolation techniques, dubbed flow line interpolation, to fill this gap. The flow line interpolation models redistribute flows between zones in a new zoning system based on the flow data in a known zoning system.

Another type of disaggregation is by categories of population. Examples include population subgroups by race, income level, or any other categorical variables. O’Kelly and Lee (2005) developed an Information Minimization (IM) method to disaggregate the total trips between a pair of zones to subtotals of trips by population subgroups. Unlike the overall commuting flow data, disaggregated commuting data by subgroups can help researchers and practitioners understand unique commuting patterns by each specific population and the roles of socio-economic factors on commuting patterns.
METHODS

Study Area

We focus on the Atlanta metropolitan area of 14 counties to examine the changing patterns of commuting trips by race or ethnicity (hereafter ethnicity) during the past two decades. Although the U.S. Census designates a 28-county region as the Atlanta Metropolitan Statistical Area, the census CTPP data typically tabulate trip data for a more compact metropolitan region. In our study area, the 14-county core region around the city is the largest common area covered in all CTPP 1990, 2000, 2008. The area covers all and beyond the region under the metropolitan area’s planning authority (the Atlanta Regional Commission).

Metropolitan Atlanta has been informally divided into two sub-regions, Atlanta-North and Atlanta-South, approximately separated by the east-west line of Interstate-20. As shown in Figure 3-1, the two central city counties, Fulton and DeKalb, are each divided into a northern half which was primarily white and a southern half which was primarily black (Jaret, Ruddiman, and Phillips 2000). Most whites live in North Fulton and suburban areas. In contrast, Latinos are concentrated in Gwinnett and Cobb counties, while also presented in Fulton and DeKalb counties. Figure 3-1 shows administrative boundaries, the highway structure, and subway lines (MARTA) of Atlanta. The study area comprised 14 counties, 601 census tracts and 958 traffic analysis zones (TAZs) in census 2000. Table 3-1 lists the changing makeup of working population in the metropolitan from 1990 to 2008. The share of Latinos among all workers has increased from about 2 percent in 1990 to nearly 10 percent in 2008, a growth of 400 percent over the 18 years of time.
Figure 3-1. Metropolitan Atlanta Area in Georgia

Table 3-1. Changing ethnoracial compositions of workers in metropolitan Atlanta 1990 -2008

<table>
<thead>
<tr>
<th></th>
<th>1990</th>
<th>2000</th>
<th>2008</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>1,051,160</td>
<td>1,185,406</td>
<td>1,239,975</td>
</tr>
<tr>
<td></td>
<td><strong>73.63%</strong></td>
<td><strong>62.24%</strong></td>
<td><strong>54.51%</strong></td>
</tr>
<tr>
<td>Black</td>
<td>320,696</td>
<td>500,483</td>
<td>682,820</td>
</tr>
<tr>
<td></td>
<td><strong>22.46%</strong></td>
<td><strong>26.28%</strong></td>
<td><strong>30.01%</strong></td>
</tr>
<tr>
<td>Latino</td>
<td>28,173</td>
<td>123,459</td>
<td>216,475</td>
</tr>
<tr>
<td></td>
<td><strong>1.97%</strong></td>
<td><strong>6.48%</strong></td>
<td><strong>9.52%</strong></td>
</tr>
<tr>
<td>Other</td>
<td>27,566</td>
<td>95,108</td>
<td>135,680</td>
</tr>
<tr>
<td></td>
<td><strong>1.93%</strong></td>
<td><strong>4.99%</strong></td>
<td><strong>5.96%</strong></td>
</tr>
<tr>
<td>Total</td>
<td>1,427,595</td>
<td>1,904,615</td>
<td>2,274,950</td>
</tr>
<tr>
<td></td>
<td><strong>100%</strong></td>
<td><strong>100%</strong></td>
<td><strong>100%</strong></td>
</tr>
</tbody>
</table>
Data

Commuting data and related demographic variables were obtained from CTPP data of 1990, 2000, and 2008 respectively, published by the United States Department of Transportation. The CTPP data based on decennial census dataset are special tabulations of workers at residence (CTPP 1), workplace (CTPP 2), and commuting flow (CTPP 3) data. The CTPP 3 provides commuting flow data between homes and workplaces at various geographic aggregation levels. In this study, we have CTPP 1990 and 2000 based on decennial censuses, and recent CTPP data based on the 3-year (2006-2008) American Community Survey (ACS) data. For the decennial CTPP data, we choose the spatial level of traffic analysis zone (TAZ) for year 1990 and census tract for year 2000 because they are the finest spatial levels available in the respective years. The 3-year ACS-based CTPP data were obtained from the website of the American Association of State Highway and Transportation at http://ctpp.transportation.org/Pages/3yrdas.aspx. The spatial granularity in ACS CTPP 3 is much coarser. We use the county-level because this is the only spatial level at which the data format is consistent with the requirements for data analysis.

In step 2 as discussed below, the method requires ancillary information for better interpolation accuracy. This study uses land cover data for the purpose. We acquired the data from Georgia Data Clearing House for the years of 1991 and 2001, and from Multi-Resolution Land Characteristics Consortium (http://www.mrlc.gov) for the year of 2006 at a spatial resolution of 30 meters. Using the Level 1 Anderson classification (http://landcover.usgs.gov/pdf/anderson.pdf), we further reclassified the data to a binary scheme as built-up and non built-up land.
Step-1. Disaggregate trip data by categorical demographic or socio-economic factors

In recognition of heterogeneous journey-to-work patterns among different population subgroups, O’Kelly and Lee (2005) first developed the Information Minimization (IM) model to disaggregate journey-to-work (JTW) data by type of occupation. Some scholarships further extended the model by gender and occupation (Sang, O’Kelly, and Kwan 2010; Kim et al. 2012). This study implements the IM model to disaggregate commuting trips by ethnicity level. The model was an extension of the traditional doubly-constrained gravity model of trip distribution, which was designed to mathematically identify the best matching trips between all pairs of origin and destination zones under the constraints of known total trips made into and out of each zone (O’Kelly and Lee 2005).

In the IM model, as shown in Equation 1, the trip volume between each pair of zones was estimated by attributes of the spatial zones and distance friction factors that were calibrated from the doubly-constrained gravity model. Equations 2 and 3 represent the inverse of row (home) and column (workplace) balancing factors, constrained by balanced home-based trip production (in Equation 4) and workplace-based trip attraction (in Equation 5).

\[
T_{ij}^k = A_i^k O_j^k B_j^k D_j^k \exp(-\beta^k c_{ij}) \quad (1)
\]

where,

\[
A_i^k = [ B_j^k D_j^k \exp(-\beta^k c_{ij})]^{-1} \quad (2)
\]

\[
B_j^k = [ A_i^k O_i^k \exp(-\beta^k c_{ij})]^{-1} \quad (3)
\]

subject to
In the equations, \( k \) is the index of population subgroup, \( T_{ij}^k \) is the total number of commuting trips between spatial zones \( i \) and \( j \) for \( k^{th} \) subgroup. \( A_i \) and \( B_j \) are balancing factors for origin \( i \) and destination \( j \). \( O_i \) is the number of workers in category \( k \) from zone \( i \). \( D_j \) is the number of workers in category \( k \) in zone \( j \). \( c_{ij} \) is the travel cost (e.g. distance or time) between zone \( i \) and zone \( j \). The study uses network distances between zone centroids as \( c_{ij} \). However, this centroid-based measure will falsely give a zero travel distance for intrazonal trips. Therefore, we apply a special-designed and well-received method for intrazonal trip distance estimation (O’Kelly and Lee 2005; Jang and Yao 2011).

### Step-2. Interpolate trip data in different years to a common spatial zoning scheme

To align multiple trip flow datasets under inconsistent spatial zoning schemes, we apply the intelligent flow line interpolation method developed by Jang and Yao (2011). The method converts flow data between pairs of spatial units from one zoning scheme to another. The intelligent flow line interpolation abandons the commonly used and oversimplified assumption of uniform distribution of events but instead uses ancillary data (land use types in this study) to inform the distribution of trip origins/destinations within respective zones. For instance, urban areas produce and attract higher traffic flows than forest and agriculture lands do. The model is mathematically expressed as follows:
In the equations, $T_{ij}^t$ is the flow volume from zone $i$ to $j$ in the target zoning scheme, $T_{kh}^s$ is the flow volume from zone $k$ to zone $h$ in the source zoning scheme. Variables $m$ and $n$ are the total numbers of origins and destination zones respectively in the source zoning scheme. $P_k$ is the proportion of built-up area located in the intersection area of zone $i$ and zone $k$ in relation to the total built-up area of zone $k$. Similarly, $P_h$ is the proportion of built-up in the intersection of zone $j$ and zone $h$ in relation to the total built-up area of zone $h$. $B_{kl}^{s\cap t}$ indicates the built-up area location within the intersection of trip origin zone $k$ (in the source zonation) and trip origin zone $i$ (in the target zonation). $B_k^s$ is the total built-up area of zone $k$. $B_{hj}^{s\cap t}$ and $B_h^s$ are defined similarly for the trip destination zones.

When converting the commuting data of three different years to a common spatial zoning scheme, ideally it is the best to choose a scheme whose spatial granularity is comparable to that of the source data such as TAZ or census tract. However, for the following two reasons, we decided to choose the Public Use Microdata Area (PUMA) as a common zoning scheme that is coarser than a TAZ and census tract yet finer than the
county level. First of all, a common characteristic of interpolation is that results are much less accurate when converting from highly aggregated (e.g. county) to disaggregated (e.g. TAZs) spatial levels. Therefore, choosing a spatial zonation whose spatial granularity is somewhere in between is a good compromise between accuracy and spatial resolution. Secondly, if a TAZ or census tract is chosen, the number of records in the matrix computation exceeds the limit allowed by the software (TransCAD) in the step of shortest path finding which is needed for visualization of flows. We modify the census Public Use Microdata Areas (PUMAs) to make it the common spatial zonation. The PUMA zonation has zones of fine sizes in the central city but some suburban and exurban PUMAs are even larger than counties in the study area. Therefore, we keep the central city PUMAs intact as they perfectly satisfy the size preference stated above. Only the suburban and exurban PUMAs are subdivided into smaller areas. To do so, we overlay these PUMAs over population centers (cities and towns) and subdivide the PUMAs so that there is exactly one population center in each new zone. Two criteria are used in the delineation process. First, the new dividing lines should coincide with existing boundary lines of census tracts. Second, we make sure each resulting zone is covered by at least 20 percent of urban or built-up land. As a result, there are 46 modified PUMAs in the study area, compared to 32 original PUMAs.

RESULTS AND DISCUSSIONS

The results of the study reveal spatio-temporal patterns of commuting trips by ethnicity in metropolitan Atlanta between 1990 and 2008. As shown in Table 3-2, albeit in different spatial granularity, a direct and rough comparison of trips in the three years
indicates a 42 percent increase (1990-2000) and a 71 percent increase (1990-2008) in total trips regardless of ethnicity. However, for Latinos, the increases were 347 percent and 668 percent in the respective two periods of time, suggesting much more rapid growth than other populations and a particularly dramatic increase in the most recent decade.

Table 3-2. Commuting trips by race/ethnicity

<table>
<thead>
<tr>
<th></th>
<th>Trips</th>
<th>% Trips</th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>939,783</td>
<td>1,158,422</td>
<td>1,225,754</td>
</tr>
<tr>
<td>Black</td>
<td>289,347</td>
<td>463,037</td>
<td>632,601</td>
</tr>
<tr>
<td>Latino</td>
<td>25,390</td>
<td>113,516</td>
<td>200,123</td>
</tr>
<tr>
<td>Other</td>
<td>24,584</td>
<td>87,158</td>
<td>123,202</td>
</tr>
<tr>
<td>Total</td>
<td>1,279,104</td>
<td>1,822,133</td>
<td>2,181,680</td>
</tr>
</tbody>
</table>

*Ethnically Divided Commuting patterns*

Before visualizing changes of commuting patterns over time, let us first take a quick look at spatial patterns of housing and job distributions in the study area. Constrained by the space limit of the paper, Figure 3-2 only gives the distributions in 2000. It shows that whites disperse in all areas in the suburbs and exurbs. Latinos and blacks, however, tend to have their own clusters of residential areas. Blacks reside in and around central city of Atlanta, with one prominent spatial expansion into the southwest suburb. Latinos are concentrated in Gwinnet County in the northeast suburb, smaller clusters in the central city, and a few other suburban areas such as the suburban city of Marietta in the northwest.

Although it is somewhat more compact than the residential distribution, the job distribution of whites is also much more dispersed than that of other ethnic groups.
Interestingly, job distribution for Latinos follows a spatial pattern that is very similar to that of its housing distribution. This may suggest that Latinos tend to live close to work. Indeed, the job-housing pattern of Latinos in Atlanta is primarily associated with job opportunities. The most Latino-concentrated area, the northeastern Gwinnett County, is a relatively recently developed suburban area. Since the 1996 Olympics in Atlanta, the high demand for low-wage labor has attracted Latinos immigrants in Gwinnett which offers many types of low wage jobs such as construction, maintenance, production, transportation, and service (ARC 2010). In the northwest suburbs, Marietta is the third largest city in the metropolitan area. Following low property taxes, a moderate cost of living and easier transportation access (I-285) so close to the city of Atlanta, Marietta has experienced a boom in the local construction industry and, thus, has attracted many Latino workers.
Figure 3-2. Population density at residence and workplace in 2000

Figure 3-3 visualizes the spatial patterns of commuting trips for each ethnicity in different years. Traditional visualization of flow data uses straight lines among origins and destinations. This method is not suitable for a large amount of trips over fine-grained spatial units, as the maps would be too cluttered to reveal patterns. In this study, we translate the trip flows to transportation networks by assigning corresponding trips between any pair of spatial zones to the shortest travel-time path between them. Because the shortest paths share many road segments, we implement a kernel density procedure to summarize trip volumes in a raster format. The cell size in use is 0.25 by 0.25 square miles. Figure 3-3 shows that Latino commuting patterns started from an insignificant
presence in 1990 and expanded greatly in both magnitude and spatial distribution over the 18 years. Both blacks’ and Latinos’ commuting trips experienced dramatic expansion from the central city into the inner suburbs. Whites have the strongest share of trips in magnitude in all three time periods and they expanded the farthest into the exurbs, yet the rate of expansion seems to be less than that of the two minority groups under study.

Figure 3-3 only provides maps in different years for a qualitative visual comparison. In the next section, our proposed two-step method is applied for quantitative measurements of changes in commuting patterns.

Figure 3-3. Commuting patterns by race/ethnicity in the three time-periods
Changing Spatial patterns of Commuting Trips

After decomposing commuting trips by ethnicity and aligning the data in different years to a common zoning scheme, we can derive the changes of trip volume between each pair of zones over time. Figure 3-4 depicts the changes by ethnicity over the last two decades. The figure reveals several interesting findings. First, Latinos are the only population whose commuting trips have increased across all areas of the study in the past decades, while both whites and blacks have overall increases but also some decreases of trips in certain areas. This is most likely attributed to the dramatic influx of Latino immigrants in the time period. Secondly, the trip decreases for blacks are small in amount (light colors) and only occur in the central city. The decreases in trips by whites spread out much more widely in the city and inner suburbs. This may be associated with several possible processes such as the residential spatial filtering which here means the process of whites moving further out and the vacated houses being filled by minorities. The process is evidenced by the declining housing prices in the inner suburbs as the percentage of minorities increases. Thirdly, the spatial patterns of increases are similar for Latinos and blacks in that both of them are more confined within the inner suburbs compared to their white counterparts. Taking Gwinnett County as an example: the increases for both minority groups penetrate into the local roads suggesting local travels between home and work, while increases for whites follow mostly highway segments suggesting passing-through trips. In addition, there is no growth of trips into the exurbs for the two minority groups, whereas such growth is evident for whites. However, the two minority subgroups do have differences in the geography of relative concentrations of trip increase. Significant shares of Latinos’ trip growth are found in Gwinnett County
and along Highway 400, while blacks’ trip growth concentrated in inner suburbs, particularly the far south end of Fulton and the far east side of DeKalb. The spatial patterns are likely attributed to the following facts: (1) the expansion of the black middle class resulted from an influx of people to the sunbelt city in the boom economy during the time period, and (2) the much enhanced availability of public transportation in the inner suburbs in Atlanta. The MARTA has expanded significantly since 2000. As shown in Figure 3-1, the MARTA covers the entire area where increases of commuting trips of the two minority subgroups are significant.

Figure 3-4. Changes in commuting trips by ethnicity in Metropolitan Atlanta over the time period 1990-2008
Increasing Travel Distances

Average length of commuting trips is calculated for each subgroup. We did not choose to calculate trip time because it depends on the transportation mode in use. First, the shortest network distances between all pairs of spatial zones are derived. Then an average distance of these trips is estimated using Equation (9).

\[ l^r = \frac{\sum t^r_{ij} d_{ij}}{\sum t^r_{ij}} \]  

where \( l^r \) is the average trip length for \( r \) ethnic group (\( r = \text{white, black, or Latino} \)), \( t^r_{ij} \) is the number of trips between spatial zones \( i \) and \( j \) made by \( r \) group, and \( d_{ij} \) is the shortest travel distance between zone \( i \) and \( j \).

Figure 3-5 plots the changing average distances for the three groups respectively. All groups show steady increases in travel distance over the two decades. The spatial expansion of the metro area and particularly the suburbanization of residential areas might be the primary causes. Comparing the three population groups, however, we found that Latinos and blacks consistently travel shorter distances for work than whites do. On average, Latinos travel slightly longer distances than blacks, yet the difference between them has been diminishing over time. In fact, the gap between travel distances of whites and minority groups has also narrowed steadily.
CONCLUSIONS

This study develops an approach to identifying spatio-temporal changes in commuting patterns by ethnicity. We discuss two innate barriers for this kind of spatio-temporal analysis. The first challenge is that the most comprehensive and popular commuting flow data (CTPP) have been provided on inconsistent sets of spatial units in different years and thus the tracking of spatial changes over time was not possible. The second challenge is the unavailability of flow data for population subgroups. The proposed approach overcomes both barriers. Our empirical study of Atlanta’s ethnically-divided commuting patterns testifies the usefulness of the approach. The spatio-temporal investigation proves that ethnicity has been a significant factor in job-housing relationships in Metropolitan Atlanta. In general, blacks and Latinos have more clustered residential patterns in the central city and inner suburbs, while whites’ housing locations are more dispersed throughout the suburbs and exurbs. Compared to blacks, Latinos’
housing and job locations are somewhat more suburbanized, thanks to the availability of low-pay job opportunities in various suburban areas. Compared to whites, Latinos and blacks tend to be more constrained by travel distance to work.

Limitations of the two techniques used in the proposed approach are also noteworthy. The first barrier of spatio-temporal analysis of commuting pattern is tackled by the flow line interpolation technique. The method has the weakness of lower estimation accuracy when interpolating from data of a highly aggregated spatial level to that of a fine spatial level. In this study, the year 2008 data is at county level, which is of much coarser spatial granularity than the levels of TAZs and census tracts in the 1990 and 2000 data. This disparity discounted the spatial resolution in the interpolation results that would otherwise be allowed by the fine-grained data of 1990 and 2000 CTPP. For future research, it will be interesting to apply the approach again on the forthcoming 2010 CTPP data for more close-up investigations. The second barrier is overcome with the IM algorithm. This method currently suffers the limitation of the number of categorical variables that can be taken into the model to disaggregate trip data. The current model can only take one variable, which was sufficient in this study as we only needed to disaggregate trips by ethnicity. However, it will be truly interesting if future research can find solutions to disaggregate flows by multiple categorical factors. It will enable many more exciting investigations of commuting patterns of stratified population groups by combinations of factors such as those by ethnicity and income level, job type, transportation mode, and so on. Such explorations will surely enhance our understanding of the dynamics of job-housing relationship in cities.
REFERENCES


(last accessed 11 September 2011).


CHAPTER 4
MEASURING RACIAL PATTERNS OF COMMUTING IN ATLANTA, GEORGIA

\[ \text{3 Jang, W. and X. Yao. To be submitted to Urban Studies.} \]
ABSTRACT

Much of segregation index analysis has dealt only with the degree of residential or employment separation between groups. This study seeks to measure segregation in the dyad of residence and workplace, rather than in residence or workplace only. This research creates a new measure of segregation based on commuting trip data and evaluates the segregation of trips by race/ethnicity. To test the application potential of this segregation index, a case study of four counties in metro Atlanta was conducted. Our results confirmed that the index offers a more realistic evaluation of segregation beyond measurements confined to spaces of residence and employment, and helps us understand spatial segregation on the basis of race, where commuting segregation exists due to the unique ethnically divided jobs and housing patterns. Keywords: segregation, urban commuting, CTPP, race/ethnicity, spatial interaction
INTRODUCTION

Both civic officials and private elites have assiduously cultivated Atlanta’s image as a “world city” of growth and innovation (Rutheiser 1996). The city hosted the 1996 Summer Olympics, and has recently been ranked as one of the top ten cities in the country for young professionals (Atlanta Business Chronicle 2010). On the surface, everything looks good. Behind the image, however, the city of Atlanta was ranked the fourth most segregated large city in the U.S. in 2000 (Frey H. William and Myers 2005). Within the city limits, Atlanta had a segregation score of 83.1, which means that 83.1 percent of the population would have to move in order to achieve perfect integration. While the metro area as a whole was found to be less starkly segregated, with a score of 68.5 percent, the city still remains at a high level of segregation between blacks and whites. The high levels of segregation in the city of Atlanta today are the result of long and complex processes of racism (Bayor 2000, Kruse 2005). The city has seen racial conflicts over where whites and blacks can live, where major structures such as stadiums and the airport should be located, the names and directions of streets, locations of major highways in the city, and areas served by public transportation. As a result, most people in Atlanta live in segregated places where people, neighborhoods, and spaces associated with them can be assigned to racial categories, e.g., white neighborhood or black school (Bayor 2000).

To study segregation problems, the majority of past research has focused on the creation and analysis of segregation indices, mostly with regard to residential space (Darden et al. 2010, Iceland and Wilkes 2006, Massey and Denton 1987, Massey and Denton 1988). Some research on segregation have focused on employment space (Aslund
and Skans 2010, Horner and Marion 2009) and school space (Allen and Vignoles 2007, Frankel and Volij 2011). The primary question is how the minority groups, especially blacks, are distributed differently from the majority group, i.e., whites, over space in the past and today. The approaches above, however, may prove misleading where commuting segregation exists due to ethnically divided jobs and housing patterns. For instance, some groups face segregation in job locations or job categories even though they are not residentially segregated, or vice-versa. The ethnically divided job-housing patterns may lead to possible spatial segregation of commuting patterns as well. Thus, the traditional segregation indices cannot give us a realistic evaluation of commuting segregation where jobs and housing are ethnically divided.

In order to properly undertake commuting segregation, this paper presents a theoretical framework of a commuting segregation index and evaluates it. It modifies the current exposure index, which is a measurement of potential contact of neighborhoods between two groups. Specifically, this paper aims to consider space between home and workplace where segregation takes place. The remainder of this paper is organized as follows. The second section provides a critical review of commonly-used segregation indices, particularly the exposure index, and discusses their limitations. Section 3 develops a new commuting segregation index based on the exposure index. Section 4 discusses our results with commuting segregation scores and sensitivity analysis, and finally, section 5 draws some conclusions and proposes some further work.
LITERATURE REVIEW

Measuring Segregation: Exposure Index

Massey and Denton (1988) reviewed 20 different indices of segregation and identified five different dimensions of segregation: evenness (the similarity of majority-minority ratio across neighborhoods), exposure (potential contact and interaction between two groups), concentration (minority group is restricted to the certain area), centralization (degree to which a group lives near the center of an urban area), and clustering (minority groups living close to each other). Their results showed that each index is somewhat similar, but also distinct. Hypersegregation (i.e., when a group is highly segregated) occurs if multiple dimensions exist. 29 major U.S. metropolitan areas were found to display white-black hypersegregation according to information provided by the 2000 Census (Wilkes and Iceland 2004).

As the primary objective in this study is to measure segregation in commuting patterns, we limit our focus to the exposure index, which attempts to measure the potential contact of neighborhoods within the same groups or between two groups. The two commonly used indices are isolation (within the same groups) and interaction (between two groups), measuring two opposite yet complementary aspects of exposure. If there are only two groups, for instance blacks and whites, the sum of isolation and interaction indices is 1.0. The exposure index ranges from 0.0 to 1.0. High values of isolation and low values of interaction indicate high levels of segregation. In contrast, low values of isolation and high values of interaction represent low levels of segregation. Lieberson (1981) defines the exposure indices as follows:
Isolation

\[ P_{xx}^* = \sum_{i=1}^{l} \left( \frac{x_i}{X} \right) \left( \frac{X}{t_i} \right) \]  

Interaction

\[ P_{xy}^* = \sum_{i=1}^{l} \left( \frac{x_i}{X} \right) \left( \frac{y_i}{t_i} \right) \]

In the Equations 1-2, \( P_{xx}^* \) is the isolation index to deal with segregation within the same group, and \( P_{xy}^* \) is the interaction index to deal with segregation between two groups. \( x_i \) and \( y_i \) are the populations of group \( X \) and group \( Y \) in zone \( i \), respectively. \( X \) is the total population in group \( X \). \( t \) refers to the total population in each zone \( i \).

The exposure index has been criticized by many researchers as being both asymmetrical and aspatial (Kaplan and Holloway 1998, Wong 2002). First, the exposure indices are not symmetrical. Suppose that the members of group \( X \) represent a small proportion of the total population of an area, while the members of group \( Y \) represent the overwhelming majority. The members of group \( X \) will have high contact with those of group \( Y \), no matter how they are distributed in the study area. This is not true to small groups such as group \( X \). Thus, the exposure of the members of group \( X \) to group \( Y \) is not the same as that of the members of group \( Y \) to group \( X \), i.e., \( P_{xy}^* \neq P_{yx}^* \). Second, the exposure indices are not spatial. In other words, the index only deals with the potential contact within predefined spatial units, but not across boundaries of the spatial units (Wong 2002). In reality, there exists the probability of contact across boundaries.

Numerous studies have attempted to deal with spatial problems imposed by the exposure indices (Morgan 1983, Reardon and O'Sullivan 2004, Wong and Shaw 2011, Wong 2002). Morgan (1983) modified the exposure index to deal with the potential contact across unit boundaries based on a distance-decay function, which captures the spatial patterns of population distribution. However, Wong (2002) argued that distance
may not capture the spatial distribution if the area is highly segregated. To overcome this problem, he applied a local spatial segregation index in order to discern the heterogeneous spatial patterns. Furthermore, in consideration of people’s mobility and activity patterns, Wong and Shaw (2011) further developed a method to measure segregation in the activity space, which can be defined as any area in which individuals interact with other groups, compared to the traditional measurement of segregation as constrained by boundaries. Using travel diary data, they demonstrate that the activity space measurement can help us understand an individual’s life experience much more comprehensively than those measurements limited to residence and employment.

*Decomposing Commuting Data By Race/Ethnicity*

Many studies make use of commuting data for detailed analysis of commuting patterns in metropolitan areas in order to understand various aspects of the spatial separation between home and work. Prior studies of commuting patterns focus mainly on excessive commuting (Horner and Murray 2002, Yang 2008), accessibility (Wang 2000), urban public transit (Pirie 1979, Sinha 2003, Yao 2007), mobility (Sanchez, Shen and Peng 2004), and spatial mismatches (Blumenberg 2004, Covington 2009). Much of the analysis of commuting patterns in the past has dealt only with homogeneous characteristics (e.g., total commuting flow) to predict overall flow patterns in metropolitan areas to help transportation planners reconstruct or maintain highways in high-traffic areas (Horner and Murray 2002, Ma and Banister 2006, Peng 1997, Sultana 2002). However, commuting trips vary by race/ethnicity, gender, and income. Research by McKenzie and Rapino (2011), for instance, determined that minority groups
(especially blacks and Hispanics) have longer commuting times than their white counterparts, due to lower rates of automobile ownership. They also found that blacks are more likely to take public transportation to work, while Hispanics are the ethnic group most likely to carpool.

The Census Transportation Planning Package (CTPP), the most widely available data source for research on commuting patterns, presents real challenges for transportation researchers due to data availability and its organization of data according to geographical units which do not remain consistent over time. For data collection and confidentiality reasons, the CTPP releases data aggregated at administrative or statistical census geographical units. The commuting data are either available for the entire population in the respective pair of spatial zones (aggregated), or disaggregated by a limited number of categorical variables such as economic status and mode of transportation. This is a problem if we want to analyze commuting patterns for different ethnic groups, which requires disaggregated trip data by ethnic category.

Much scholarly work has attempted to address the challenges of disaggregating commuting data by categories of population such as race, gender, and occupation in order to discern heterogeneous commuting patterns in a city (Kim et al. 2012, O’Kelly and Lee 2005, Sang, O’Kelly and Kwan 2010, Horner 2007, Li, Corcoran and Burke 2010). O’Kelly and Lee (2005) first developed the information minimization (IM) algorithm to decompose commuting patterns by type of occupation. The primary principle of the IM is that aggregated commuting trips can be disaggregated with known detailed information. For instance, trips can be further categorized as white trips, black trips, Hispanic trips, and Asian trips. Sang et al (2010) implemented this approach for a case study of detailed
commuting patterns by gender and job types in Rochester, Minnesota. The method produced results showing that women have more jobs in administration and healthcare support while men are concentrated in transport and production. Kim (2012) also analyzed commuting variations associated with gender and occupation in Hamilton, Ohio and demonstrated that occupation has a greater effect on commuting patterns than gender.

The IM method in the Equations 3-7 below is based on the traditional doubly-constrained gravity model, which mathematically estimates trips by category (e.g., race, gender, occupation), between all pairs of origin (home) and destination (workplace). Equations 4 and 5 represent the balancing factors, such that home and workplace are balanced with regard to production of trips from home, by category, and attraction of trips to workplace, by category. In equations 6 and 7, both the sum of the trips produced from origin and the sum of the trips ending in destination are required to match with observed numbers of workers by race $k$ respectively.

$$T_{ij}^k = A_i^k O_i^k B_j^k D_j^k \exp(-\beta^k c_{ij})$$  \hspace{1cm} (3)

where

$$A_i^k = [B_j^k D_j^k \exp(-\beta^k c_{ij})]^{-1}$$  \hspace{1cm} (4)

$$B_j^k = [A_i^k O_i^k \exp(-\beta^k c_{ij})]^{-1}$$  \hspace{1cm} (5)

subject to

$$\sum_j T_{ij}^k = O_i^k \forall i$$  \hspace{1cm} (6)

$$\sum_i T_{ij}^k = D_j^k \forall j$$  \hspace{1cm} (7)

Equations 3-7, $k$ refers to category (race). $i$ and $j$ are the indices of origin ($i$) and destination ($j$). $A_i$ and $B_j$ are the balancing factors associated with origins and
destinations respectively. $O_i$ and $D_j$ are the numbers of workers in category $k$ from zone $i$ and $j$. $c_{ij}$ is the travel cost between zone $i$ and zone $j$. $T_{ij}$ is the total number of commuting trips between $i$ and $j$. $T_{ij}^k$ is the total number of commuting trips by subgroup $(k)$ from zone $i$ and zone $j$. The traditional doubly-constrained gravity model neglects intrazonal trips ($i = j$) with a zero value. In reality, there exist a great numbers of intrazonal trips than those of interzonal trips in most urban areas. O’Kelly and Lee (2005) modified the intrazonal distance based on the square root of area divided by pi ($\pi$) in order to estimate intrazonal trips.

**METHODOLOGY**

*Study Area and Data*

The study area is four of Atlanta’s core counties: Cobb, DeKalb, Fulton, and Gwinnett (See Figure 4-1). They are most highly populated counties in the metro area. As of the 2000 Census, the 1,361,593 workers lived in these four counties comprised 66.1 percent of the 2,060,422 workers living in the Metro Atlanta area. The ethnic composition of workers living in these four counties was white (55.6 percent), black (31.3 percent), Latino (7.37 percent), Asian (3.9 percent), and other race (1.86 percent). Commuting trips between home and workplace within the four counties were 1,250,709, out of 1,990,994 total trips. Due to the huge amount of trip data, we only selected these four counties, which contain 62.8 percent of all commuting trips in metro Atlanta. To prepare for the study of segregation among Hispanic and other ethnoracial groups, we categorize the population as Hispanic (Latino), Non-Hispanic White (white), Non-
Hispanic Black (black), and Non-Hispanic Asian (Asian). We exclude the category of “other race” due to its small proportion of the total working population.

Figure 4-1. Four counties in Metro Atlanta, Georgia, 2000

The primary data in this study is from the Census Transportation Planning Package (CTPP), which is a set of special tables from decennial census surveys, sponsored by the Department of Transportation (DOT). The CTPP provides information on demographic or socio-economic patterns, e.g., race, gender, income, occupation. There are three parts of the CTPP: residence (CTPP 1), workplace (CTPP 2) and journey-to-work flow (CTPP 3). Whereas both CTPP 1 and 2 provide detailed information at a disaggregated level, CTPP 3 provides general information, e.g., total number of trips, or mean travel time, at an aggregated level. This study uses the CTPP data from the year 2000. There are two reasons for using the 2000 CTPP: first, the 2010 CTPP data is not yet available, as it is scheduled for release in early 2013. Second, recent CTPP data based
on the 3-year (2006-2008) American Community Survey (ACS) are of coarse spatial granularity, e.g., county or Public Use Microdata Areas (PUMAs), as opposed to the fine-grained data from 2000, e.g., census tract and Traffic Analysis zone (TAZ). The CTPP data can be obtained from the Bureau of Transportation Statistics (BTS) website at http://www.transtats.bts.gov/. Due to the lack of data regarding trips by subgroups, e.g., according to income, gender, and race/ethnicity, based on the CTPP data, we apply the information minimization (IM) method to estimate commuting trips by ethnic groups, e.g., white trips, black trips, etc.. This process is completed by through application of Equations 3-7.

*Modify Exposure Indices for Trip Data*

We develop the modified isolation/exposure indices based on trip data to measure segregation in commuting patterns (i.e., we treat the spaces which join home to work as spaces which are also subject to segregation), and evaluate segregation of trips by ethnicity. Unlike traditional segregation indices which deal with residential and/or employment space, the new indices will observe commuting space in which segregation may occur. Commuting space in this research is defined as space between a reference space, i.e., occurs in a pair of trips between home and work, and a control space, i.e., occurs in all inbound trips from home to workplace.

Let us explain how commuting trips work over space. Figure 4-2 shows commuting trips between seven zones, i.e., zone 1 through zone 7, which fall in 49 possible types of origin-destination (OD) pairs. Each zone can be treated as residential or employment space. Trip A in Figure 4-2 takes place from zone 3 (home) to zone 4
(workplace) where there exist three inbound trips such as trip B from home and trip C and trip D to workplace respectively. To measure commuting segregation, first we consider a reference space (zones 3 and 4) where a trip takes place between origin and destination. Then we treat a comparison space (zones 2, 5, and 7) where all inbound trips occur from the origin and to the destination of the reference group. This procedure allows us to fully cover the space where segregation occurs in zones 3 and 4.

Figure 4-2. Commuting space: a reference space in trip A from home to workplace (left) and a control space with associated trip A (right)

Flows can be represented as an extended matrix (See Figure 4-3). Whereas each row identification (ID) number (i) represents origin space, and each column ID represents destination space (j), each cell in origin and destination shows a trip \( T_{ij} \). Suppose that \( x \) is a black trip \( T_{ij}^x \) and \( y \) is a white trip \( T_{ij}^y \). As shown in Figure 4-3, trip A \( T_{34}^x \) is a black trip between origin (zone 3) and destination (zone 4). Zone 3 to zone 4 in trip A is associated with three white trips from home \( T_{53}^y \) and to workplace \( T_{24}^y \) and \( T_{74}^y \) respectively. To determine the number of trips made by a reference group, e.g., black
trips, as well as the percentage of total trips represented by that reference group, we divide trip $A (T_{34}^x)$ by the total black trips between home ($i$) and workplace ($j$). In a control group (white trips), we divide all inbound white trips ($T_{53}^y$) from home by all inbound total trips ($T_i$) from home. The same principle is applied to the proportion rate of white trips to workplace, dividing all inbound white trips ($T_{24}^y, T_{74}^y$) by the total white and black trips ($T_j$) to workplace.

Figure 4-3. Extended flow matrix (7 * 7 example)

Finally, the new trip exposure index attempts to measure segregation that takes place in the commuting space. Equation 8 is a trip isolation index, which measures the degree of potential contact of trips within group X, i.e., the likelihood that a black trip will come into contact with another black trip. In contrast, Equation 9 is a trip interaction index to measure the probability of trips in group X, i.e., black trips, interacting with those in group Y, i.e., white trips. If there are only two trips (black and white), then sum
total of the exposure of black trip to black trip and the exposure of black trip to white trip is 1.0, i.e., $P_{ij}^{xx} + P_{ij}^{xy} = 1$. Trip exposure index is not symmetrical. For instance, the potential contact of black trips to white trips is not the same as that of white trips to black trips, i.e., $P_{ij}^{xy} \neq P_{ij}^{yx}$. Thus, the probability of the exposure of black trips to white trips is not the same as the probability of exposure of white trips to black trips. The first fraction of the trip isolation index is the average number of trips of the reference group, i.e., black trips, and the second bracket with two fractions is the average of both inbound trips from origin to destination in the control group (in this case, black trips). For the control group, we summarize two proportion rates of control spaces (all inbound trips from origin and to destination associated with the reference group), and make an average of them. The trip interaction index is similar to the trip isolation index with the exception of the control group (white trips). The trip isolation/interaction indices are mathematically represented as follows:

$$
\begin{align*}
\bar{P}_{ij}^{xx} &= \frac{1}{I} \sum_{i=1}^{I} \sum_{j=1}^{J} \left( \frac{T_{ij}^{x}}{T^{x}_i} + \frac{T_{ij}^{x}}{T^{x}_j} \right) \quad (8) \\
\bar{P}_{ij}^{xy} &= \frac{1}{I} \sum_{i=1}^{I} \sum_{j=1}^{J} \left( \frac{T_{ij}^{y}}{T^{y}_i} + \frac{T_{ij}^{y}}{T^{y}_j} \right) \quad (9)
\end{align*}
$$

Where

$$
T^{x}_i = \frac{1}{I} \sum_{i=1}^{I} \sum_{j=1}^{J} T_{ij}^{x} \quad (10)
$$
\[ T_i = \sum_{j=1}^{J} T_{ij}^x + T_{ij}^y \quad (11) \]
\[ T_j = \sum_{i=1}^{I} T_{ij}^x + T_{ij}^y \quad (12) \]

In Equations 8-12, \( P_{ij}^{xx} \) is the trip isolation index and \( P_{ij}^{xy} \) is the trip exposure index. \( i \) and \( j \) refer to indices of origin (home) and destination (workplace). \( T \) is the trip by race/ethnicity. \( T^x \) and \( T^y \) are trips of group X and group Y, respectively. \( T^x \) is the total number of trips in group X. \( T_i \) and \( T_j \) refer to the total number of trips of group X and Y in zones \( i \) and \( j \), respectively.

RESULTS

The indices of Trip Isolation and Interaction

Before dealing with segregation indices, let us first take a quick look at the racial composition of workers in the four selected Atlanta counties in 2000 (See Table 4-1). Whites are the majority in the four counties, both as workers (62.3 percent) and as residents (58.2 percent), with the exception of DeKalb county, in which blacks make up 50.5 percent of all residents. Whites tend to reside in the suburbs, particularly those of Cobb and Gwinnett, while most Atlanta blacks live and work in Fulton and DeKalb. Latinos and Asians are also predominantly suburban populations, with a majority of these groups both living and working in Gwinnet where Latinos make up 9.4 percent of all residents and 8.7 percent of all workplace, compared to Asians in Gwinnett making up 6.2 percent of residents and 4.5 of workplace respectively.
Table 4-1. Racial compositions of workers in Four counties in Atlanta, 2000

<table>
<thead>
<tr>
<th>County</th>
<th>Residence</th>
<th>Workplace</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>White</td>
<td>Black</td>
</tr>
<tr>
<td>Cobb</td>
<td>230,025</td>
<td>57,570</td>
</tr>
<tr>
<td></td>
<td>(72.0)</td>
<td>(18.0)</td>
</tr>
<tr>
<td>DeKalb</td>
<td>125,559</td>
<td>168,520</td>
</tr>
<tr>
<td></td>
<td>(37.6)</td>
<td>(50.5)</td>
</tr>
<tr>
<td>Fulton</td>
<td>204,753</td>
<td>139,245</td>
</tr>
<tr>
<td></td>
<td>(54.1)</td>
<td>(36.8)</td>
</tr>
<tr>
<td>Gwinnett</td>
<td>217,210</td>
<td>39,590</td>
</tr>
<tr>
<td></td>
<td>(71.3)</td>
<td>(13.0)</td>
</tr>
<tr>
<td>Total</td>
<td>777,547</td>
<td>404,925</td>
</tr>
<tr>
<td></td>
<td>(58.2)</td>
<td>(30.3)</td>
</tr>
</tbody>
</table>

*Note: Numbers in parentheses ( ) are percentages of the total number of workers.*

The results of the study reveal segregated commuting patterns in the four selected counties of Atlanta. We measured commuting segregation with the exposure indices (isolation/interaction) based on Equations 8 and 9, as well as home and work segregation based on Equations 1 and 2. First, the isolation indices are shown in Table 4-2.

Comparing the four race/ethnicity groups, white trips (71.8 percent) are the most highly segregated among the four counties. Among the minority groups, black trips (49.7 percent) are the most highly segregated, and followed by Latino (14.0 percent) and Asian trips (5.5 percent). The home isolation score is lower than that of home and JTW isolation. In other words, the residential patterns of workers in four counties of Atlanta are much more segregated than their jobs or patterns of commuting.
The interaction indices of home, work, and JTW in Table 4-3 show how each ethnic trip interacts with those of other ethnic groups. Asians (94.5 percent) are much more likely to interact with members of other ethnic groups than any other ethnicity. Whites are the ethnic group with the most highly segregated commuting patterns (28.2 percent), followed by blacks (50.3 percent), and Latinos (86.0 percent). The black-white trip interaction index (42.1 percent) is twice as large as the white-black trip interaction index (18.8 percent), which means that black workers are much more to cross paths with whites on their commuting routes than vice versa. Especially, the high values of the Latino/Asian-white trip interaction indices (64.4 percent and 58.1 percent, respectively) result from the relatively low percentages of Latinos (7.6 percent of residents, and 6.6 percent or workers) and Asians (3.9 percent of residents, and 3.4 percent of workers) in the four counties. This small numbers of Latinos/Asians, as a percentage of the total population, leads to higher potential contact between groups, regardless of how they are distributed in the study area. Analysis of segregation in spaces of home, work, and JTW shows that residential space in these four counties (i.e., Cobb, DeKalb, Fulton, and Gwinnett) of Atlanta is much more highly segregated than employment or commuting space.
### Table 4-3. Interaction indices of home, work, and JTW

<table>
<thead>
<tr>
<th>Race (From/To)</th>
<th>Home</th>
<th>Job</th>
<th>JTW</th>
<th>Race (From/To)</th>
<th>Home</th>
<th>Job</th>
<th>JTW</th>
</tr>
</thead>
<tbody>
<tr>
<td>White / Others</td>
<td>0.247</td>
<td>0.347</td>
<td>0.282</td>
<td>Latino / Others</td>
<td>0.788</td>
<td>0.919</td>
<td>0.860</td>
</tr>
<tr>
<td>/ Black</td>
<td>0.140</td>
<td>0.247</td>
<td>0.188</td>
<td>/ White</td>
<td>0.499</td>
<td>0.625</td>
<td>0.581</td>
</tr>
<tr>
<td>/ Latino</td>
<td>0.065</td>
<td>0.066</td>
<td>0.061</td>
<td>/ Black</td>
<td>0.234</td>
<td>0.257</td>
<td>0.238</td>
</tr>
<tr>
<td>/ Asian</td>
<td>0.043</td>
<td>0.034</td>
<td>0.034</td>
<td>/ Asian</td>
<td>0.055</td>
<td>0.036</td>
<td>0.042</td>
</tr>
<tr>
<td>Black / Others</td>
<td>0.351</td>
<td>0.648</td>
<td>0.503</td>
<td>Asian / Others</td>
<td>0.923</td>
<td>0.957</td>
<td>0.945</td>
</tr>
<tr>
<td>/ White</td>
<td>0.268</td>
<td>0.554</td>
<td>0.421</td>
<td>/ White</td>
<td>0.631</td>
<td>0.620</td>
<td>0.644</td>
</tr>
<tr>
<td>/ Latino</td>
<td>0.058</td>
<td>0.061</td>
<td>0.056</td>
<td>/ Black</td>
<td>0.186</td>
<td>0.267</td>
<td>0.218</td>
</tr>
<tr>
<td>/ Asian</td>
<td>0.024</td>
<td>0.033</td>
<td>0.026</td>
<td>/ Latino</td>
<td>0.106</td>
<td>0.070</td>
<td>0.083</td>
</tr>
</tbody>
</table>

**Sensitivity Analysis – Tract vs. Small Places**

Sensitivity analysis is performed to check how well our model deals with sensitivity of the size of area units. For instance, the larger area units tend to have the smaller the value of segregation score. We apply two different boundaries – census tract and “small place,” as defined by the U.S. Census. There are two or three “small places” within each census tract, and a tract is thus generally twice or third the size of a small place. In our study, there are 440 census tracts and 1,242 small places listed by the metropolitan Planning Organization (MPO) in the four counties in Atlanta. As shown in Table 4-4, the isolation index 1 is based on Equation 1, and the isolation index 2 is based on Equation 8. The isolation index 1 score between tract and small places is lower than that of the isolation index 2. The isolation index 2 in Equation 8 considers all possible spaces between origin and destination, treating spaces from all inbound trips, from both residence and workplace areas. By contrast, the isolation index 1 in Equation 1 only considers the space between origin and destination. Thus, the trip exposure indices we provide can help to deal with the problems of sensitivity of the size of area units.
Table 4-4. Sensitivity analysis of isolation index based on tract and small place

<table>
<thead>
<tr>
<th>Race (From / To)</th>
<th>Trip Isolation Index 1</th>
<th>Trip Isolation Index 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Tract</td>
<td>Place</td>
</tr>
<tr>
<td>White / White</td>
<td>0.784</td>
<td>0.840</td>
</tr>
<tr>
<td>Black / Black</td>
<td>0.645</td>
<td>0.711</td>
</tr>
<tr>
<td>Latino / Latino</td>
<td>0.213</td>
<td>0.327</td>
</tr>
<tr>
<td>Asian / Asian</td>
<td>0.083</td>
<td>0.135</td>
</tr>
</tbody>
</table>

Note: Index 1 is the trip isolation index based on Equation 1.
Index 2 is the trip isolation index based on Equation 8.
Tract refers to census tracts, and Place refers to small places.
Difference is the ratio between census tracts and small places.

Visualization of Trip Isolation

Figure 4-4 shows trip isolation by race/ethnicity. AB is the trip isolation from origin (home) to destination (workplace). BA is the trip isolation from origin (workplace) to destination (home). Due to the difficulty of visualization of trip isolation by tracts (leading to 193,600 pairs of trips) on the map, we aggregate pairs of trips in census tracts to those in super districts, which are sub-county geographic areas defined by the Atlanta Regional Commission (ARC). Super districts are larger than census tracts, but smaller than counties. Commuting trips between 32 different pairs of super districts have 1,024 possible directions (inbound and outbound flows), which lead to the sum total of trip isolation as 1.0. The result is that each trip scores very low on the trip isolation index because each 1,024 possible scores will make the sum total 1.0, making it difficult to compare levels of segregation between trips. The maximum value of outbound trip isolation in this case is 1/1000 (0.001), which implies a very low rate of segregation. Even so, Figure 4-4 tells us how trip isolation differs by race/ethnicity in general. For instance, there exists some black trip isolation clustering in South Atlanta, compared to
Hispanic trip isolation clustering in the city of Atlanta. White trip isolation mostly occurs in North Atlanta, and Asian trip isolation exists in North Fulton.

If we focus upon black trip isolation (See Figure 4-5), the trip isolation index helps us to understand how commuting segregation relates to residential and employment segregation. Whereas black workers mostly reside in South Atlanta (on the left map), they mostly work in North Fulton and DeKalb, as well as in the City of Atlanta itself (on the right map). Overlaying black trip isolation scores as straight lines on residential and employment segregation maps, it reveals that black workers create their own isolated clusters in the limited residential and employment space which is occupied by them, but that black trip isolation is much more likely to be the result of residential as opposed to employment segregation.
CONCLUSIONS

The majority of past research concerning the creation and analysis of segregation indices has mostly dealt with segregation in residential space (Massey and Denton 1987, Morgan 1983, Massey and Denton 1988, Iceland and Wilkes 2006, Dawkins 2004), employment space (Aslund and Skans 2010, Leonard 1987, Horner and Marion 2009) and school space (Gorard 2009, Frankel and Volij 2011, Allen and Vignoles 2007). However, the approaches above may provide misleading results where segregation takes place between residence and workplace instead of either in residence or workplace. To tackle the issue, we provided a new measure of commuting segregation, namely the trip isolation/interaction index, which calculates segregation between a reference space, i.e., occurs in a pair of trips between home and work, and a control space, i.e., occurs in all
inbound trips from home to workplace, instead of treating the space as a pair of trips between them.

By applying this approach to the analysis of four metro Atlanta counties, our results have confirmed that segregation also takes place in commuting space as well as in residential and employment spaces. The results show that whites tend to have the most highly segregated commuting patterns, followed by blacks, Latinos, and Asians. Among minority groups, Asians are much more likely to share their commuting space with members of other ethnic groups, while blacks had the most highly segregated commuting patterns of any minority group. Sensitivity analysis was performed to check the sensitivity of our model to the relative size of boundaries, e.g., tracts and small places. The results confirmed that trip isolation and interaction indices, which consider all inbound trip space associated with both origin and destination, are performed well in comparison with the traditional trip exposure indices, which only consider the space between origin and destination. Using black trip isolation as an example, we found that black commuting segregation is much more likely to be the result of residential segregation than employment segregation. In other words, the concentration of blacks into a limited amount of residential space is the primary cause of commuting segregation in certain areas such as the city of Atlanta and South DeKalb.

Due to the huge amount of trip data (440 tracts leading to 193,600 pairs of trips), the visualization of trip isolation/interaction presents a significant challenge. Although we provided trip isolation maps, using the super district as the common unit, it is still difficult to determine which pairs of trips are most likely to cross paths with other pairs of trips. For future research, it would be interesting to visualize the trip exposure index in a
web-based context with a dynamic query. By applying this model, a user will easily be able to investigate commuting segregation within specific areas, as along with rates of residential and employment segregation. Plus, the 2010 census transportation planning package (CTPP) will be released in early 2013. Thus, it would be also interesting to compare commuting segregation in 2000 with that of 2010, and to determine which types of commuting segregation have increased or decreased over space and time.
REFERENCES


Li, T., J. Corcoran & M. Burke. 2010. Investigating the changes in journey to work
patterns for South East Queensland - a GIS based approach. In Australasian
Transport Research Forum ATRF Paper Archive.

cities.

Reviews, 26, 749-767.

Massey, D. S. & N. A. Denton (1987) Trends in the residential segregation of Blacks,

Forces, 67, 281-315.


Morgan, B. S. (1983) A distance-decay based interaction index to measure residential
segregation. Area, 15, 211-217.

excess commuting and jobs -- housing balance. Environment & Planning A, 37,
2233-2252.

1215.

Pirie, G. H. (1979) Measuring accessibility: a review and proposal. Environment and
Planning A, 11, 299-312.


CHAPTER 5
CONCLUSIONS AND FUTURE RESEARCH

CONCLUSIONS

This dissertation investigates commuting patterns by ethnicity over space and time. It deals with two fundamental difficulties presented by the organization of data in the Census Transportation Planning Package (CTPP), namely inconsistent boundaries and unavailability of trip data by subgroups. It also explores the issue of racial segregation related to commuting patterns. Through applications of the methods developed in this dissertation, researchers can achieve a more in-depth understanding of spatial mismatch (separation) between jobs and housing over space and time.

The CTPP data are inconsistent with respect to geographic boundaries such as census tracts and traffic analysis zones (TAZs) over time. In response to this problem, the research in Chapter 2 proposes the concept of flow line interpolation and develops first models for it, namely the areal-weighted and the intelligent flow line interpolation models. The model aligns trip data on inconsistent spatial units (e.g. census tracts vs. TAZs) to a common spatial framework. This research is innovative and significant because it takes an important step in matching and analyzing inconsistent commuting data. It enables researchers to further investigate the spatiotemporal changes of commuting patterns.
Enabled by the developed flow line interpolation models, chapter 3 analyzes spatiotemporal changes in commuting patterns by ethnicity over time in Atlanta. To measure commuting trips by ethnicity, this research applies the aforementioned two-step analysis, which combines information minimization and flow interpolation methods. Of particular interest are the commuting patterns of the Latino population because the rapid influx of this group has resulted in a marked change in the composition of the population in metro Atlanta. A case study is performed to track changing commuting patterns for whites, blacks, and Latinos in metro Atlanta over the last two decades (1990-2008). The results show that ethnicity has been a significant factor in job-housing relationships in metro Atlanta. This research is significant for its contribution to our understanding of the spatial dynamics of jobs/housing relationships over space and time. Unlike the traditional dichotomous approach (black-white), by highlighting the commuting patterns of the Latino population, the two-step analysis performed here can give a more realistic representation of the spatial mismatch of jobs and housing by ethnicity. Still, this research has some limitations. First, it focuses on commuting trips by one category (ethnicity), and provides insufficient understanding of how multiple variables affect commuting patterns in U.S. metro areas. For example, it would be interesting to investigate trip patterns by gender and income level. Second, the accuracy of the interpolation method is sensitive to the spatial granularity levels of the source and target zonations. Future research can be conducted to consider more about the sensitivity and to develop ways in coping with it.

As informed by the research in Chapter 3, commuting patterns are divided by ethnicity. Chapter 4 focuses on the analysis of racial or ethnical segregation in the
commuting space. Unlike traditional segregation studies which focus exclusively on segregation in spaces of residence or employment, this research approaches the segregation in the commuting space which couples job and housing spaces. The research designs a so-called trip exposure index to assess the level of racial segregation between a reference space, i.e., occurs in a pair of trips between home and workplace, and a control space, i.e., occurs in all inbound trips from home and to workplace, instead of treating the space as a pair of trips between them. The new index can deal with spatial segregation where commuting segregation exists in job and housing patterns. Sensitivity analysis confirms that the new index, which considers the space of all inbound trips, both in terms of origin and destination, is capable of making adjustments to the relative size of boundaries. This study is significant to our understanding of urban racial segregation when coupled job-housing spaces are considered. Overlaying the commuting segregation index map onto the traditional residential/employment segregation map can gives us a more comprehensive evaluation of segregation. This study still has a limitation due to the usage of a one-time snap shot (year of 2000). With forthcoming 2010 CTPP (expected in early 2013), the trip exposure index can be applied again to identify which commuting segregation scores have increased or decreased in metropolitan areas in the U.S..

Lastly, previous studies are unable to undertake a spatiotemporal trajectory analysis of commuting patterns due to data inconsistency and technical barriers presented by the CTPP. To overcome these problems, this research has developed the following methodologies: the flow line interpolation, the two-step analysis of commuting patterns, and the commuting segregation index. Through application of these methods, this dissertation contributes to efforts at the detailed examination of changing commuting
patterns by race/ethnicity, with particular focus on Latinos, while also developing techniques for the measurement of commuting segregation. Some potential applications of methods developed and used in this dissertation can be extended to other domains such as air traffic, the migration and movement/consumption of goods. The scope of this dissertation can be also applied to other targets such as (a) the ageing baby-boomer generation, as trip analysis applied to this group can be used to measure access to quality-of-life related amenities and resources, and (b) cities with far lower indices of segregation.

FUTURE RESEARCH

Future research will aim at three categories, namely: (1) close-up investigation of commuting trips by multiple categorical factors, e.g., low-income Latino trips with the 2010 CTPP data, (2) dynamic visualization of commuting data on the web, and (3) application of the methods (flow interpolation, information minimization, and commuting segregation index) to deal with analysis of the changing commuting patterns of the ageing baby-boomer generation.

First, the research in Chapter 3 only dealt with one variable, i.e., ethnicity. Unique patterns of commuting trips can also be analyzed with combinations of variables, such as ethnicity and income level, job type, and mode of transportation. Thus, it will be interesting to apply the two-step approach again in order to reveal complex commuting patterns by use of multiple variables in combination. For instance, if we investigate the patterns by ethnicity and income, we may understand better whether income has a greater effect on commuting patterns than ethnicity or vice versa. Such results will enhance our
understanding of the dynamics of the jobs-housing relationship and consequently assist in
decision making when dealing with urban problems. Combined with multiple categorical
factors, the two-step approach can be re-applied with the new 2010 CTPP data, which is
said to be due for release early next year. It will provide commuting data on a finer scale
(e.g., TAZs) than that were used in this dissertation research. The new TAZ CTPP will
allow us to perform analysis at a higher level of accuracy than that achieved in chapter 3
of this dissertation, when comparing commuting patterns by ethnicity during the last two
decades (1990-2010).

Second, the huge amount of commuting data makes it very difficult to present for
visualization on a map. For instance, in 2000 metro Atlanta contained 660 census tracts,
resulting in a total of 435,600 possible pairs of trips. The current visualization of trip data
mostly uses a straight line, e.g., Euclidean distance, between origin and destination.
Visualization is difficult with this approach, as pairs of trips overlap one another. Some
research has applied a shortest-path algorithm for assigning commuting trips on the
network (Li et al. 2010). A problem with this approach is that commuting trips do not
always take the shortest path. To overcome this challenge, future research will develop a
web-based commuting model through a dynamic query and GIS framework. Instead of
showing all pairs of trips, a map in this model can be shown according to the user’s
choice, e.g., Latino trips in the center of the city or female trips in suburban areas. Thus,
this model will be able to investigate commuting trips by demographic or socio-economic
variables efficiently and effectively.

Third, the methodological framework in this dissertation can be further applied to
analysis of access to quality-of-life related amenities and resources, as well as emergency
preparedness, of the ageing baby boomer generation. The increasing numbers of baby 
boomers (born between 1946 and 1964) reaching retirement age, combined with longer 
life spans, will profoundly impact the demographic and socioeconomic structures of 
American society (Goulias et al. 2007, Gray-Graves, Turner and Swan 2011, McGuigan 
2010, Miranda-Moreno and Lee-Gosselin 2008, Pickett 2007, Yan, Shuang Yann and 
Tao 2009). However, urban sprawl leads to difficulties for aging people in accessing 
social services, recreation, and shopping centers where mass transit is not available. 
Connecting the target population to such facilities is one of the key issues in improving 
their quality of life. It will also be crucial for planners to develop efficient approaches for 
dealing with emergency preparedness for the aging populations. For instance, rapid 
response time is crucial for the survival of strokes, mostly suffered by the elderly. By 
applying network analysis combined with the methods developed in this dissertation, we 
can further develop methods for determining the shortest and fastest routes between the 
target populations and essential health care facilities.
REFERENCES


APPENDIX

LIST OF ACRONYMS

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACS</td>
<td>American Community Survey</td>
</tr>
<tr>
<td>ARC</td>
<td>Atlanta Regional Commission</td>
</tr>
<tr>
<td>CTPP</td>
<td>Census Transportation Planning Package</td>
</tr>
<tr>
<td>DOT</td>
<td>Department of Transportation</td>
</tr>
<tr>
<td>IM</td>
<td>Information Minimization</td>
</tr>
<tr>
<td>IRCA</td>
<td>Immigration Reform and Control Act</td>
</tr>
<tr>
<td>JTW</td>
<td>Journey to Work</td>
</tr>
<tr>
<td>LULC</td>
<td>Land Use and Land Cover</td>
</tr>
<tr>
<td>MARTA</td>
<td>Metropolitan Atlanta Rapid Transit Authority</td>
</tr>
<tr>
<td>MAUP</td>
<td>Modifiable Area Unit Problem</td>
</tr>
<tr>
<td>MPO</td>
<td>Metropolitan Planning Organization</td>
</tr>
<tr>
<td>MRLC</td>
<td>Multi-Resolution Land Characteristics</td>
</tr>
<tr>
<td>OD</td>
<td>Origin-Destination</td>
</tr>
<tr>
<td>PUMA</td>
<td>Public Use Microdata Area</td>
</tr>
<tr>
<td>RMSE</td>
<td>Root Mean Square Error</td>
</tr>
<tr>
<td>SMH</td>
<td>Spatial Mismatch Hypothesis</td>
</tr>
<tr>
<td>SQL</td>
<td>Structure Query Language</td>
</tr>
<tr>
<td>T</td>
<td>TAZ</td>
</tr>
<tr>
<td>----</td>
<td>---------</td>
</tr>
<tr>
<td>V</td>
<td>VBA</td>
</tr>
</tbody>
</table>