ACCESSIBILITY AND BUILT ENVIRONMENT EFFECTS ON TRANSIT USE

ACCESSIBILITY AND BUILT ENVIRONMENT EFFECTS ON TRANSIT USE

By MD. MONIRUZZAMAN, Bachelor of Urban and Regional Planning

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Arts

MASTER OF ARTS (2011) (School of Geography and Earth Sciences)	McMaster University Hamilton, Ontario, Canada

TITLE	Accessibility and Built Environment Effects on Transit Use
AUTHOR	Md. Moniruzzaman, Bachelor of Urban and Regional Planning (Bangladesh University of Engineering & Technology)
SUPERVISOR	Dr. Antonio Páez, Associate Professor School of Geography and Earth Sciences
NUMBER OF PAGES	viii, 73

ABSTRACT

A critical factor in transit mode share analysis is the level of accessibility to transit services. The objective of this study is to investigate the relationships between accessibility and the built environment, and the use of transit. To do so, the proportion of transit users is modeled as a function of socio-economic characteristics, transit level of service, and built environment characteristics. While accessibility to transit has been the object of previous research, accessibility *by* transit is a factor that has received only limited attention in prior transit modal share research. The case study is the city of Hamilton, and the geographic unit of analysis is the Dissemination Area. A logistic model for proportions with a spatial filter (for spatial autocorrelation) and an over-dispersion parameter is found to provide the best fit and statistical properties. The results of analysis at the meso-scale show that accessibility by transit contributes positively to usage of transit. The possibility that factors at the micro-scale may also influence use of transit, suggests the development of a methodology, based on the use of spatial filtering, to systematically select walkability audit sites. The proposed methodology is demonstrated by means of a case study of neighborhoods in Hamilton. Statistical analysis of walkability audit information shows that the proposed selection strategy can be used to better target limited resources for field-based work, and produce valuable insights into the micro-level factors that may affect transit use.

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to Dr. Antonio Páez for his continuous supervision, guidance, motivation and support over the entire duration of my Master program in McMaster University towards successful completion of my thesis.

I would also like to acknowledge Dr. Darren M. Scott for his expert opinion about the technology used to implement walkability audit tools in this study. In addition, I would like to thank Cathy Moulder, Director of Library Services, Maps, Data and GIS at McMaster University for providing the most updated geographic files to carry out this study. I would also like to thank Jim Dahms, Manager, Transit Planning, Hamilton Street Railway Company for providing transit frequency information.

I would like to give special thank Ron Dalumpines, PhD candidate in the School of Geography and Earth Sciences for helping me in various GIS problem solving. I would also like to thank all my friends and colleagues from my lab for their support throughout the course of my graduate program. I also would like to express my sincere gratitude to my parents, who always encouraged and supported me in all my difficulties. Lastly, the support from my wife, Jarin Ahsan Esita is priceless and without her support it would be impossible to accomplish many achievements.

TABLE OF CONTENTS

Abstract..iii

Acknowledgements iv

Table of contents v

List of tables vii

List of figures viii

Chapter 1 Introduction.. 1

1.1. Built environment and travel behaviour 1

1.2. Research challenges... 4

1.3. Objectives of the study 5

1.4. Organization of the thesis 5

Chapter 2 Accessibility to Transit, by Transit, and Mode Share.......................... 7

2.1. Introduction 7

2.2. Background.. 10

2.3. Case study and data 13

2.4. Methods 18

  2.4.1. Measuring accessibility by transit 18

  2.4.2. Logistic model for proportions.. 18

  2.4.3. Over-dispersion and spatial autocorrelation.. 20

  2.4.4. Eigenvector-based spatial filtering.. 22

2.5. Results and discussion... 26

2.6. Conclusions 31

Chapter 3 Selecting Case Sites for Walkability Audits...................................... 33

3.1. Introduction 33

3.2. Prior research... 35

3.3. Methods 39

3.4. Case study.. 42

  3.4.1. Context 42

  3.4.2. Selection of sites.. 44

3.4.3. Audit instrument and technology 46
3.4.4. Walkability audits.. 47
3.5. Results and discussion... 48
3.6. Summary and conclusions.. 59
Chapter 4 Conclusions... 62
References 66

LIST OF TABLES

Table 2.1 Variable definition and descriptive statistics 17
Table 2.2 Logistic regression parameter estimation in three different specifications 28
Table 3.1 Results of logistic model for proportion of commuting trips by transit 43
Table 3.2 Audit items and summary of independence tests 50
Table 3.3 Uses in segment: contingency table 52
Table 3.4 Degree of enclosure: contingency table 53
Table 3.5 Setbacks: contingency table 53
Table 3.6 Building height: contingency table 54

LIST OF FIGURES

Figure 3.1 Example of a generic model 40
Figure 3.2 Examples of random and spatially autocorrelated residuals 41
Figure 3.3 Paired control cases 42
Figure 3.4 Spatial filter with audit locations 45
Figure 3.5 Electronic version of PEDS 47
Figure 3.6 Low volume street segment (top) and high volume street segment (bottom).... 51
Figure 3.7 Examples of street segments with mixed land uses (residential and commercial), high enclosure and small setbacks 55
Figure 3.8 Examples of street segments with single land use, low enclosure, large setbacks, and low-rise buildings 56

Chapter 1 Introduction

1.1. Built environment and travel behaviour

Studies pertaining to the association of built environment and travel behaviour have been a topic of great interest over the past decade (TRB and Institute of Medicine, 2005). Transport researchers have strived to answer the question of how different forms of urban development affect the travel behaviour of individuals. More specifically, a question of interest has been to figure out what type of urban design can promote walking and transit use and thereby reduce automobile dependence.

From this point of view, the concept of development that is capable of accommodating different forms of travel has been raised. This includes neo-traditional design, Transit Oriented Design, and Smart Growth. Neo-traditional design is now a part of the larger movement known as The New Urbanism. The New Urbanism is an international movement to raise our quality of life and standard of living by creating or altering the design of built environment into more compact, mixed-use, walkable communities composed of “the same components as conventional development, but assembled in a more integrated fashion, in the form of complete communities. These contain housing, work places, shops, entertainment, schools, parks, and civic facilities essential to the daily lives of the residents, all within easy walking distance of each other” (Demiroz, 2005, p. 22).

Smart Growth (also known as Compact City) is a general term with relatively similar concept of the New Urbanism. Smart Growth offers policies that concentrate urban growth as more mixed used, compact, walkable, and transit oriented land use development. Smart Growth policies are exercised to make a more vibrant and liveable community by making the community more competitive for new businesses, providing alternative places for shopping, and working, creating a better sense of place, providing more jobs for its residents, increasing property values, preserving open space, controlling growth, improving safety, and, overall, improving quality of life. The difference between Smart Growth and New Urbanism is in the geographic scale, for instance, Smart Growth policies are more applicable at the regional scale whereas New Urbanism is at the neighborhood scale.

Transit Oriented Development (TOD) is a particular category of New Urbanism and Smart Growth with more emphasis on transit use, walking, and cycling. It can be defined as residential and commercial centers designed to maximize people's access by transit and non-motorized transportation, and with other features to enhance transit ridership. It is a simple way to shift some of the car trips to transit by land use mixing and non-motorized transportation improvements. High quality transit service has the potential to benefit accessibility and agglomeration in high density urban centers which supports the function of these regions. For example, activities/sites may be reached by walking or cycling if they are physically closer, and relatively distant activities may be reached by public transit. On the other hand, automobile dependent transportation conflicts with high density urban areas as they are more space sensitive and require more land for roads and parking facilities (Boroski et al., 2002; Voith, 1998).

There is a belief that adopting the form of compact and mixed-use design advocated by Transit Oriented Design, New Urbanism, and Smart Growth, will result in increased pedestrian activity and transit use, and decreased dependency on the automobile. Previous studies (e.g. Hsiao et al. (1997)) already indicate that areas with higher land use density and grid street patterns provide better pedestrian access to transit than areas with irregular street patterns and lower land use density.

In addition to the issues typically associated with automobility, such as congestion, delays, and pollution, in recent years, the public health importance of physical activity has added urgency to the study of the relationships between the built environment and travel behaviour. Active travel (walking and cycling), and walking as a complement to using transit, are seen as important elements of active lifestyles. Active transport has potential health benefits in terms of reducing the risk of developing obesity, osteoporosis, depression, and some other chronic diseases (Gilmour, 2007; Tjepkema, 2006; Warburton et al., 2006). Suitable built environments are recognized as an important pre-condition for people to become more active. On the other hand, decentralized land use patterns push people towards the use of the automobile and reduced physical activity levels. There is a need, therefore, to develop a better understanding of the factors that influence travel behaviour, in particular with respect to use of transit, walking, and cycling.

1.2. Research challenges

Previous studies of the built environment and travel behaviour, more specifically transit shares, have been based on geo-referenced information. This allows researchers to consider spatial variables related to the built environment, and spatial relationships such as accessibility to transit. A factor that has received relatively limited attention is accessibility by transit, the main output of the transit system. In addition to providing opportunities, the use of spatial data also presents some analytical challenges, including the presence of spatial autocorrelation. Spatial autocorrelation is known to inflate or deflate the significance of regression coefficients, a fact compounded in models of proportions by over-dispersion. The cumulative effect leads to weaker, potentially misleading estimates of the travel behaviour. In this study, transit modal shares are modeled as a function of accessibility by transit in addition to local accessibility, built environment, and socio-economic variables. Furthermore, the effect of autocorrelation is treated using a spatial filtering approach while accounting for over-dispersion.

Modeling transit shares at the level of statistical areas can provide important information about the influence of accessibility and the built environment on travel behaviour at that scale of analysis. However, there is increasing interest on the effect of environments at more micro-level (i.e. pedestrian) scales. A challenge to develop a better understanding of micro-level environments is that data are seldom collected in a sufficiently broad and systematic fashion, and purposive data collection efforts tend to be time-consuming and resource-intensive. The use of a spatial filtering approach to model transit shares suggests the development of a method to systematically select walkability audit sites, in order to collect information about pedestrian-scale design features. The method can be applied to reduce the cost and duration of the project.

1.3. Objectives of the study

The objectives of this thesis are the following:

1) To assess the association of built environments and transit modal share.

2) To incorporate global accessibility (i.e. accessibility by transit) as a variable in the model along with transit level of service, built environments, and socio-economic characteristics.

3) To apply the concept of eigenvector based spatial filter to capture the effect of spatial autocorrelation as well as control for over-dispersion parameter for providing sharper inference to the model.

4) To develop an approach, based on the use of spatial filter, to systematically select walkability audit sites in order to collect pedestrian-scale design information.

1.4. Organization of the thesis

The remainder of this thesis is organized as follows.

Chapter 2 presents the modeling of transit modal share, using Dissemination Areas as the unit of analysis, against various socio-economic, transit level of service, and built environments characteristics. The case study is the City of Hamilton, in Canada. Substantively, the focus of the study is on accessibility by transit. Technically, the focus is on capturing spatial autocorrelation and reducing the deleterious effects of over-dispersion. The results show that the model can provide sharper inference than alternatives that ignore either of these two statistical issues. Finally, more reliable results regarding accessibility by transit are generated, while the influence of some built environmental variables is disproved.

The focus of Chapter 3 is to propose a method to systematically select walkability audit sites. Walkability audits aim at obtaining information about micro-scale built environments that may have been missed by the meso-scale model, conducted at the level of statistical areas in Chapter 2. The core of the method is the spatial filtering approach used to control spatial autocorrelation. The filter, it is argued, captures relevant but omitted variables that follow a spatial pattern. Eight dissemination areas are selected using the proposed model-based strategy, and walkability audits conducted to collect the micro-scale design features of the selected areas. Statistical analysis (contingency tables and chi-squared tests of independence) are used to investigate the potential differences between areas with larger or smaller transit shares than expected according to the model.

Chapter 4, the last of this thesis, summarizes all the findings, clearly identifies the contributions of the thesis, and presents recommendations for future research.

Chapter 2 Accessibility to Transit, by Transit, and Mode Share

2.1. Introduction

Transit systems play an important role in urban public policy. They are increasingly seen as an essential element in policy packages that aim to reduce congestion, make more efficient use of road space, reduce pollution, and keep the lid on increasing energy consumption in the transportation sector (May et al., 2000). As such, they appear to be key to achieve overall environmental, economic, and social sustainability objectives.

Despite their potential contributions to urban policy objectives, transit systems suffer from the effects of a vicious circle. Falling patronage reduces revenue, and makes it more difficult for providers to offer a competitive transportation alternative without substantial subsidies. This vicious circle has in part been facilitated by urban development patterns that favour private mobility. Although these patterns have been evident for most of the past half century in North America, public transportation in Canada did not require subsidies until the 80s. Until 1975, revenue exceeded operating costs; for instance, in 1975, revenue was more than twice the cost of operation. This situation changed in the following decades. In 1985, the direct operation cost was slightly greater than the revenue and it continued to become worse in 1995 when direct operation cost was nearly double than the revenue. In 2005 operating costs in Canada exceeded revenue by a factor of almost three (American Public Transit Association, 2010). Despite substantial and increasing subsidies, transit ridership declined across the country from 1991 and finally stabilized in the later years of the decade (Statistics Canada, 1999). Overall, the market share of transit travel across Canada, the proportion of workers who used the bus or subway to get to and from work, remained steady at about 12% between 1992 and 2005.

Several reasons can be advanced to explain this low uptake of public transit. A critical factor in transit mode share analysis is the level of accessibility to transit services (Murray, 2003; Ziari et al., 2007). Accessibility denotes the ease with which a given destination can be reached from an original location using a particular form of mobility (Handy and Niemeier, 1997). Accessibility to transit facilities has been explored in the past. This includes examinations of the built environment and levels of service of transit (Estupiñán and Rodriguez, 2008; O'Sullivan and Morrall, 1997; Rodriguez et al., 2009). Urban form is also believed to have an impact, since density thresholds must be met before transit becomes a feasible proposition (Cervero and Kockelman, 1997). An additional factor that remains to be more fully explored in the literature on transit patronage and modal shares is accessibility by transit. Analysis of mode shares with a focus on accessibility has tended to consider primarily the local conditions at the point of origin of trips, in other words, the ease of entering the system. This ease, it is contended, must be contrasted against the ease of reaching attractive destinations, in other words, the accessibility produced by the transit systems. In the case of commuting, this means the potential of reaching jobs using public transit. It is proposed then that while being able to find transit facilities locally is important (i.e. accessibility to transit), the places and opportunities that can be reached by transit (i.e. accessibility by transit) is also an important factor that has only recently begun to receive attention.

In this chapter, the implications of accessibility to transit and by transit for mode shares in the city of Hamilton, Canada, are investigated. The transit provider in the city is Hamilton Street Railway (HSR), and data are collected to indicate mode shares for the case of commuting to work. Socio-economic and demographic population attributes, transit level of service, and built environment variables are the explanatory variables in this study. Of particular interest for this research are the effects of accessibility to transit, and accessibility by transit, defined as the number of jobs that can be reached using HSR services. Modal share is analyzed in this research using a generalized linear (i.e. logistic) model for proportions.

This chapter makes the following contributions. Firstly, 'accessibility by transit' is calculated using a geographical information system; this variable is introduced as an explanatory factor in logistic regression models of modal share. Secondly, logistic regression is sensitive to the presence of over-dispersion, which makes the estimated variance to be artificially small, which leads to misleading inference. This condition may be caused to some extent by lack of independence. Thus, the second contribution of the chapter is to estimate a model that jointly considers over-dispersion and spatial autocorrelation, a manifestation of lack of independence in spatial data. Treatment of spatial autocorrelation is by means of eigenvector-based spatial filter (Griffith, 2000, 2003, 2004), used to remove autocorrelation from the residuals of logistic regression model. The results from the analysis in this chapter suggest that estimating an over-dispersion parameter alleviates to some extent the inflation of the variance, but that resolving the autocorrelation situation leads to even sharper inference. This, it is argued, has important implications for model-based policy prescriptions.

2.2. Background

The factors that influence transit patronage and shares have been examined in previous research. These factors include the effect of demographic and socio-economic characteristics of travelers. With the exception of Loutzenheiser (1997) who found that higher income persons using BART in California are more likely to walk to transit stations to commute to central city jobs, income has mostly been shown to have a negative association with transit ridership (Hess, 2009; Hsiao et al., 1997; Taylor and Miller, 2003). A negative association with income was also found by Greenwald and Boarnet (2001) when regressed against non-work walking trips and by Cervero and Kockelman (1997) with non-work trips by all means except non-single occupant vehicle and non-work trips by non-personal vehicle.

Improving the level of service of transit has been shown to attract more commuters towards transit. Presence of transit stops, headway/service frequency, route density, presence of dedicated bus routes, and availability of bus schedule have been considered in previous transit ridership models. Most studies in general indicate a positive relationship of these factors with ridership or walking to transit stations (Alshalalfah and Shalaby, 2007; Cervero et al., 2009a; Cervero et al., 2009b; Johnson, 2003; Taylor et al., 2009). Taylor and Miller (2003) also show that service quality variables are positively and significantly related to ridership.

In addition to the level of service, a topic of current interest in the literature is the association between accessibility to transit services and use of transit. Accessibility to transit is influenced by factors relating to the built environment, usually described in terms of density, diversity, and design, i.e. the 3Ds of built environment (Cervero and Kockelman, 1997). The available evidence indicates that higher densities support public transit more than lower densities (Cervero et al., 2009a; Johnson, 2003; Taylor and Miller). Higher mixture of land use (Cervero et al., 2009b; Estupiñán and Rodriguez, 2008; Ewing, 1995; Frank and Pivo, 1995; Kitamura et al., 1997; Loutzenheiser, 1997), vertical mixed-use close to transit services (Johnson, 2003), and higher land use density (Greenwald and Boarnet, 2001; Hsiao et al., 1997) tend to support travel choices that include transit, walking, and car sharing. Accessibility to transit is enhanced by design factors, such as the presence of sidewalks, sidewalk width, sidewalk density, street density, and intersection density (e.g. Cervero and Kockelman, 1997; Cervero et al., 2009b; Guo, 2009; Hess et al., 1999; Kitamura et al., 1997; Rodriguez and Joo, 2004). Likewise, walking distance to and from the transit stop is negatively associated with transit ridership and pedestrian access to transit (Alshalalfah and Shalaby, 2007; Cervero et al., 2009a; Greenwald and Boarnet, 2001; Hess, 2009; Kitamura et al., 1997; Loutzenheiser, 1997; O'Sullivan and Morrall, 1997; Olszewski and Wlbowo, 2005; Sanchez, 1999; Zhao et al., 2003). Alshalalfah and Shalaby (2007) have also found that transit users are willing to walk further to access subway lines than to access bus/streetcar routes.

Finally, of particular interest for this research is accessibility by transit, defined as the number of jobs within a determined time frame that can be accessed by transit.

Accessibility by transit has been extensively studied in the past. For instance, the effect of job accessibility by transit on employment participation (e.g. duration of welfare, employment status, and employment retention) has been studied, among others, by Sanchez (1999), Thakuriah and Metaxatos (2000), Cervero et al. (2002), Sanchez et al. (2004), Yi (2006), Alam (2009), and Matas et al. (2010). Transit-based accessibility has been compared to accessibility by car in other research, including Blumenberg and Ong (2001), Shen (2001), Hess (2005), Kawabata and Shen (2006, 2007), Kawabata (2009), and Benenson et al. (Benenson et al., 2010). Decision support systems have been developed to help improve accessibility by transit, for example, by Grengs (2004), Kamatu et al. (2007), Minocha et al. (2008), and Lei and Church (2010). On the other hand, accessibility by transit as a factor that influences transit ridership and shares has only recently begun to receive attention. For instance, Chow et al. (2007) found that the percentage of commuters in traffic analysis zones traveling by transit increases as regional accessibility to employment increases. Brown and Thompson (2008), in their study of transit patronage in Atlanta and its relationship to employment decentralization, found that the decline in transit patronage can be reduced by improving accessibility to decentralized jobs. Finally, Yanmaz-Tuzel and Ozbay (2010) found a positive relationship between city-wide monthly employment and transit ridership in New Jersey.

In the following analysis, the relationship between socio-economic and demographic variables, accessibility to transit, and accessibility by transit, on transit shares in Hamilton is investigated by using logistic regression.

2.3. Case study and data

Transit ridership varies significantly across urban areas in Canada. In large urban areas where the service is more accessible to commuters, the proportion of commuter by transit tends to be higher. Approximately 20% of workers in Canada's six largest metropolitan areas used the bus or subway for part or all of their commute in 1992 and 2005 (Statistics Canada, 2006). The trends, moreover, have been positive if modest. According to the Canadian Census, the proportion of transit commuters increased slightly by 0.90% from 1996 to 2001; in contrast, the proportion of car users decreased by 0.67% over the same period. This slight reversal of fortunes, however, has not been observed everywhere. The region of interest, the city of Hamilton, in Ontario, is a case in point. Following decades of sprawling development, Hamilton had a lower rate of transit commuters at 9.40% and 9.28%, in 1996 and 2006 respectively, compared to the national average at 10.13% and 11.03%. Thus, while some small gains have been observed elsewhere, the situation in Hamilton continues to be that of a declining trend of transit commuters. As the city approaches a cross-road, including new transit investment plans[1], it is important to improve our understanding of the factors that make transit an attractive proposition to commuters.

Data for this study were collected from different sources. For land use information, the Hamilton parcel map for 2009 was obtained from the GIS Planning and Analysis Section, Planning and Economic Development Department, City of Hamilton. Other built

---

[1] Metrolinx: Vision, Goals and Objectives towards Sustainable Transportation (http://www.metrolinx.com/Docs/WhitePapers/WhitePaper1.pdf)

environment information such as number of street intersection, length of streets, length of sidewalk , and square footage of buildings have been provided by the GIS Services, Information Technology Services, City of Hamilton. Transit information was collected from the Hamilton Street Railway (HSR). The Hamilton Street Railway Company (HSR) is the Transit Division of the City of Hamilton, Public Works Department. A geographic file was made available that describes all bus routes for the Hamilton Street Railway (HSR) public transit system, as well as all the bus stops associated with each route. In addition, information for transit frequency at each bus stop level was provided. There are total thirty-three transit routes of HSR. Among these, three routes operate only in the summer. For this reason, these routes were not considered during preparation of transit related variables. Finally, counts of population, employment, residential dwelling, median household income, and commuter statistics by different modes of transportation were collected from the 2006 Census at the Dissemination Area (DA) level. This is the smallest publicly available Census geography.

Built environment variables were processed based on the above sources, using a Geographical Information System to extract the built environment variables at the DA level. For calculating the variable ‘square footage to parcel ratio’, parcels with no structure i.e. parcels with zero square footage have been excluded. All built environment variables have been defined as densities or counts. Land use mix, on the other hand, is a composite variable which was first used in the travel behavior research by Cervero (1988) and further developed and used by Frank and Pivo (1995) and Cervero and Kockelman (1997). Land use mix is defined as a function of the different land uses within a given area using an entropy formulation as follows:

$$-\frac{1}{\ln(N)}\sum P_n \ln(P_n) \qquad 2.1$$

where P_n is the proportion of land in the area used by type n, and N is the total number of uses. For calculating land use mix in this paper, five types of land uses have been considered, namely residential, commercial, office, institutional, and parks/open spaces. Land use mix always has a range from 0 to 1. It informs us about the homogenous land use within the DA with values (or “with a value”) closer to 0 and the diversified land use with values closer to 1.

In many cases, the most proximate HSR facility may be across a DA boundary. In order to account for availability of HSR services in contiguous DAs spatially lagged versions of these variables are used for the analysis in addition to level of service (e.g. HSR route and density) and built environment (e.g. sidewalks) variables. A spatially lagged density variable X^L is defined as follows:

$$X_i^L = \frac{\sum w_{ij} X_j}{\sum w_{ij} A_j} \qquad 2.2$$

where the lag of the variable at DA i depends on the value of the variable at DAs j, X_j, and the areas at those locations A_j. A spatial weight w_{ij} takes the value of 1 if DAs i and j are contiguous, and zero otherwise or if $i=j$. The lag of variables considered in this study are namely, HSR route density, HSR stop density, HSR frequency density, street density, intersection density, and sidewalk density.

After preparing all variables, DAs with zero transit commuters and zero HSR stops were removed from the dataset. Descriptive statistics of all variables after removing DAs with zero transit commuters and zero HSR stops along with the definition of the variables are given in Table 2.1.

Table 2.1 Variable definition and descriptive statistics.

		Descriptive Statistics			
Variable name	Variable definition	Min	Max	Mean	Standard Deviation
Pop_Den	Population density (number of population within DA/ DA area in sq.km)	0.0	80922.0	4777.8	5390.3
logMHI	Logarithm of (median household income × 10^{-4})	-9.2	3.0	1.6	1.1
Stop_Den	Density of HSR stops (number of HSR stops within DA/ DA area in sq.km)	0.0	759.0	23.5	41.1
$Stop\_Den^L$	Lag of the density of HSR stops	0.0	230.0	20.0	24.8
Route_Den	Density of HSR routes (length of HSR route in km within DA/ DA area in sq.km)	0.0	85.0	3.8	6.7
$Route\_Den^L$	Lag of the density of HSR routes	0.0	55.0	3.5	5.7
Freq_Den	Density of HSR frequency (frequency of HSR within DA/ DA area in sq.km)	0.0	42010.0	1328.0	2473.0
$Freq\_Den^L$	Lag of the density of HSR frequency	0.0	12795.0	1175.5	1502.3
Near_Dis	The nearest distance of HSR stop from the centroid of DA in meter	9.5	13612.0	389.4	1119.6
T_Employment	Number of total employment within DA	115.0	2525.0	488.0	243.7
Access_E	Number of jobs that can be reached from a DA × 10^{-4}	0.0	17.0	1.3	1.5
LUM	Land use mix	0.0	1.0	0.3	0.2
SF_P_Ratio	Square footage of building within DA divided by plot area DA area within DA	0.0	1.0	0.3	0.1
SW_Den	Sidewalk density (length of sidewalks in km within DA/DA area in sq.km)	0.0	62.0	23.3	11.9
SW_Den^L	Lag of sidewalk density	0.0	52.0	16.8	10.7
Inter_Den	Intersection density (number of intersection within DA/ DA area in sq.km)	0.0	1034.0	61.7	69.6
$Inter\_Den^L$	Lag of the intersection density	0.5	172.0	36.1	26.3
Street_Den	Street density (length of streets in km within DA/ DA area in sq.km)	0.5	25.0	12.7	4.9
$Street\_Den^L$	Lag of street density	1.0	20.0	10.0	4.7
Dwell_Den	Dwelling density (number of dwelling within DA / DA area in sq.km) × 10^{-4}	0.0	4.0	0.2	0.3

2.4. Methods

2.4.1. Measuring accessibility by transit

A key variable in this analysis is accessibility by transit. In order to create this variable, a network dataset was created for each unique HSR route, including the stop information and direction of travel. Stop information gives the point of entry into the transit system and direction is used to calculate travel time to different destinations. Service areas were then developed for each of 2,936 bus stops, using a threshold of 30 minutes travel time and an average travel speed of 23 km per hour, but ignoring transfer time. The 30 min threshold was selected based on the average commute time by transit as calculated from the 2005 General Social Survey of Canada. Travel speed was obtained as the average speed of selected HSR routes in both directions, based on published timetable data. Service areas for each bus stop then were overlaid on the DA boundaries, and DAs reachable within the area coverage were identified. Combined with the employment statistics for DAs, the final outcome of the service area analysis was, for each stop, the number of jobs accessible from a given stop: in other words, a cumulative opportunities accessibility measure. These values at the stop level were further processed to eliminate duplicates (i.e. the same destination DA can be reached from two or more stops within an origin DA). Finally, all unique values were aggregated to produce the measures of accessibility to employment for each DA.

2.4.2. Logistic model for proportions

The outcome variable of interest is the proportion of transit commuters in DA *i*, relative to all commuters in DA *i*. The appropriate approach to model proportions is to estimate a generalized linear regression. The model assumes that the outcome variable Y_i (the number of commuters by transit) is the result of a process with an underlying categorical variable (e.g. transit and non-transit). The underlying variable *j* at *i* (y_{ij}) is dichotomous and takes the value of 1 (i.e., if traveler *j* commutes by transit) and 0 (i.e., if the traveler uses other modes of transportation). Given *n* travelers at *i*, the probability of observing exactly *k* transit users is given by the binomial distribution with parameter π (the probability of success):

$$P(Y_i = k) = \binom{n_i}{k_i} \pi^k (1-\pi)^{n-k} \qquad 2.3$$

The moments of this distribution are $E[Y_i]=n\pi$ and $var[Yi]=n\pi(1-\pi)$. Given π, it is possible to estimate Y_i. A link function allows us to relate the mean of the distribution function (i.e. $E[Y_i]$) to a linear set of predictors. The logit link function is given by the ratio of the odds of success:

$$\text{logit}(\pi) = \log_e\left[\frac{\pi}{1-\pi}\right] = \beta_0 + \sum_{j=1}^{k} \beta_j X_j \qquad 2.4$$

The inverse transformation of the equation above is the well-known logistic function for modal share analysis (Ortúzar and Willumsen, 2001):

$$\pi = \frac{e^{\beta_0 + \sum_{j=1}^{k} \beta_j X_j}}{1 + e^{\beta_0 + \sum_{j=1}^{k} \beta_j X_j}} \qquad 2.5$$

The probability of success (and consequently E[Y_i], the number of successes) can be calculated based on estimable coefficients β_0 and β_j. The coefficients are estimated using maximum likelihood techniques.

2.4.3. Over-dispersion and spatial autocorrelation

As noted above, an important assumption is that the outcomes of the underlying variable are independent and identically distributed. Under such assumption, the mean and variance of the distribution depend on a single parameter (π), and the theoretical value of the variance is fixed by the estimation of the mean. In practice, the variance can be larger or smaller than its nominal (i.e. theoretically expected) value, leading to situations of over- and under-dispersion. Under-dispersion is extremely rare, but over-dispersion is more common, with some maintaining that nominal dispersion is in fact the exception (e.g. McCullagh and Nelder, 1989). Over-dispersion can be observed due to variations in success probabilities, which implies correlation between binary responses. This can be observed by noting that the variance of the process is given by:

$$\operatorname{var}(Y_i) = \sum_{j=1}^{n_i} \operatorname{var}(y_{ij}) + \sum_{j=1}^{n_i} \sum_{k \neq j} \operatorname{cov}(y_{ij}, y_{ik}) \qquad 2.6$$

The first term of the right-hand side is $n\pi(1-\pi)$, that is, the nominal variance. If the responses are correlated, the second term will be different from zero, and the variance of Y_i

will depart from its nominal value, i.e., $var(Y_i) = \varphi n\pi(1-\pi)$, where φ is an over-dispersion parameter as follows (p is the number of parameters in the model):

$$\varphi = \frac{1}{n-p}\sum\frac{(y-n\pi)^2}{n\pi(1-\pi)} \qquad 2.7$$

Clearly, when φ>1 (i.e. there is over-dispersion), the nominal variance will be under-estimated, and the nominal interval of confidence will be too small. As a consequence some coefficients may appear to be significant when they are not.

Over-dispersion can happen due to clustering or hierarchical effects, for instance if the same individual is observed repeated times. In a spatial setting, a variable can be correlated across geographical units. This is the well-known autocorrelation effect discussed in spatial analysis (Cliff and Ord, 1969; Cliff and Ord, 1973, 1981; Getis, 2007, 2008), whereby the values of a spatially distributed variable display patterns of similarities (positive autocorrelation) or dissimilarities (negative autocorrelation) across space.

A way to deal with over-dispersion is to use the calculated over-dispersion factor to correct the intervals of confidence. While this can improve inference, the correction does not address the underlying causes of over-dispersion. Depending on the degree of over-dispersion, this factor can have an important effect on significance testing. If over-dispersion is caused by correlation between spatial responses, a possibility is to incorporate autocorrelation as part of the model. This can be done in a more or less straightforward fashion in the analysis of normally distributed variables by means of the auto-normal model (Griffith, 2000) widely used in spatial econometric and statistics (Anselin, 1988; Griffith, 1988). In the case of generalized linear models, an alternative is to adopt a spatial filtering approach to remove residual autocorrelation by judiciously using latent map patterns. This is described next.

2.4.4. Eigenvector-based spatial filtering

Spatial autocorrelation in a variable Y is commonly assessed by means of Moran's I coefficient, defined as (see Griffith and Layne, 1999):

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} \left(y_i - \bar{y}\right)\left(y_j - \bar{y}\right)}{\sum_i \left(y_i - \bar{y}\right)^2} \qquad 2.8$$

where $\bar{y}$ is the mean of Y and w_{ij} is a spatial weight that codifies spatial relationships between spatially referenced observations, as noted above. It can be seen that the coefficient is a function of the spatial auto-covariance, normalized by the variance and the total connectivity in the system, relative to its size n. The spatial weights w_{ij} can be arranged in matrix form to give a spatial lag operator $\mathbf{W}$.

A spatial filtering approach has been proposed to capture spatial autocorrelation that depends on the eigenvector analysis of the spatial lag operator (Griffith, 2003). The eigenvector filtering approach is a nonparametric technique. It creates synthetic variables that act as surrogates of missing variables in the regression equation to remove the effect of spatial autocorrelation in the residuals. Eigenvectors for a spatial filter are derived from the following square matrix obtained by transforming the spatial lag operator using a projection matrix:

$$\left(\mathbf{I}-\frac{\boldsymbol{11'}}{n}\right)\mathbf{W}\left(\mathbf{I}-\frac{\boldsymbol{11'}}{n}\right) \qquad 2.9$$

where **I** is the identity matrix, $\boldsymbol{1}$ is an $n\times 1$ vector of ones, and n is the number of areal units. Expression 2.9 is in fact part of the numerator of Moran's *I* coefficient. The eigenvectors are mutually orthogonal and uncorrelated. Following the orthogonality property of eigenvector analysis, a remarkable feature of the eigenvectors obtained from Equation 2.9, described by Tiefelsdorf and Boots (1995) and Griffith (2004), is that each eigenvector represents a distinct map pattern, and the associated eigenvalues correspond to their degree of autocorrelation as measured by Moran's *I*. Hence, the first eigenvector, $\boldsymbol{E}_1$, is the set of numerical values that has the largest value of *I* achievable by any possible set of numerical values, for the arrangement of locations given the spatial lag operator. The second eigenvector is the set of numerical values that has the largest achievable *I* by any set of numerical values that is uncorrelated with $\boldsymbol{E}_1$. This sequential construction of eigenvectors continues through $\boldsymbol{E}_n$, which is the set of numerical values that has the largest negative value of *I* achievable by any set of numerical values that is uncorrelated with the preceding n-1 eigenvectors.

As mentioned above, to capture spatial autocorrelation the eigenvector based spatial filtering approach provides synthetic variables as proxies to the missing variables along with all explanatory variables, therefore, the expression 2.4 can be rewritten as:

$$\log_e\left[\frac{\pi}{1-\pi}\right]=\beta_0+\sum_{j=1}^{k}\beta_j X_j+\phi S \qquad 2.10$$

where $\boldsymbol{S}$ is a spatial filter constructed using a subset of the n eigenvectors chosen by some supervised selection criteria. Since the eigenvectors are uncorrelated, one possible procedure to select eigenvectors for the filter is based on a forward stepwise search procedure as follows:

1. Initialize an index value i=1 and an empty vector for the spatial filter $\boldsymbol{S}$=[ ]; set $\mathbf{X}^{+}$=$\mathbf{X}$.
2. Select eigenvector $\boldsymbol{E}_i$ as a candidate for inclusion in the model, and estimate the model $\boldsymbol{Y} = f\left(\left[\mathbf{X}^{+}, \boldsymbol{E}_i\right], [\boldsymbol{\beta}, \theta]\right) + \boldsymbol{\varepsilon}$
3. If coefficient θ is significant at a pre-determined level (e.g. $p\leq0.10$), then synthesize the eigenvector and the existing filter: $\boldsymbol{S}$=$\boldsymbol{S}$ +$\theta\boldsymbol{E}_i$, and continue to step (4), otherwise, return to step (3)
4. Re-set matrix $\mathbf{X}^{+}$ as follows: $\mathbf{X}^{+}$=[$\mathbf{X}$, $\boldsymbol{S}$], and estimate the model $\boldsymbol{Y} = f\left(\mathbf{X}^{+}, \boldsymbol{\beta}\right) + \boldsymbol{\varepsilon}$.
5. Determine the degree of autocorrelation in the estimated residuals; if the normalized Moran's coefficient is greater than 0.5, set i=i+1 and return to 3, otherwise end. The spatial filter is vector $\boldsymbol{S}$.

This stop criterion followed above is slightly less stringent than the normalized value of 0.1 used by Tiefelsdorf and Griffith (2007), but already associated with a non-significant level of residual autocorrelation (p>0.40). Using a stop criterion of 0.5 instead of 0.1 means that the search will be somewhat more parsimonious and the filter will likely use fewer eigenvectors.

Construction of the filter can be achieved with simple linear combinations of eigenvectors, as discussed by Boots and Tiefelsdorf (2000). These authors note that for the case of combining two eigenvectors, the resulting map pattern is associated with a value of *I* that is a weighted average of the two eigenvector autocorrelation coefficients. Further, the level of autocorrelation of the new vector must be contained in the interval defined by that of the two original eigenvectors. A proposal is to define the new eigenvector, $\boldsymbol{E}_0$, as:

$$\boldsymbol{E}_0 = \sin(\theta)\boldsymbol{E}_j + \cos(\theta)\boldsymbol{E}_k \qquad 2.11$$

for two eigenvectors $\boldsymbol{E}_j$ and $\boldsymbol{E}_k$ and some angle θ. This specification results in a level of autocorrelation:

$$I_0 = sin^2(\theta) I_j + cos^2(\theta) I_k \qquad 2.12$$

Equivalently, consider the following linear combination:

$$\boldsymbol{E}_0 = a\boldsymbol{E}_j + b\boldsymbol{E}_k \qquad 2.13$$

with:

$$\sin(\theta) = \sqrt{\frac{a^2}{a^2+b^2}} \text{ and } \cos(\theta) = \sqrt{\frac{b^2}{a^2+b^2}} \tag{2.14}$$

This implies that:

$$\begin{aligned} I_0 &= \frac{n}{\boldsymbol{1}'\mathbf{W}\boldsymbol{1}} \frac{\left(\left(a\boldsymbol{E}_j + b\boldsymbol{E}_k \right)' \left(\mathbf{I} - \frac{\boldsymbol{11}'}{n} \right) \mathbf{W} \left(\mathbf{I} - \frac{\boldsymbol{11}'}{n} \right) \left(a\boldsymbol{E}_j + b\boldsymbol{E}_k \right) \right)}{a^2+b^2} \\ &= \frac{a^2}{a^2+b^2} I_j + \frac{b^2}{a^2+b^2} I_k \end{aligned} \tag{2.15}$$

The bivariate linear combination can be easily extended to a multivariate linear combination of the form:

$$\boldsymbol{E}_0 = \sum_{q=1}^{Q} \delta_q \boldsymbol{E}_q \tag{2.16}$$

for any subset Q of the $\boldsymbol{N}$ eigenvectors. The autocorrelation value of $\boldsymbol{E_0}$ is given by:

$$\frac{1}{\sum_{q=1}^{Q} \delta_q^2} \sum_{q=1}^{Q} \delta_q^2 I_q \tag{2.17}$$

The above shows that a filter can be produced with an arbitrary level of autocorrelation.

2.5. Results and discussion

Spatial filtering, since it is based on conventional estimation routines once that the eigenvector analysis has been completed, can be easily implemented in commercial statistical packages. A suite of models is compared here that: 1) ignore over-dispersion and autocorrelation; 2) estimate the over-dispersion parameter without the spatial filter; and 3) jointly estimate a dispersion parameter and use spatial filtering to control for autocorrelation. A specification search strategy was conducted beginning with an exhaustive model that included all explanatory factors, the significance of which was reassessed based on the search of eigenvectors for the spatial filter. Table 2.2 presents the model estimation results.

Table 2.2 Logistic regression parameter estimation in three different specifications.

	(1) No OD parameter and spatial filter		(2) OD parameter without spatial filter		(3) OD parameter with spatial filter	
Variables	Coefficients	p-value	Coefficients	p-value	Coefficients	p-value
Constant	-2.54230	0.0000	-2.52545	0.0000	-2.27885	0.0000
Socio-economic characteristics						
Pop_Den	0.22681	0.0321	-	-	-	-
logMHI	-0.15273	0.0000	-0.15426	0.0000	-0.09714	0.0001
Transit Level of Service						
Stop_Den	-0.00236	0.0000	-	-	-	-
$Stop\_Den^L$	-0.02903	0.0000	-0.02979	0.0000	-0.00838	0.0109
Route_Den	-0.00830	0.0000	-0.01015	0.0404	-	-
$Route\_Den^L$	0.02791	0.0000	0.02839	0.0258	-	-
Freq_Den	0.00004	0.0000	-	-	-	-
$Freq\_Den^L$	0.00047	0.0000	0.00049	0.0000	0.00016	0.0057
Transit accessibility						
Near_Dis	-0.00010	0.0000	-0.00010	0.0210	-0.00013	0.0000
T_Employment	-	-	-	-	0.00018	0.0037
Access_E	0.06439	0.0000	0.06588	0.0010	0.02670	0.0453
Built environment characteristics						
LUM	-	-	-	-	-	-
SF_P_Ratio	0.22681	0.0321	-	-	-	-
SW_Den	-0.00213	0.0322	-	-	-	-
SW_Den^L	-	-	-	-	-	-
Inter_Den	-0.00034	0.0100	-	-	-	-
$Inter\_Den^L$	-	-	-	-	-	-
Street_Den	0.01424	0.0000	0.01379	0.0478	-	-
$Street\_Den^L$	0.01858	0.0000	0.01851	0.0205	-	-
Dwell_Den	0.41141	0.0000	0.58641	0.0000	0.38747	0.0000
Spatial filter						
Spatial Filter (14 eigenvectors)	-	-	-	-	1.00000	0.0000
N	761		761		761	
pseudo-R2	0.307		0.301		0.450	
φ	1 (assumed)		3.601 (estimated)		3.077 (estimated)	
Moran's I (z)	6.225 (p<0.0000)		6.195		0.292	

The results indicate that under the assumption of the nominal level of dispersion and ignoring spatial autocorrelation, most variables are significant at the conventional 0.01 and 0.05 probability levels. The only exceptions are total employment, land use mix, and lag of sidewalk and intersection density. The results of the model would overall be comparable to previous findings in transit accessibility and ridership research. The results, however, are suspect since possible over-dispersion and significant residual autocorrelation (z-score of Moran's coefficient is 6.225) are known to affect the variance and thus lead to misleading inference. In the second model specification, the over-dispersion parameter was estimated jointly along with the other logistic regression parameters. The estimated value of the dispersion parameter is high (3.601). This, of course, has important implications for inference. Once that over-dispersion is taken into account, and the standard errors are corrected, the significance of several variables is lost. This model, however, still displays a significant level of spatial autocorrelation, with a z-score of Moran's coefficient of 6.195.

As seen in the table, residual autocorrelation has been successfully removed: examination of the residuals by means of Moran's I (z=0.292) shows that residual autocorrelation is no longer significant. In addition to capturing spatial autocorrelation, the filter also helps to reduce the value of the over-dispersion parameter from 3.601 to 3.077. The final model has a pseudo-R^2 value 0.450, which compares favourably to the fit of the previous models. In terms of inference, seven parameters are statistically significant in the final model of Table 2.2. In comparison, the first and second models had sixteen and ten significant variables respectively. Clearly, ignoring the parallel issues of over-dispersion and autocorrelation would cause to overstate the relevance of a number of variables in this analysis, in particular with respect to level of service and the built environment. Any prescriptions based on these findings would likely be misleading and possibly ineffective.

The substantive findings of the analysis are now discussed in this section. In agreement with previous studies, it is found that median household income has a negative effect on the share of transit (Cervero and Kockelman, 1997; Greenwald and Boarnet, 2001; Hess, 2009; Hsiao et al., 1997; Taylor and Miller, 2003). This suggests that as household income increases, the members of that family will have more alternatives available, most likely private mobility, and will as a consequence use less transit. In a model with only lagged stop density, the variable has positive sign which was the priori expectation from previous studies (Alshalalfah and Shalaby, 2007; Cervero et al., 2009a; Cervero et al., 2009b; Johnson, 2003; Taylor et al., 2009). The nearest distance of bus stop from the centroid of the DA has the expected negative sign. Transit commuters usually walk from their home to the nearest transit stop. As this distance increases, they become less willing to walk to the transit stop (Alshalalfah and Shalaby, 2007; Cervero et al., 2009a; Greenwald and Boarnet, 2001; Hess, 2009; Kitamura et al., 1997; Loutzenheiser, 1997; O'Sullivan and Morrall, 1997; Olszewski and Wlbowo, 2005; Sanchez, 1999; Zhao et al., 2003). Also, as expected, dwelling density has positive sign which indicate transit ridership will be higher in areas with higher dwelling density (Cervero et al., 2009a; Johnson, 2003; Taylor and Miller, 2003). Higher density is one of the preconditions of Transit Oriented Development (TOD). A TOD neighbourhood typically has a center with a transit station or stop (train station, metro station, tram stop, or bus stop), surrounded by relatively high-density development with progressively lower-density development spreading outwards from the center. TODs generally are located within a radius of one-quarter to one-half mile (400 to 800 meters) from a transit stop, as this is considered to be an appropriate scale for pedestrians. Finally, accessibility by transit displays a positive sign, indicating that transit shares increase in the measure that jobs are made accessible by the transit system. Thus, in addition to density of the built environment and proximity of stops (accessibility to transit), transit ridership depends to a significant extent of the ability of the system to produce accessibility. This would indicate that local solutions are important, but that a more global perspective of the accessibility that the system generates needs to be considered too.

2.6. Conclusions

Modeling aggregated transit shares is an important element to plan services. A focus of previous research has been on accessibility to transit as an explicative factor of mode shares. In this chapter, it was argued that accessibility by transit is a factor in need of more attention. Investigation of transit shares for the city of Hamilton, Canada, was conducted by means of a logistic regression model for proportions, using a spatial filtering approach to control for spatial autocorrelation. This study illustrates the use of eigenvector-based spatial filtering to deal with autocorrelation in non-normal models, and demonstrates the effect of autocorrelation and over-dispersion in the logistic model.

The results of this case study indicate that use of spatial filters can effectively remove spatial autocorrelation and reduce the degree of over-dispersion in the model. In this case study, the spatial filter was composed of a linear combination of 14 eigenvectors to capture moderate-to-strong positive spatial autocorrelation. With regards to the variable 'accessibility by transit', the results show that it contributes positively to usage of transit. This variable retained its significance in the final model, even after other measures of the level of service and the built environment lost theirs. The model also confirms previous findings regarding mode share of transit. Dwelling density is important for increasing the mode share of transit. This is the only built environment variable which is significant after applying the spatial filtering approach. Frequency of transit service also can help to increase transit mode share, when considering the overall levels of service in neighbouring DAs.

It is worthwhile to note at this point that the results of this study are contingent on the scale of analysis, in this case Dissemination Areas. Previous research suggests that analysis at smaller scales can help to improve our understanding of accessibility, both when entering and leaving the transit system (e.g. Rodriguez et al., 2009; Werner et al., 2010). In the present case, a direction for further research is to use the results of the spatial filtering procedure as an exploratory tool to search for omitted relevant variables. Some of these variables may well be related to factors at the level of pedestrian environment. This is addressed in the following chapter.

Chapter 3 Selecting Case Sites for Walkability Audits

3.1. Introduction

Representative characteristics of the built environment thought to influence travel behavior are commonly summarized by means of Cervero and Kockelman's 3Ds: Density, Diversity, and Design (1997). Many of the variables typically used to represent these characteristics were considered in the analysis of transit shares presented in Chapter 2. As is usually the case, statistical analysis of transit shares in Hamilton is conditioned by the scale of the analysis. Studies of travel behavior often have to grapple with the issue of scale, either implicitly, depending on data availability (e.g. the smallest available Census geography), or explicitly, if further data aggregations are desired or recommended (e.g. Boarnet and Sarmiento, 1998; Guo and Bhat, 2007). When aggregated data are used, a distinct possibility is that some information about the built environments will be lost, especially in terms of micro-level design factors such as sidewalk width, path materials, and number of curb cuts, and more qualitative attributes such as street cleanliness (e.g. Frank et al., 2003; Saelens et al., 2003). These are all potential factors that affect walking environments, and yet, that are seldom collected by local governments or other agencies in sufficiently wide and systematic ways (Parmenter et al., 2008). Even if available, measurements and analysis of administrative-level data may be insufficient to capture micro-scale effects, due to current limitations in the understanding of geo-spatial perception (e.g. mental maps) and a lack of canonical representation approaches.

In order to create inventories and assessments of the local environments at the pedestrian scale, a suite of methods has been developed in recent years under the label of walkability audits. A walkability audit is a data collection instrument used to itemize and assess different aspects of the local, pedestrian-level environment, including sidewalks, crossing aids, traffic calming devices, and other pedestrian-supportive design features. Numerous audit instruments exist that can provide rich and detailed information about walking environments (Clifton et al., 2007; Dannenberg et al., 2005; Day et al., 2006; Millington et al., 2009; Moudon and Lee, 2003; Pikora et al., 2002). A common trait of existing instruments is that they require substantial amounts of field work, as auditors walk the streets conducting visual inspections of the environment and completing inventories of items of interest. Considering the usually limited resources available to researchers, this has prompted an emerging interest in the development of methods to allow better targeting of efforts.

The objective of this chapter is to propose a model-based approach to select case sites for conducting walkability audits. In very simple terms, the proposal is to identify sites of interest based on the inspection of the unexplained part of a model of some relevant travel behavior. It is hypothesized that systematic under- and over-estimation of the behavior is related to micro-level factors that affect pedestrian mobility; in other words, it is anticipated that better walking environments will be found for locations where the model tends to under-estimate walking-related travel behaviors, and vice-versa. The proposed approach is implemented following the results presented in Chapter 2, for neighborhoods in the city of Hamilton, Canada. The underlying conjecture is that walkable environments can increase the use of transit. The resulting model is then used to identify DAs where the share of transit is under-estimated (the actual share is higher than predicted by the model) or over-estimated (the actual share is lower than the predicted share). Based on the examination of patterns of under- and over-estimation, eight DAs are selected for walkability audits, four at each level (under- and over-estimated). Walkability audits are then conducted, and the data analyzed using contingency tables and χ^2 tests of independence.

The results of the analysis of audit data reveal a number of items commonly found in DAs with higher-than-expected use of transit, including mixed uses, compact development, and availability of bus stops. Other attributes are found to associate with lower-than-expected use of transit, including single land use development. Differences in these attributes lend credence to this conjecture that higher transit shares than predicted by the model correspond to neighborhoods that provide better walking environments. The experiment thus demonstrates that the model-based selection strategy for walkability audit sites proposed in this chapter can be used to more effectively target limited resources for field-based work, and produce valuable insights into the micro-level factors that affect active travel.

3.2. Prior research

Previous research has found that pedestrian-friendly walking environments are conducive for people to walk with increased frequency (Cervero and Kockelman, 1997; Cervero et al., 2009b; Guo, 2009; Hess et al., 1999; Kitamura et al., 1997; Rodriguez and

Joo, 2004). Many studies assess the pedestrian environment in terms of sidewalk density, street density, and street connectivity, widely available data items that can be conveniently processed using a geographic information system. Increasingly, however, there is a need to collect environmental data in more detail, to reflect the fact that walking, as a slow mode, implies that pedestrians experience the environment in a qualitatively different way (Moudon and Lee, 2003). Walkability audit instruments help fill the gap left by other statistics, as they can be used to collect micro-level information about pedestrian environments. Two relevant aspects of these instruments are their resolution and extent (Moudon and Lee, 2003). Resolution refers to the precision of the items considered for the audits, whereas the extent refers to the geographic area where the audit takes place. To date most of the studies pertaining to walkability audits have dealt with the issue of resolution, and the development of user friendly, efficient, and reliable instruments (Clifton et al., 2007; Dannenberg et al., 2005; Day et al., 2006; Millington et al., 2009; Pikora et al., 2002).

The required extent of audit tools, in contrast, has not yet received much attention, perhaps because some instruments are still in their developmental stages (Brownson et al., 2009; Sallis, 2009). However, considering the time and cost involved in conducting comprehensive audits over entire study areas, there is currently an interest in devising more effective protocols to assist in the determination of the best extent for a project. McMillan et al. (2010) state that the principal limitation of walkability audits is “the time and cost involved in data collection” (p. 1). Clarke et al. (2010) note that “in-person audits are highly resource intensive and costly, making them prohibitive for many studies.” (p. 1224).

These opinions are echoed by Rundle et al. (2011), who also report that audits "are time-consuming and expensive to conduct largely because of the costs of travel" (p. 95). Among the walkability audits available, the Pedestrian Environment Data Scan (PEDS) instrument (Clifton et al., 2007) requires on average 3-5 minutes per segment to audit on foot, compared to 10 minutes for the Analytic Audit Tool of Saint Louis University (Hoehner et al., 2005), 20 minutes for the Irvine Minnesota Inventory (Day et al., 2006), and 30 minutes for the Walking Suitability Assessment Form (Emery et al., 2003). In the case of a new project, the total project cost must also consider the time required for training, reliability testing, and other initial items. When these costs are included, the total cost can be substantially greater than the nominal cost of conducting an audit.

In principle, the cost of implementing walkability audits can be reduced by sampling streets within neighborhoods, by sampling neighborhoods, or a combination of these two approaches. In a study that assessed the effect of sampling streets on the reliability of audit data, McMillan et al. (2010) collected data comprehensively on five variables (presence of sidewalks, observer ratings of attractiveness and safety for walking, connectivity, and number of traffic lanes) for neighborhoods of 400 m and 800 m around a specific location. Then, they compared the results to those obtained from sampling street segments at 75%, 50%, and 25% levels. Based on the results of this analysis, these authors conclude that a sample of 25% residential street segments within 400 m radius of a residence is adequate for the pedestrian built environment, although a fuller sample of arterial street segments may be required due to higher heterogeneity.

A different possibility is to reduce the cost of implementation by reducing the cost of doing field work. Along these lines, a study by Clarke et al. (2010) proposes the use of Google Street View as a way to limit the cost of objectively measuring pedestrian environments. After selecting neighborhoods for the audits, Clarke et al. conducted the audits virtually using geo-spatial technology provided by Google. In order to test this approach, pre-existing data collected using the same instrument as part of a previous project were used to assess the reliability of these "virtual" audits, relative to the "in-person" audits. The results indicate that a virtual audit is able to provide reliable information about recreational facilities, local food environment, and general land uses. However, care must be taken for more finely detailed observations, for instance, presence of garbage, litter, or broken glass. A similar approach is used by Rundle et al. (2011) to find that 54.3% of items audited have high levels of concordance in term of percentage agreement for categorical measures and spearman rank-order correlation for continuous measures.

There is a key difference between the method proposed here and the methods of McMillan et al. (2010), Clarke et al. (2010), and Rundle et al. (2011). These three approaches can help to control the cost of conducting walkability audits by reducing the number of street segments visited (after sampling within a neighborhood), or by entirely eliminating the need for in-person visits (in the case of virtual audits). However, neither of these approaches provides guidelines for selecting neighborhoods for the audits. A random selection of neighborhoods may fail to capture the wide range of attributes in an entire study area, and as a consequence, the results may be inaccurate or even misleading.

Selection based on one or two attributes of areas (e.g. household income, level of education) risks missing other confounding factors. Therefore, assuming that even at a reduced cost audits for a moderately large geographic area will be expensive, the intent of the approach proposed here is to provide a way to sample neighborhoods for walkability audits in a systematic way that can account for a wide range of confounding factors. Implementation of the audits can subsequently, as desired, proceed by sampling street segments (McMillan et al., 2010), or by conducting virtual site visits (Clarke et al., 2010; Rundle et al., 2011). The proposed approach is described next.

3.3. Methods

Multivariate statistical models are frequently estimated using meso-scale (administrative) data in order to assess the effect of various variables on travel behavior. If Y is the behavior of interest, these models are generally composed of two elements, namely the observed/explained and unobserved/unexplained parts, as follows:

$$Y_i = f\left(\boldsymbol{X}_i, \boldsymbol{\beta}\right) + \varepsilon_i \qquad 3.1$$

In equation 3.1, $\boldsymbol{X}_i$ is a vector of (observed) explanatory variables, $\boldsymbol{\beta}$ is a vector of estimable coefficients, and $\boldsymbol{\varepsilon}_i$ is a residual term (unobserved). Depending on the nature of the dependent variable Y, the model can be a linear regression (for continuous variables) or a generalized linear model (for instance, a logistic regression for proportions). Regardless of the specific form of the model, a key assumption is that the residual terms are random, and thus uncorrelated across units $\boldsymbol{i}$. In practice, this assumption is violated with sufficient frequency in the case of modeling with spatial data, as to have given rise to a specialized set of techniques to deal with situations where residuals are spatially correlated (Anselin, 1988; Cliff and Ord, 1973, 1981; Griffith, 1988; Haining, 1990).

Spatial autocorrelation of the residuals is often attributed to common, but unobserved, factors that influence the process. Consider for example the situation illustrated in Figure 3.1, which depicts a simple generic model (a bivariate regression) with seemingly well-behaved residuals. Clearly, the observations above the regression line are under-estimated by the model (the actual behavior is more frequent than predicted), and observations below the regression line are over-estimated (the actual behavior is less frequent than predicted). Depending on the actual spatial distribution of the observations, the residuals could be independent (Figure 3.2a) or spatially autocorrelated (Figure 3.2b) – an indication that a relevant variable with a spatial pattern has been omitted (e.g. a variable common in the north of the region, but absent in the south).

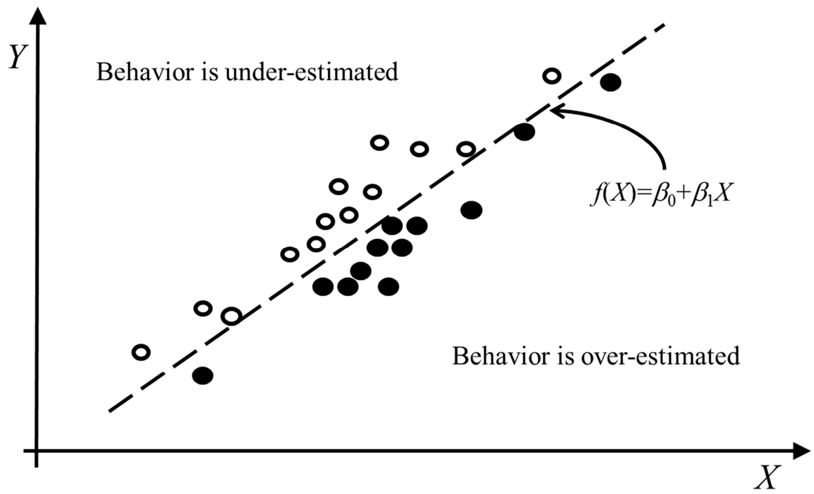

Figure 3.1 Example of a generic model.

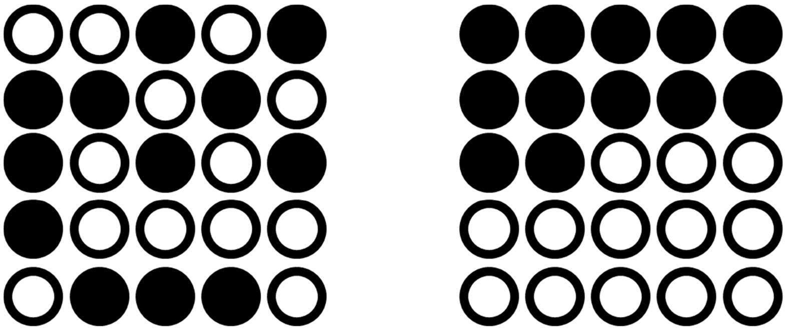

Figure 3.2 Examples of random and spatially autocorrelated residuals.

The working hypothesis is that residual spatial autocorrelation is caused, at least in part, by the omission of relevant micro-scale variables (i.e. variables at the level of the pedestrian experience) that follow a spatial pattern. Specifically, in the case of under-estimation, the conjecture is that the omitted variables are related to elements of the environment that facilitate/encourage the behavior (e.g. greater connectivity of walkable paths, cleanliness), whereas variables that hinder/discourage the behavior are related to over-estimation of the behavior (e.g. lack of sidewalks or amenities). The similarity of attributes among spatially proximate street segments is in fact the operational basis of the sampling procedure proposed by McMillan et al. (2010), and is an effect that has been noted as well by Agrawal et al. (2008).

Given a statistical model of a travel behavior of interest, a proposal is given to select sites for walkability audits based on the examination of the unexplained part of the model. An attractive feature of this proposal is that the unexplained part is obtained after controlling for a suitable set (based on availability) of explanatory/confounding factors. Assuming that the residuals are independent of the variables included in the model, the cases can be selected based on different levels of the dependent variable only, since potential confounders are accounted for in the model. Thus, if the focus of the project is on factors that facilitate the behavior, then sites can be selected from the set of observations that the model under-predicts. If the focus is on the factors that hinder the behavior, then selection can be made from the set of observations that the model over-predicts. When control cases are sought, pairs of observations at different levels of the behavior can be identified, one each from the under-estimated and over-estimated sets (see Figure 3.3).

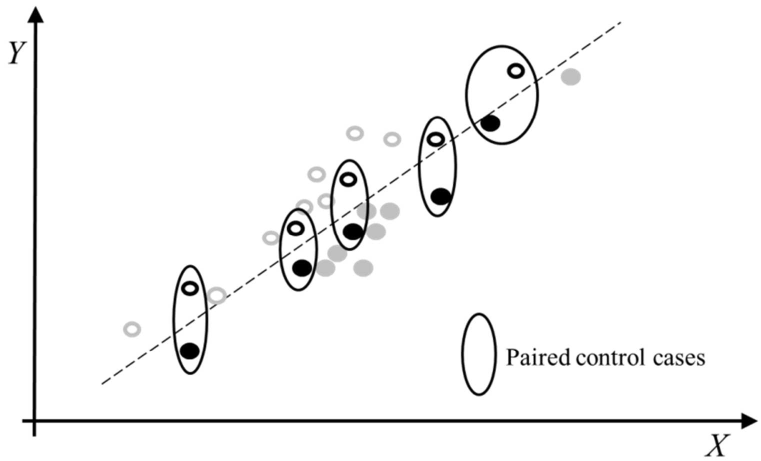

Figure 3.3 Paired control cases.

3.4. Case study

3.4.1. Context

A model of the share of transit as a mode of transportation for the journey to work, third model specification of Table 2.2, is the basis of meso-scale analysis in Chapter 3. The resulting model is repeated in Table 3.1, where it can be seen that residual spatial autocorrelation has been successfully removed by means of a filter composed of fourteen eigenvectors and an over-dispersion factor is estimated to correct the variance of the model.

The spatial filter can be examined at this point to identify positive values (observations that the model under-estimates and that therefore need a positive correction) and negative values (observations that the model over-estimates). Please note that variables that were not significant in the specification search of Chapter 2 are not shown in the table (for a full description of variables tested see Table 3.1).

Table 3.1 Results of logistic model for proportion of commuting trips by transit.

Variable	Coefficient	p-value
Constant	-2.27885	0.0000
Socio-economic characteristics		
logMHI	-0.09714	0.0001
Transit Level of Service		
Stop_Den^L	-0.00838	0.0109
Freq_Den^L	0.00016	0.0057
Transit accessibility		
Near_Dis	-0.00013	0.0000
T_Employment	0.00018	0.0037
Access_E	0.02670	0.0453
Built environment characteristics		
Dwell_Den	0.38747	0.0000
Spatial filter		
Filter (14 eigenvectors)	1.00000	0.0000
Summary Statistics		
n=761		
pseudo-R^2=0.450		
Moran's I (Z)=0.292		
Over-dispersion parameter (estimated)=3.077		

3.4.2. Selection of sites

Of the 761 DAs included in the analysis, 308 have negative values of the spatial filter, and 453 have positive values (Figure 3.4). Given resource constraints, eight DAs were selected for walkability audits. Four DAs were chosen from the set with positive values of the spatial filter, and four from the set with negative values, or in other words, four matching pairs for control. In order to select the eight DAs, firstly the spatial filter was divided into five quantiles. Two DAs were then chosen from amongst the bottom 20% and two from the top 20% of values. Four more DAs were taken from the middle quantiles, again two with positive and two with negative values. To ensure that the pairs were matching, careful attention was paid to the values of transit user and transit modal share modal share when selecting the DAs. All four pairs of DAs have approximately matching transit shares as a primary selection criterion, and wherever possible also approximately matching numbers of transit users. In general, multiple DAs were found that met the criteria at each level of transit users and transit modal share. To form the matching pairs, cases were randomly selected from available DAs. In other words, selection of cases was random within strata. In the present case, we randomly selected one matching pair, for bottom and upper quantiles, and two matching pairs for the middle quantile.

Figure 3.4 Spatial filter with audit locations.

3.4.3. Audit instrument and technology

Walkability audits were conducted using the Pedestrian Environment Data Scan instrument (PEDS, (Clifton et al., 2007)). PEDS was selected because of its relatively low data collection time and high resolution. The instrument was developed in such a fashion that it can measure various built and natural environmental features pertaining to walking in an efficient and reliable way (Clifton et al., 2007). The spatial unit of analysis for auditing pedestrian environments using the PEDS tool is the road or pathway segment. In this study, only road segments were audited. Further, PEDS resource materials were initially developed as a pencil and paper instrument, but later adapted for use with handheld technology (Clifton et al., 2007). For this study, PEDS was also adapted from its original pencil and paper format to digital format using ArcPad Application Builder 7.0.1.

The electronic version of PEDS greatly facilitates data entry. Most of the attributes in PEDS are categorical, with the exception of sidewalk connectivity, number of lanes, and posted speed limit. For categorical variables, categories can be selected by clicking on the drop-down arrow. Few of the attributes could have more than one answer. Check-boxes were provided for such attributes (Figure 3.5). The digital version of PEDS audit then transferred into a mobile device which can support ArcView geographic information file. This technology eliminates the time required for data entry. It also provides improved quality of data and reduces errors (Clifton et al., 2007).

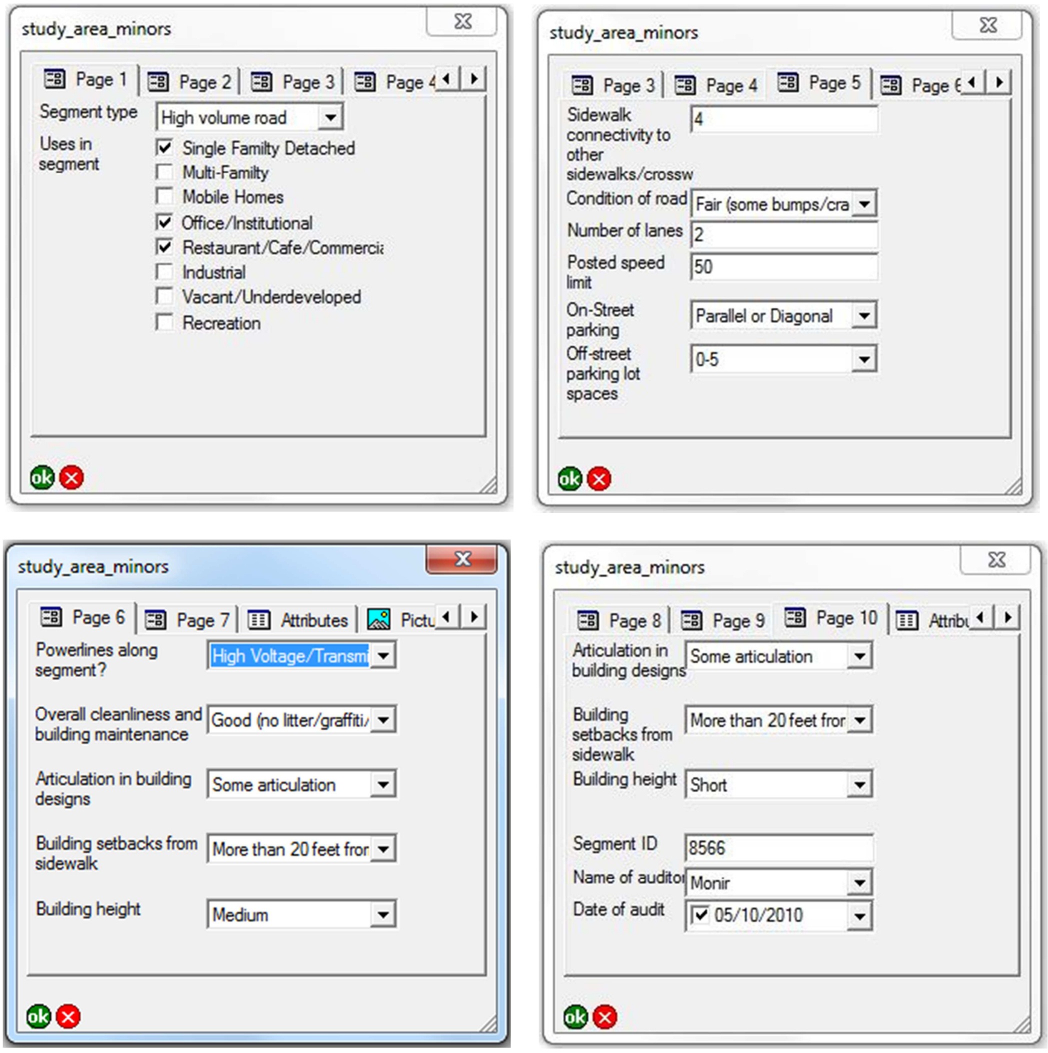

Figure 3.5 Electronic version of PEDS.

3.4.4. Walkability audits

Audits of all selected DAs were conducted by the author in early October, 2010, with support from three graduate research assistants. The research assistants received training from the author using the training materials provided by Clifton et al. (2007). In addition, a pilot audit was conducted before the final audits were launched. Street network information was collected in electronic format by the City of Hamilton (GIS Services, Information Technology Services). Exhaustively all street segments within the eight DAs were audited. In total, 119 street segments were audited. Coincidentally, out of the 119 segments, 60 were from DAs with a positive filter value and 59 are from DAs with negative values. Average length of the street segments is 137.14 meters. In addition to completing the audits, segments were photographed for off-site examination.

3.5. Results and discussion

The results of the walkability audits were transferred to a database for analysis. In addition to segment type, PEDS organizes the audit items by category, of which there are four (environment, pedestrian facility, road attributes, and walking environment), and two subjective assessments to be completed by the auditor (segment is attractive/safe for walking). In order to analyze the information contained in the audits, contingency tables were prepared. Contingency tables categorized street segment data by the sign of the filter of their respective DAs. Two classifications were used: positive, for segments in DAs with higher-than-expected use of transit (which are anticipated to display more walkability traits); and negative, for segments in DAs with lower-than-expected use of transit (which are anticipated to have fewer walkability attributes).

With the exception of sidewalk connectivity, which was audited as a continuous variable, the attributes were categorical. Sidewalk connectivity was categorized for analysis using a contingency table. All contingency tables were assessed by means of the χ^2 test of independence, whereby the observed frequency of cases in each cell in the table is compared to the expected frequency of cases under the null hypothesis of independence.

Table 3.2 shows the summary of the tests of independence. As shown there, the null hypothesis was rejected for nine audit items, six of which correspond to the walking environment category. The table also summarizes the main findings from the analysis of contingency tables. As seen in the table, six of the significant attributes lend support to the proposed conjecture that DAs with higher-than-expected use of transit would display more walkability traits.

First, street segments were categorized as low volume and high volume. A low volume segment is defined as a street with a single lane, or two lanes without clear road markings to demarcate the lanes, and little or infrequent traffic. Streets with more than one lane and some traffic are classified as high volume. Figure 3.6 shows examples of low and high volume street segments. Low volume segments were observed with significantly higher frequency in DAs with a negative value of the spatial filter (lower than expected transit usage); whereas high volumes were observed more in DAs with positive filter values (higher than expected transit usage).

Table 3.2 Audit items and summary of independence tests.

Item	Independence test	Notes (+: positive filter; -: negative filter)
Segment type	Significant*	(+) More high volume segments, fewer low volume segments; (-) More low volume segments, fewer high volume segments.
A. Environment		
Uses in segment	Significant*	(+) More mixed uses segments, fewer single use segments; (-) More single use segments, fewer mixed uses segments.
Slope	n.s.	
Segment intersections	n.s.	
B. Pedestrian Facility		
Path condition/maintenance	n.s.	
Path obstructions	n.s.	
Buffers	n.s.	
Sidewalk width	n.s.	
Curb cuts	n.s.	
Sidewalk connectivity	Significant*	(+) Fewer segments with high connectivity, more segments with no connectivity; (-) More segments with high connectivity, fewer segments with no connectivity.
C. Road Attributes		
On street parking	n.s.	
Traffic Control Devices	n.s.	
Crosswalks	n.s.	
D. Walking Environment		
Roadway/path lighting	n.s.	
Trees shading walking area	Significant*	(+): More segments with none/very few trees, fewer segments with some/dense greenery; (-) More segments with some/dense greenery, fewer segments with non/very few trees.
Degree of enclosure	Significant*	(+) More segments with high enclosure, fewer segments with low enclosure; (-) Fewer segments with high enclosure, more segments with low enclosure.
Cleanliness	Significant*	(+) More segments with low cleanliness score, fewer segments with high cleanliness score; (-) More segments with high cleanliness score, fewer segments with low cleanliness score.
Articulation in building designs	n.s.	
Building setbacks	Significant*	(+) More segments with smaller setbacks, fewer segments with large setbacks; (-) Fewer segments with smaller setbacks, more segments with large setbacks.
Building height	Significant*	(+) More segments with multi-floor buildings, fewer segments with low-rise buildings; (-) Fewer segments with multi-floor buildings, more segments with low-rise

		buildings.
Bus stops	Significant*	(+) More segments with a bus stop, fewer segments with no bus stop; Fewer segments with a bus stop, more segments no bus stop.
Subjective Assessment		
Segment is attractive for walking	n.s.	
Segment is safe for walking	n.s.	

* $p \leq 0.05$; n.s.: not significant.

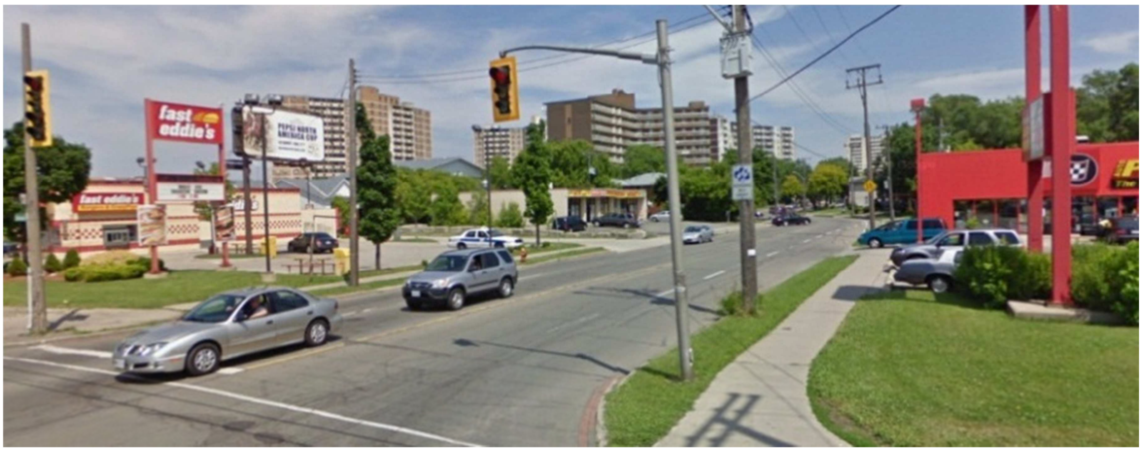

Figure 3.6 Low volume street segment (top) and high volume street segment (bottom).

The attribute of land uses in the segment was audited as a checklist with different types of land uses (e.g. housing, office/institution, industrial, recreational etc.). For analysis, the attribute was reclassified as single use, if only one type of land use was observed in the segment, and mixed if two or more land use types were observed. The contingency table for this item appears in Table 3.3. It can be seen that segments with mixed uses were significantly more common in DAs where transit usage is more common than predicted by the model (35 segments versus 22.2 expected). In contrast, DAs with a negative value of the filter, those lower uses of transit than predicted by the model, had significantly more segments characterized by single uses (50) than expected (37.2).

Table 3.3 Uses in segment: contingency table.

		Uses in segment		
		Single use	Mixed uses	Total
Positive spatial filter (more transit use than predicted)	Count	25	35	60.0
	Expected	37.8	22.2	
Negative spatial filter (less transit use than predicted)	Count	50	9	59.0
	Expected	37.2	21.8	
	Total	75	44	119.0

Two items, the degree of enclosure and size of setbacks, are indicators of compact development. Enclosure is the space between contiguous buildings, and high enclosure indicates that buildings sit closer together. A setback is the separation between the building and the sidewalk. The contingency tables for these two items are shown in Tables 3.4 and 3.5. As seen in the tables, DAs with positive spatial filters tend to be more compact, both in terms of the frequency of segments with higher enclosure and smaller setbacks. DAs with negative filters (i.e. lower than expected use of transit) display the opposite tendency, towards segments with buildings with little or no enclosure and especially larger setbacks.

Table 3.4 Degree of enclosure: contingency table.

		Degree of enclosure		
		Highly enclosed	Little/no enclosure	Total
Positive spatial filter	Count	18	42	60.0
	Expected	12.6	47.4	
(more transit use than predicted)				
Negative spatial filter	Count	7	52	59.0
	Expected	12.4	46.6	
(less transit use than predicted)				
	Total	25	94	119.0

Table 3.5 Setbacks: contingency table.

		Setbacks		
		> 20’ from sidewalk	< 20’ of sidewalk	Total
Positive spatial filter	Count	33	27	60.0
	Expected	44.9	15.1	
(more transit use than predicted)				
Negative spatial filter	Count	56	3	59.0
	Expected	44.1	14.9	
(less transit use than predicted)				
	Total	89	30	119.0

Another indicator of the built environment is building height. This item was recorded in the audit as low- (1-2 stories), medium- (3-5 stories), and high-rise (taller than 5 stories). Due to the small number of high-rise buildings, the original three categories were re-coded as low-, and medium/high-rise. As seen in Table3.6, areas where use of transit is over-predicted by the model tend to have more segments with low-rise buildings, whereas taller buildings are a significantly more common sight in DAs where transit shares are higher than predicted by the model.

Table 3.6 Building height: contingency table.

		Building height		
		Low-rise	Medium/High-rise	Total
Positive spatial filter	Count	53	7	60.0
	Expected	56.0	4.0	
(more transit use than predicted)				
Negative spatial filter	Count	58	1	59.0
	Expected	55.0	4.0	
(less transit use than predicted)				
	Total	111	8	119.0

The attributes concerning land uses, building enclosure, setbacks, and building height, are illustrated in Figures 3.7 and 3.8.

Figure 3.7 Examples of street segments with mixed land uses (residential and commercial), high enclosure and small setbacks.

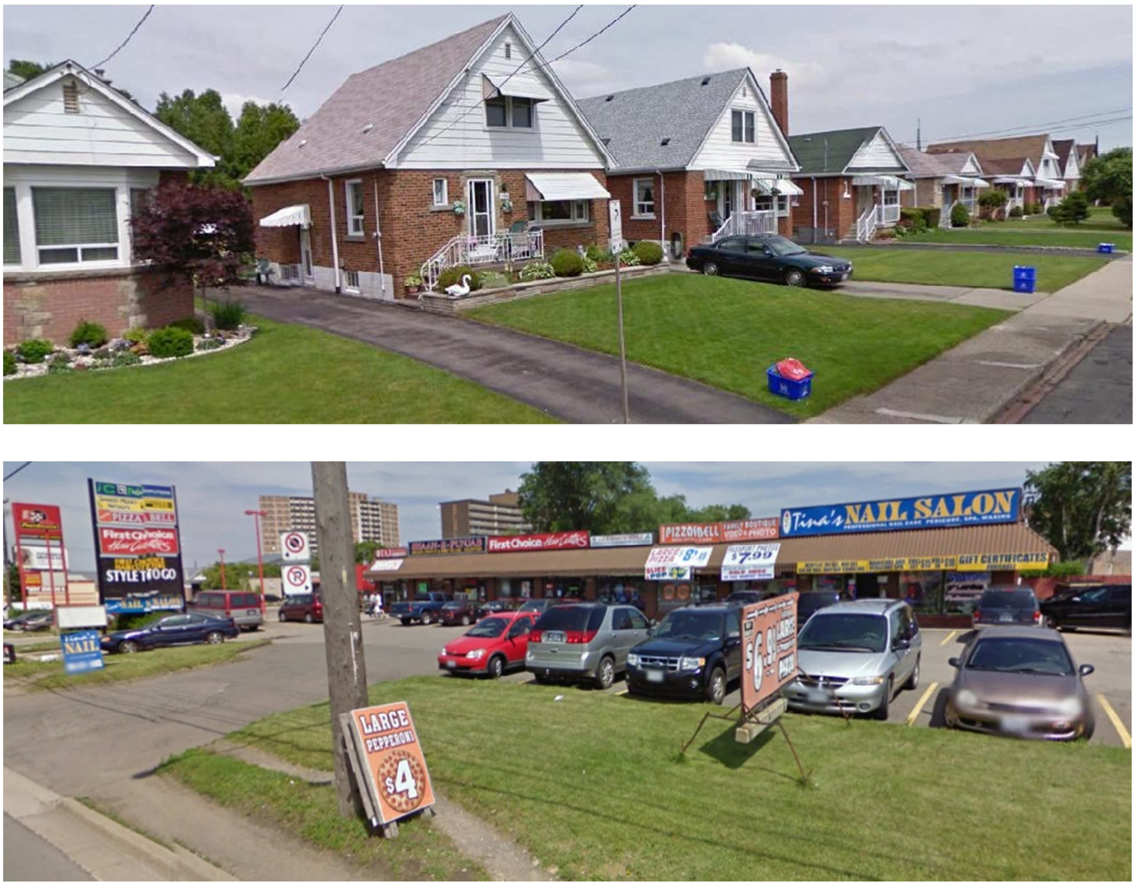

Figure 3.8 Examples of street segments with single land use, low enclosure, large setbacks, and low-rise buildings.

Figure 3.7 presents two examples of segments with mixed residential and commercial land uses, buildings that in general have smaller setbacks and higher degrees of enclosure, with little or no space between them. According to the analysis of audit data, these are all attributes associated with greater use of transit than predicted by the model. Figure 3.8 in contrast shows two examples of segments characterized by a single land use, in one case residential, in the other commercial. In addition to the monotonous streetscape in terms of uses, these segments have much larger setbacks, the buildings have little or no enclosure, and they are consistently low-rise along the segment. These are, according to the analysis of the contingency tables, attributes observed with significantly higher frequency in DAs with lower transit usage compared to the shares predicted by the model.

The last item in the category of walking environment is the presence of bus stops. This item was audited as bus stop with shelter, bus stop with bench, bus stop with signage only, or no bus stop. However, due to low variability in the first three categories, this item was reclassified according to whether a bus stop exists in the segment, or not. The analysis indicates that segments with a bus stop are significantly more common in DAs where transit usage is higher than predicted by the model, whereas the converse is true for DAs where the model over-predicts the transit share.

While the previous six items align well with the prior expectations about walkable environments, three significant attributes (sidewalk connectivity, presence of trees and shading, cleanliness) provide somewhat counter-intuitive results.

Sidewalk connectivity, the extent to which sidewalks are connected across streets, was categorized in three groups, i.e., no connectivity, low connectivity, and high connectivity. The analysis shows that there is no difference between observed and expected counts in terms of low connectivity for DAs with higher- and lower-than-expected use of transit. However, the number of segments where sidewalks have five or more connections is greater than expected in the DAs with negative values of the spatial filter, as opposed to DAs with positive values. The number of trees shading the walking area was audited as none or very few (less than 25% covered), some (25-75% covered), and many/dense (more than 75% covered). Due to the low frequency of segments with many trees, this item was re-coded into two classes: none/very few and some/many/dense. Contrary to the expectation that trees might increase the attractiveness of a street segment for walking, trees are more commonly found in street segments in DAs where the usage of transit is lower than predicted by the model. A similar result is found for the attribute of cleanliness. Overall cleanliness and building maintenance was coded as poor/fair (much or some litter/graffiti/broken facilities), and good (no litter/graffiti/broken facilities). According to the contingency table, there were significantly more segments with good cleanliness rating in DAs where the use of transit is lower than expected, and vice-versa.

In summary, the weight of the evidence tends to favor the initial conjecture, namely, that micro-scale factors in areas where transit is chosen as a mode of transportation more commonly than predicted by the model, would be more supportive of active travel. The overall picture that emerges of an area where transit is used more frequently than expected, corresponds to a pedestrian environment with higher volume streets, mixed land uses and more compact development (smaller setbacks and higher enclosure), and street segments where bus stops are available. This is all consistent with previous research indicating that mixed land uses and pedestrian-oriented designs reduce car trips and encourage people to walk, bike, and use transit (Cervero and Kockelman, 1997; Cervero et al., 2009b; Estupiñán and Rodriguez, 2008; Ewing, 1995; Frank and Pivo, 1995; Kitamura et al., 1997; Loutzenheiser, 1997). In contrast, it is found that areas where transit is used less frequently than expected, tend to be have low volume streets, more segments characterized by single uses, and less compact development (greater setbacks, little or no enclosure, and low-rise buildings) that is less-well serviced by transit (fewer segments have a bus stop).

The results with respect to sidewalk connectivity, presence of greenery, and overall cleanliness, initially seem counter-intuitive, but ultimately, it could be argued, are not unreasonable due to possible correlations with other factors. For instance, segments with mixed uses, which according to the analysis of this chapter and previous research are more supportive of walking, will tend to attract more people and other forms of traffic as well, such as commercial vehicles loading or unloading goods. This may plausibly reduce the overall cleanliness of the street segments and/or their sidewalks. Single land use segments, in particular residential, may be pristine, but as suggested by the analysis, also less conducive for walking. Another example is the number of trees in a segment. Since the analysis suggests that smaller setbacks are frequently seen in DAs where transit is used with higher than expected frequency, this necessarily means that the segments have less space to accommodate vegetation. Thus, a final caveat is in order. Correlation between attributes (negative in these two cases) is more appropriately seen as indicative of preferable attributes (mixed land uses over cleanliness), than a path to explain causality (cleaner streets leading to less walking).

3.6. Summary and conclusions

In this chapter a novel approach was proposed to systematically select sites for walkability audits. This approach is based on the conjecture that the unexplained portion of a meso-scale model of travel behavior can be partially attributed to missing attributes at the micro-scale level. Examination of the spatial pattern of the unexplained component of the model can assist in the selection of sites for conducting audits. An attractive feature of this approach is that sites can be selected after controlling for an arbitrary number of confounding factors in the model. The unexplained part can then reasonably be assumed to be independent of any variables included in the model. The use of a spatial filtering approach, moreover, means that the pattern is undisturbed by genuinely random variation.

A case study was presented to provide a proof of principle of the proposed approach. Based on a model of transit shares at the DA level in the City of Hamilton, eight DAs (out of a total of 761 DAs) were selected, based on the examination of the spatial filter retrieved for the model. Walkability audits of these DAs were conducted using PEDS, covering a total of 119 segments. Audit data were summarized in contingency tables and analyzed using χ^2 test of independence. The conjecture of this chapter that the patterns of over- and under-estimation of model predictions can be related to micro-scale environments is supported by the analysis of transit shares and the attributes of the pedestrian environment in neighborhoods across the City of Hamilton.

Substantively, it was found that DAs where transit use was higher than predicted tend to have more active, diverse, and compact street segments, with good transit access. These attributes seem to be accompanied by lower connectivity, fewer trees, and lower overall cleanliness and building maintenance ratings. In contrast, DAs with lower than expected use of transit tend to be more homogeneous and less compact, but have more connected segments, trees, and higher cleanliness ratings. An intriguing finding is that the majority of significant factors in the analysis of micro-scale environments belong to the walking environment class in the PEDS instrument.

The case study provides persuasive evidence that the proposed model-based selection strategy can be used to better target limited resources for field-based work, and that it is capable of producing valuable insights into the micro-level factors that affect travel behavior.

Chapter 4 Conclusions

The objective of this thesis has been to examine the way in which accessibility and the built environment affect the use of transit, using as a case study the Hamilton Census Metropolitan Area in Canada. This examination was based on information about the proportion of transit users at the level of Dissemination Areas in the city. As explanatory factors, a number of population socio-economic characteristics, transit level of service, accessibility by transit, and built environment characteristics were considered. The thesis was composed of two major elements. In Chapter 2, the focus was on transit modal shares, and the use of a generalized linear (logistic) model for proportions. It is well-established that the analysis of spatial data, such as that considered in the examination of transit that is the focus of the thesis, is characterized by spatial effects, in particular spatial autocorrelation. Spatial autocorrelation is a statistical nuisance. In particular, autocorrelation may compound the effect of over-dispersion, and thus lead to misleading inference. In this work, autocorrelation was treated by means of the eigenvector-based spatial filtering approach of Griffith (2000, 2003, 2004). The results from the analysis in Chapter 2 suggests that estimating an over-dispersion parameter alleviates to some extent the inflation of the variance i.e. over-dispersion parameter, but resolving the autocorrelation situation leads to even sharper inference.

Three sets of models were compared as follows: 1) ignore over-dispersion and autocorrelation; 2) estimate the over-dispersion parameter without the spatial filter; and 3) jointly estimate a dispersion parameter and use spatial filtering to control for spatial autocorrelation. From the final model i.e. joint estimation of dispersion parameter and use spatial filter to control for autocorrelation, it was found that along with the previously explored built environment and transit level of service variables, 'accessibility by transit' is also an important determinant of transit modal share analysis. Accessibility by transit contributes positively to usage of transit. This variable retained its significance in the final model, even after other measures of the transit level of service and the built environments dropped from the analysis. The model also confirms previous findings regarding modal share of transit. Dwelling density is important for increasing the mode share of transit. This is the only built environment variable which is significant after applying the spatial filtering approach. Frequency of transit service can also help to increase transit mode share, although the analysis of level of service appears to be relevant at a meso-scale, considering the overall levels of service beyond dissemination areas.

The presence of autocorrelation, which was considered a statistical nuisance in Chapter 2, was also an opportunity for discovery in Chapter 3. In this chapter, a novel method was proposed to systematically select sites for conducting further in-depth field work. The approach is based on the examination of the spatial filter used to remove autocorrelation in the model of transit shares. The proposed approach is based on the conjecture that the unexplained portion of a meso-scale model of travel behavior can be partially attributed to missing attributes at the micro-scale level. Examination of the spatial pattern of the unexplained component of the model can assist in the selection of sites for conducting audits. An attractive feature of this approach is that sites are selected after controlling for an arbitrary number of confounding factors in the model. The unexplained part can reasonably be assumed to be independent of any variables included in the model. The use of a spatial filtering approach, moreover, means that the pattern is undisturbed by genuinely random variation. As a proof of principle, eight dissemination areas were chosen based on the values of the spatial filter, and the transit shares of the zones. Walkability audits using the Pedestrian Environment Data Scan (PEDS) were conducted within the systematically selected dissemination areas. The paper version of PEDS was converted into electronic version using ArcPad Application Builder 7.0.1 and then transferred into a handheld device to conduct the walkability audit in the field.

From this study it was found that dissemination areas where transit usage is higher than predicted tend to have more active, diverse, and compact street segments, with good transit access. These attributes seem to be accompanied by lower connectivity, fewer trees, and lower overall cleanliness and building maintenance ratings. In contrast, dissemination areas with lower than expected use of transit tend to be more homogeneous and less compact, but have more connected segments, trees, and higher cleanliness ratings. The case study demonstrates that the proposed model-based selection strategy can be used to better target limited resources for field-based work, and produce valuable insights into the micro-level factors that may affect travel behavior.

In terms of future research, the methodology proposed in Chapter 3 is very general and has broad applicability. For instance, the meso-scale model used in the case study is of transit shares for the journey to work, and therefore is specific to a mode of transportation and a population segment. The specific origin of the spatial filter used to select walkability audit sites could well restrict its generality. It is reasonable to think that some of the built environment attributes that a commuter walking to a transit stop may consider important, could in fact be irrelevant to a person walking for leisure, and vice-versa. An intriguing avenue for future research therefore is to use models of different travel behaviors (e.g. walking for leisure, auto ownership) complemented by walkability audits, in order to assess whether some features of the built environment influence particular forms of mobility, and on the contrary if they affect more mobility more generally. In a similar way, the methodology proposed here can be applied to different contexts, for instance other cities, or to specific population segments (e.g. children or seniors).

References

Agrawal, A.W., Schlossberg, M., Irvin, K., 2008. How Far, by Which Route and Why? A Spatial Analysis of Pedestrian Preference. *Journal of Urban Design* 13, 81 - 98.

Alam, B.M., 2009. Transit Accessibility to Jobs and Employment Prospects of Welfare Recipients Without Cars A Study of Broward County, Florida, Using Geographic Information Systems and an Econometric Model. *Transportation Research Record* 2110, 78-86.

Alshalalfah, B.W., Shalaby, A.S., 2007. Case study: relationship of walk access distance to transit with service, travel, and personal characteristics. *Urban Planning and Development* 133, 114-118.

American Public Transit Association, 2010. *The Public Transportation Fact Book, 61st Edition*. APTA, Washington, D.C.

Anselin, L., 1988. *Spatial Econometrics: Methods and Models*. Kluwer Dordrecht.

Benenson, I., Martens, K., Rofe, Y., 2010. Measuring the Gap Between Car and Transit Accessibility Estimating Access Using a High-Resolution Transit Network Geographic Information System. *Transportation Research Record* 2144, 28-35.

Blumenberg, E., Ong, P., 2001. *Cars, buses, and jobs - Welfare participants and employment access in Los Angeles*.

Boarnet, M.G., Sarmiento, S., 1998. Can land-use policy really affect travel behaviour? A study of the link between non-work travel and land-use characteristics. *Urban Studies* 35, 1155-1169.

Boots, B., Tiefelsdorf, M., 2000. Global and local spatial autocorrelation in bounded regular tessellations. *Journal of Geographical Systems* 2, 319-348.

Boroski, J., Faulkner, T., Arrington, B., Mori, S., Parker, T., Mayer, D., 2002. Statewide Transit-Oriented Development Study, Parking and TOD: Challenges and Opportunities (Special Report). Business, Transportation and Housing Agency, California Department of Transportation.

Brown, J.R., Thompson, G.L., 2008. The relationship between transit ridership and urban decentralisation: Insights from Atlanta. *Urban Studies* 45, 1119-1139.

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

Cervero, R., 1988. Land-use mixing and suburban mobility. *Transportation Quarterly* 42, 429-446.

Cervero, R., Jin, M., Miller, M.A., 2009a. *Direct ridership model of bus rapid transit in Los Angeles County*. UC Berkeley Center for Future Urban Transport, Berkeley.

Cervero, R., Kockelman, K., 1997. Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D-Transport and Environment* 2, 199-219.

Cervero, R., Sandoval, O.S., Landis, J., 2002. Transportation as a stimulus of welfare-to-work - Private versus public mobility. *Journal of Planning Education and Research* 22, 50-63.

Cervero, R., Sarmiento, O.L., Jacoby, E., Gomez, L.F., Neiman, A., 2009b. Influences of built environments on walking and cycling: lessons from Bogota. *International Journal of Sustainable Transportation* 3, 203-226.

Chow, L.F., Zhao, F., Liu, X., Li, M.T., Ubaka, I., 2007. Transit ridership model based on geographically weighted regression. *Transportation Research Record* 1972, 105-114.

Clarke, P., Ailshire, J., Melendez, R., Bader, M., Morenoff, J., 2010. Using Google Earth to conduct a neighborhood audit: Reliability of a virtual audit instrument. *Health & Place* 16, 1224-1229.

Cliff, A., Ord, J.K., 1969. The Problem of Spatial Autocorrelation, in: Scott, A. (Ed.), *London Papers in Regional Science*. Pion, London, pp. 25-55.

Cliff, A.D., Ord, J.K., 1973. *Spatial Autocorrelation*. Pion, London

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

Clifton, K.J., Smith, A.D.L., Rodriguez, D., 2007. The development and testing of an audit for the pedestrian environment. *Landscape and Urban Planning* 80, 95-110.

Dannenberg, A.L., Cramer, T.W., Gibson, C.J., 2005. Assessing the walkability of the workplace: A new audit tool. *American Journal of Health Promotion* 20, 39-44.

Day, K., Boarnet, M., Alfonzo, M., Forsyth, A., 2006. The Irvine-Minnesota inventory to measure built environments development. *American Journal of Preventive Medicine* 30, 144-152.

Demiroz, Z., 2005. *The applications of new urbanist communities in Istanbul, Turkey: a success of a failure?* Cornell University.

Emery, J.E., Crump, C., Bors, P., 2003. Reliability and Validity of Two Instruments Designed to Assess the Walking and Bicycling Suitability of Sidewalks and Roads. *American Journal of Health Promotion* 18, 38-46.

Estupiñán, N., Rodriguez, D.A., 2008. The relationship between urban form and station boardings for Bogota's BRT. *Transportation Research Part A-Policy and Practice* 42, 296-306.

Ewing, R., 1995. Beyond density, mode choice, and single-purpose trips. *Transportation Quarterly* 49, 15-24.

Frank, L., Pivo, G., 1995. Impacts of mixed use and density on utilization of three modes of travel: single-occupant vehicle, transit and walking. *Transportation Research Record* 1466, 42-52.

Frank, L.D., Engelke, P.O., Schmid, T.L., 2003. *Health and community design: the impact of the built environment on physical activity*. Island Press, Washington, D.C.

Getis, A., 2007. Reflections on spatial autocorrelation. *Regional Science and Urban Economics* 37, 491-496.

Getis, A., 2008. A history of the concept of spatial autocorrelation: A geographer's perspective. *Geographical Analysis* 40, 297-309.

Gilmour, H., 2007. Physically active Canadians, *Health Reports*. Statistics Canada, pp. 45-66.

Greenwald, M.J., Boarnet, M.G., 2001. Built environment as determinant of walking behavior - Analyzing nonwork pedestrian travel in Portland, Oregon. *Land Development and Public Involvement in Transportation*, 33-42.

Grengs, J., 2004. Measuring change in small-scale transit accessibility with geographic information systems - Buffalo and Rochester, New York. *Transportation Research Record* 1887, 10-17.

Griffith, D.A., 1988. *Advanced Spatial Statistics: Special Topics in the Exploration of Quantitative Spatial Data Series*. Kluwer, Dordrecht.

Griffith, D.A., 2000. A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems* 2, 141-156.

Griffith, D.A., 2003. *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization?"<L*. Springer-Verlag, Berlin.

Griffith, D.A., 2004. A spatial filtering specification for the autologistic model. *Environment and Planning A* 36, 1791-1811.

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

Guo, J.Y., Bhat, C.R., 2007. Operationalizing the concept of neighborhood: Application to residential location choice analysis. *Journal of Transport Geography* 15, 31-45.

Guo, Z., 2009. Does the pedestrian environment affect the utility of walking? A case of path choice in downtown Boston. *Transportation Research Part D-Transport and Environment* 14, 343-352.

Haining, R., 1990. *Spatial data analysis in the social and environmental sciences*. Cambridge University Press, Cambridge.

Handy, S., Niemeier, D., 1997. Measuring Accessibility: An Exploration of Issues and Alternatives. *Environment and Planning A* 29, 1175-1194.

Hess, D.B., 2005. Access to employment for adults in poverty in the Buffalo-Niagara region. *Urban Studies* 42, 1177-1200.

Hess, D.B., 2009. Access to public transit and its influence on ridership for older adults in two U.S. cities. *Journal of Transport and Land Use* 2, 3-27.

Hess, P.M., Moudon, A.V., Snyder, M.C., Stanilov, K., 1999. Site Design and Pedestrian Travel. *Transportation Research Record* 1674, 19.

Hoehner, C.M., Ramirez, L.K.B., Elliott, M.B., Handy, S.L., Brownson, R.C., 2005. Perceived and objective environmental measures and physical activity among urban adults. *American Journal of Preventive Medicine* 28, 105-116.

Hsiao, S., Lu, J., Sterling, J., Weatherford, M., 1997. Use of Geographic Information System for Analysis of Transit Pedestrian Access. *Transportation Research Record* 1604, 50-59.

Johnson, A., 2003. Bus transit and land use: illuminating the interaction. *Public Transportation* 6, 21-39.

Kamatu, J.N., Hutchins, T.N., Wagner, A.G., Lambert, J.H., 2007. Mapping and accessibility analysis of Virginia transit systems. *2007 IEEE Systems and Information Engineering Design Symposium*, 230-233.

Kawabata, M., 2009. Spatiotemporal dimensions of modal accessibility disparity in Boston and San Francisco. *Environment and Planning A* 41, 183-198.

Kawabata, M., Shen, Q., 2006. Job accessibility as an indicator of auto-oriented urban structure: a comparison of Boston and Los Angeles with Tokyo. *Environment and Planning B-Planning & Design* 33, 115-130.

Kawabata, M., Shen, Q., 2007. Commuting inequality between cars and public transit: the case of the San Francisco Bay Area, 1990-2000. *Urban Studies* 44, 1759-1780.

Kitamura, R., Mokhtarian, P.L., Laidet, L., 1997. A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. *Transportation* 24, 125-158.

Lei, T.L., Church, R.L., 2010. Mapping transit-based access: integrating GIS, routes and schedules. *International Journal of Geographical Information Science* 24, 283-304.

Loutzenheiser, D.R., 1997. Pedestrian Access to Transit Model of Walk Trips and their Design and Urban Form. *Transportation Research Record* 1604, 49.

Matas, A., Raymond, J.L., Roig, J.L., 2010. Job Accessibility and Female Employment Probability: The Cases of Barcelona and Madrid. *Urban Studies* 47, 769-787.

May, A.D., Shepherd, S.P., Timms, P.M., 2000. Optimal transport strategies for European cities. *Transportation* 27, 285-315.

McCullagh, P., Nelder, J.A., 1989. *Generalized Linear Models*. Chapman & Hall, London.

McMillan, T.E., Cubbin, C., Parmenter, B., Medina, A.V., Lee, R.E., 2010. Neighborhood sampling: how many streets must an auditor walk? *International Journal of Behavioral Nutrition and Physical Activity* 7.

Millington, C., Thompson, C.W., Rowe, D., Aspinall, P., Fitzsimons, C., Nelson, N., Mutrie, N., Sparcoll, 2009. Development of the Scottish Walkability Assessment Tool (SWAT). *Health & Place* 15, 474-481.

Minocha, I., Sriraj, P.S., Metaxatos, P., Thakuriah, P., 2008. Analysis of transit quality of service and employment accessibility for the greater Chicago, Illinois, Region. *Transportation Research Record*, 20-29.

Moudon, A.V., Lee, C., 2003. Walking and bicycling: an evaluation of environmental audit instruments. *The Science of Health Promotion* 18, 21-37.

Murray, A.T., 2003. A coverage model for improving public transit system accessibility and expanding access. *Annals of Operations Research* 123, 143-156.

O'Sullivan, S., Morrall, J., 1997. Walking distances to and from light-rail transit stations. *Transportation Research Record* 1538, 26.

Olszewski, P., Wlbowo, S.S., 2005. *Using equivalent walking distance to assess pedestrian accessibility to transit stations in Singapore.*

Ortúzar, J.D., Willumsen, L.G., 2001. *Modelling Transport*. Wiley, New York.

Parmenter, B., McMillan, T., Cubbin, C., Lee, R.E., 2008. Developing Geospatial Data Management, Recruitment, and Analysis Techniques for Physical Activity Research. *Journal of the Urban and Regional Information Systems Association* 20, 13-19.

Pikora, T.J., Bull, F.C.L., Jamrozik, K., Knuiman, M., Giles-Corti, B., Donovan, R.J., 2002. Developing a reliable audit instrument to measure the physical environment for physical activity. *American Journal of Preventive Medicine* 23, 187-194.

Rodriguez, D.A., Brisson, E.M., Estupinan, N., 2009. The relationship between segment-level built environment attributes and pedestrian activity around Bogota's BRT stations. *Transportation Research Part D-Transport and Environment* 14, 470-478.

Rodriguez, D.A., Joo, J., 2004. The relationship between non-motorized mode choice and the local physical environment. *Transportation Research Part D-Transport and Environment* 9, 151-173.

Rundle, A.G., Bader, M.D.M., Richards, C.A., Neckerman, K.M., Teitler, J.O., 2011. Using Google Street View to Audit Neighborhood Environments. *American Journal of Preventive Medicine* 40, 94-100.

Saelens, B.E., Sallis, J.F., Frank, L.D., 2003. Environmental correlates of walking and cycling: Findings from the transportation, urban design, and planning literatures. *Annals of Behavioral Medicine* 25, 80-91.

Sallis, J.F., 2009. Measuring Physical Activity Environments: A Brief History. *American Journal of Preventive Medicine* 36, S86-S92.

Sanchez, T.W., 1999. The connection between public transit and employment - The cases of Portland and Atlanta. *Journal of the American Planning Association* 65, 284-296.

Sanchez, T.W., Shen, Q., Peng, Z.R., 2004. Transit mobility, jobs access and low-income labour participation in US metropolitan areas1. *Urban Studies* 41, 1313-1331.

Shen, Q., 2001. A spatial analysis of job openings and access in a US metropolitan area. *Journal of the American Planning Association* 67, 53-68.

Statistics Canada, 1999. Passanger Bus and Urban Transit Statistics, Ottawa.

Statistics Canada, 2006. The 2005 General Social Survey - Cycle 19 Time Use: Public Use Microdata File Documentation and User Guide, Product No. 12M0019-GPE.

Taylor, B.D., Miller, D., Analyzing the determinants of transit ridership using a two-stage least square regression of a national sample of urbanized areas, Washington, D.C.

Taylor, B.D., Miller, D., 2003. Analyzing the determinants of transit ridership using a two-stage least square regression of a national sample of urbanized areas, *82th Annual Meeting of the Transportation Research Board*, Washington, D.C.

Taylor, B.D., Miller, D., Iseki, H., Fink, C., 2009. Nature and/or Nurture? Analyzing the Determinants of Transit Ridership across US Urbanized Areas. *Transportation Research Part A-Policy and Practice* 43, 60-77.

Thakuriah, P., Metaxatos, P., 2000. Effect of residential location and access to transportation on employment opportunities. *Transportation Research Record* 1726, 24-32.

Tiefelsdorf, M., Boots, B., 1995. The Exact Distribution of Moran I. *Environment and Planning A* 27, 985-999.

Tiefelsdorf, M., Griffith, D.A., 2007. Semiparametric filtering of spatial auto correlation: the eigenvector approach. *Environment and Planning A* 39, 1193-1221.

Tjepkema, M., 2006. Adult obesity, *Health Reports*. Statistics Canada, pp. 9-25.

TRB and Institute of Medicine, 2005. Does the Built Environment Influence Physical Activity? Examining the Evidence. Transportation Research Board, Washington, D.C.

Voith, R., 1998. The Downtown Parking Syndrome: Does Curing the Illness Kill the Patient? *Business Review* 1, 3-14.

Warburton, D.E., Nicol, C.W., Bredin, S.S., 2006. Health benefits of physical activity: the evidence. *Canadian Medical Association Journal* 174, 801-809.

Werner, C.M., Brown, B.B., Gallimore, J., 2010. Light rail use is more likely on "walkable" blocks: Further support for using micro-level environmental audit measures. *Journal of Environmental Psychology* 30, 206-214.

Yanmaz-Tuzel, O., Ozbay, K., 2010. Impacts of Gasoline Prices on New Jersey Transit Ridership. *Transportation Research Record* 2144, 52-61.

Yi, C., 2006. Impact of public transit an employment status - Disaggregate analysis of Houston, Texas. *Transportation Research Record*, 137-144.

Zhao, F., Chow, L.F., Li, M.T., Ubaka, I., Gan, A., 2003. Forecasting transit walk accessibility - Regression model alternative to buffer method. *Transit: Planning and Development, Management and Performance, Marketing and Fare Policy, and Intermodal Transfer Facilities - Public Transit*, 34-41.

Ziari, H., Keymanesh, M.R., Khabiri, M.M., 2007. Locating stations of public transportation vehicles for improving transit accessibility. *Transport* 22, 99-104.