Walking in the City of Ottawa: Pedestrian Volume and its Relationship with Walkability

by

Marc Bouchard

A thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

In partial fulfillment of the requirements for the

M.Sc. degree in Geography

Department of Geography, Environment and Geomatics

Faculty of Arts

University of Ottawa

Table of Contents

Tables

Figures

Equations

Abstract

Walkability indices are currently used for a wide range of research and commercial applications. Few studies have examined the relationship between walkability indices and measured pedestrian volume or walking rates, nor explored moderators of pedestrian volume such as weather. With 14 years of traffic study data from the City of Ottawa, a spatial auto-regressive (SAR) multi-level model (MLM) was used to understand the proportion of variance in walking explained by the commercial Walk Score® index and selected weather variables. Modeling revealed that a significant proportion of pedestrian volume at a given location in Ottawa, including its spatial lag, was explained by the corresponding Walk Score® value and its spatial lag (51.45%). Furthermore, weather expressed as a combination of 'felt' temperature, presence or absence of precipitation, and percent cloud cover, accounted for 2.79% of the variance in walking. These findings indicate that walkability indices may provide value as cost-effective engineering and urban planning tools.

Résumé

Les indices de marchabilité sont présentement utilisés pour un large éventail d'applications en recherche et au niveau commercial. Peu d'études ont examiné le rapport entre les indices de marchabilité et le volume de trafic piétonnier, ni les modérateurs du trafic piétonnier tel que les conditions météorologiques. Avec des données d'études de trafic acquises par la ville d'Ottawa sur un période de 14 ans, un modèle multi-niveau auto-régressif spatial a été utilisé pour comprendre la proportion de variance qui s'explique par l'indice de marchabilité commercial Walk Score®, ainsi qu'une sélection de variables météorologiques. La modélisation a révélé qu'une proportion importante de la variance du volume de trafic piétonnier à un endroit donné à Ottawa, y compris son décalage spatial, s'explique par la valeur de l'index Walk Score® correspondante ainsi que le décalage spatial de cette dernière (51,45%). De plus, les conditions météorologiques, exprimées par une combinaison de la température 'ressentie', de la présence ou de l'absence de précipitations, et du pourcentage de couverture nuageuse, ont compté pour 2,79% de la variance du trafic piétonnier. Ces résultats indiquent que les indices de marchabilité ont le potentiel d'apporter une valeur ajoutée en tant qu'outils rentables pour l'ingénierie et la planification urbaine.

Acknowledgements

I am immensely grateful to my supervisor Dr. Michael Sawada who, first of all, believed in my thesis idea, but also took me under his wing as a mature graduate student. His feedback, keen insights and guidance were instrumental in making this project a reality. I couldn't have asked for a better supervisor, and my only regret is to not have taken the opportunity to meet with him more often while developing this project.

I also want to thank my committee members, Dr. Konrad Gajewski and Dr. Elizabeth Kristjansson, who also provided valuable feedback on my initial proposal, which shaped the project I have here today.

This work would not have been possible without the support of staff at the City of Ottawa. Kornel Mucsi and Gill Wilson were supportive of my project and got their teams involved in assisting me. In particular, Ibrahim Conteh and Linda Marin took the time to meet with me, put together a presentation, and review their internal data collection procedures and data structure. They and their team also worked to compile and supply the data to me in a raw electronic format. The City provided free access to 14 years-worth of valuable pedestrian traffic data, without which this project would not exist.

Stantec Consulting Ltd was also supportive while I worked on this project part-time. They provided the flexibility and some financial compensation needed for me to complete my M.Sc. coursework requirements while juggling a heavy work load. Similarly, my current employer SLR Consulting Ltd., has been very flexible and has allowed me to juggle my work and studies as seamlessly as possible.

Finally, I would like to thank my wife Veronika Huta, who along this journey was wholeheartedly supportive of my efforts, and the sacrifices they entailed for our family. She was, in fact, my biggest cheerleader. At times when I doubted myself and questioned my life choices, she reassured me and gave me the encouragement I needed to find the strength to carry on, even when the choices were difficult. Her statistical expertise, especially with hierarchical linear modeling, was absolutely key to the successful outcome of this work. She is a gifted teacher, and though the husband-teacher dynamic is far from ideal, the self-learning of these more advanced statistical techniques was made significantly easier thanks to her.

Chapter 1: Introduction

Cities are grappling with escalating traffic congestion and infrastructure costs associated with increasing population density in major urban centres. Many cities worldwide are systematically working on incorporating more efficient multi-modal solutions emphasizing human-powered transport (Pucher, Buehler, & Seinen, 2011). However, forecasting demand and associated level of service for pedestrians continues to be a challenge due to a scarcity of comprehensive data and models of pedestrian behaviour (Aultman-Hall, Lane, & Lambert, 2009; Dhanani, Tarkhanyan, & Vaughan, 2017), and the high cost of collecting pedestrian volume data. One potential avenue to enhance our understanding of pedestrian behaviour would be to leverage the concept of walkability. Relatively few studies have examined the relationships between walkability indices or their component variables and the empirically measured usage of the pedestrian network (Hajrasouliha & Yin, 2015; Ozbil, Peponis, & Stone, 2011; Hillier & Iida, 2005). There is an identified need to validate the aggregated effect of walkability indices on actual walking (Maghelal & Capp, 2011). This investigation aims to quantify the relationship of walkability with measured pedestrian volumes in the City of Ottawa, Ontario, Canada, while accounting for the role of other important predictors and moderators, including weather and time of day and week, using a multi-level modeling (MLM) approach.

Walkability

Walkability is generally recognized as a measure of how friendly a location is for walking (NHTSA, 2004). There is a large body of literature pertaining to walkability and its relationship with safety and health outcomes. There is also a significant body of work on the auditing and validation of walkability indices either with field data acquisition (Clifton, Smith, & Rodriguez, 2007) or with corroborating spatial data, for example using Google Street View (Lee & Talen, 2014). Many definitions of walkability exist even within the same geography, and they are often context-specific. Several are commercial in nature (Walk Score, 2015; Walkonomics, 2014; Maponics LLC, 2015), while many more walkability measures have been developed for research purposes (Duncan, Aldstadt, Whalen, & Melly, 2013; Frank, et al., 2010).

A number of walkability-related research programs are underway in Ottawa. A concept of City of Ottawa neighbourhoods was developed to study health outcomes (Parenteau M.-P., 2008). More recently, the relationship between physical activity, obesity and neighbourhood physical factors from the Ottawa Neighbourhood Study (ONS) was explored, combining the concepts of walkability with the neighbourhood model developed by Parenteau et al (Prince S.A. K. E.-M., 2011; Prince S.A. K. E.-M., 2012). An intervention study examined the impact of physical activity increase to overcome structural barriers to physical activity, based on Walk Score® (WS) and walkability from the ONS (Riley D.L., 2013). Data collection and research are on-going at the University of Ottawa with ONS partners to develop and audit improved measures of walkability that include field-measured subjective variables (Lafontaine, Sawada, & Kristjansson, 2017), as well as deep learning neural networks, thereby establishing stronger links between walkability and health outcomes. However, there are currently no documented studies in Ottawa relating walkability indices and component variables with empirically-measured pedestrian volumes.

There is currently no internationally-recognized standard operational definition of walkability. Because of the relatively fine-grained nature of the pedestrian environment, and the multitude of physical, environmental and social factors that are thought to affect walking behavior, the amount and quality of data that can be realistically collected to provide a measure of walkability varies greatly from one location to another. Furthermore, walkability indices alone are not necessarily proxy measures for actual pedestrian activity. In that sense, walkability could be interpreted as a 'potential' pedestrian activity of a certain mode or type, e.g. recreational or utilitarian. Whether or not that potential is realized is mitigated by a multitude of factors, such as weather, infrastructure, or pedestrian intent or purpose.

The commercial WS walkability index was selected for this study since it can be sampled at any location within Ottawa and is free of cost. The measure is static in time (sampled in March 2017) and is biased towards utilitarian walking (e.g. commuting and shopping), being largely based on proximity to amenities (e.g. grocery stores, markets, restaurants, retail, parks, etc), while accounting for network properties (intersection density and block length) (Walk Score Advisory Group, 2017). This characteristic of the index may reduce association with leisure-based walking trips given the importance of pedestrian purpose, or intent, in explaining walking flows (Lee & Moudon, 2006).

Pedestrian Volume

Counting to assess pedestrian presence, flow or volume in the context of the study of public spaces and urban design was pioneered by Jan Gehl in Copenhagen in the late 1960s and has received significant attention by municipal governments in Europe to this day (Gehl, 1996). The importance of pedestrian and walkability research in North America has become more prominent since the mid-1990s. North American municipal, regional and national regulatory agencies and professional associations have developed several design and planning guidance documents to address the previous void in formal pedestrian planning. In North America, pedestrian counts are typically used to determine the level-of-service (LOS) needed for pedestrians, such as sidewalk dimensions and crosswalk needs, to assess the need for pedestrian signals, and to time those signals. For instance, pedestrian LOS for sidewalks and pathways is defined by the United States Department of Transportation (USDOT) by empirically-derived functions of pedestrian volume (USDOT FHA, 2006). Design guidance for intersection crossings, including lighting and signals, was also developed by the USDOT (USDOT FHA, 2004). Design criteria for pedestrian facilities are further defined in the USDOT Highway Capacity Manual (USDOT FHA, 1998). More recently, guidance on the development of bicycle and pedestrian plans is provided in the Statewide Pedestrian and Bicycle Planning Handbook (USDOT FHA, 2014). Pedestrian facility design guidelines are also set by AASHTO and ITE (AASHTO, 2004; ITE, 2010). The ITE document supports the AASHTO guidance by providing more specific design criteria for walkable environments, including criteria typical of walkability indices like intersection density, but also level-of-service design criteria like sidewalk widths and building setbacks. In Ontario, the Ontario Ministry of Transportation defines pedestrian LOS in the Ontario Traffic Manual, based on measured pedestrian volumes (MTO, 2010). The design guidance is focused on traffic signals and integration of pedestrians with vehicular traffic only in terms of safety. Pedestrian facility design criteria in Ontario are determined based on a net 8-hour adjusted pedestrian volume (MTO, 2012). Pedestrian volume is currently the instrument of choice in North America to assess LOS requirements given its widespread adoption.

A number of North American municipalities have completed keystone pedestrian plans and studies in recent years. Notable among them are the City of New York's Pedestrian Level of Service Study (City of New York, 2006), Sacramento's new pedestrian LOS measures

(Sacramento Area Council of Governments, 2011), the City of Toronto's Walking Strategy (City of Toronto, 2009), and the City of Ottawa's draft pedestrian plan (OPP) (City of Ottawa, 2013).

In 2013, 55.2% of Canadians aged 12 and older were active during their leisure time (Statistics Canada, 2014). Based on the National Household Survey of 2011, 5.7% of Canadians walk to work. In Ottawa specifically, 7.1% of workers commute on foot (Statistics Canada, 2013). The City of Ottawa has recently put a strong emphasis on pedestrian transportation, having signed an international Charter for Walking (Walk21, 2006) in 2011, with a goal to increase the pedestrian modal share. The plan elaborates on the various benefits of walking and sets targets for morning peak period pedestrian volume for 2031.

This study utilizes hourly pedestrian count, or volume, data collected by the City of Ottawa and its contractors between 1999 and 2013. The data is considered secondary, as the locations and timing of the data collection, having taken place prior to conception of the current study, do not necessarily align with its objectives. Hence, the data may contain biases in time and space (e.g. emphasis on rush-hour daytime collection for automotive traffic studies), rather than exhibiting ideal statistical properties for the intended Multi-Level Model (MLM).

Predictors and Moderators of Walking

There is general agreement in the literature that weather has an important effect on pedestrian activity levels in public places. Weather may explain from 10 to 60% of the variance of pedestrian presence or activity (de Montigny, Ling, & Zacharias, 2012; Aultman-Hall, Lane, & Lambert, 2009; Zacharias, Stathopoulos, & Wu, 2001; Zacharias, Stathopoulos, & Wu, 2004). Air temperature has also been shown to have a direct impact on walking tempo (speed) (Rotton, 1990). Furthermore, winter conditions can impact walking and/or walkability (Chapman, Nilsson, Larsson, & Rizzo, 2017; Clarke, et al., 2017). It is therefore important to control for weather in the study of the relationship between walkability and pedestrian volume.

Human perception of comfortable weather is elastic; people develop both physiological and psychological expectations about their local weather conditions and adjust their behaviours accordingly (Knez, Thorsson, Eliasson, & Lindberg, 2009). The relationship between weather conditions and human comfort is complicated by variations in individual protection (i.e. clothing insulation, water resistance, wind resistance, style, etc.) (Błażejczyk, Epstein, Jendritzky, Staiger, & Tinz, 2012), and the effect of temperature on subjective and objective human thermal comfort is moderated by other weather parameters, notably humidity, wind and solar radiation (Maras, Schmidt, Paas, Ziefle, & Schneider, 2016). Furthermore, seasonal conditions tend to moderate the impact of temperature and precipitation on walking in northern climates where there is presence of snow on the ground in winter (Aultman-Hall, Lane, & Lambert, 2009).

Humidex (Environment and Climate Change Canada, 2015) and Windchill (Environment and Climate Change Canada, 2017) are indices developed by Environment and Climate Change Canada (ECCC) that are commonly used in Canada to assess human comfort with respect to temperature, while accounting for wind (extreme cold) and/or humidity (extreme heat). Humidex is not an ideal measure of perceived temperature as it accounts for dry bulb temperature and humidity but not solar radiation and wind. Similarly, Windchill only accounts for dry bulb temperature and windspeed, and not humidity and solar radiation. Therefore, neither of these models of human comfort with respect to temperature separately account for the full range of conditions or for all relevant weather variables. Wet Bulb Globe Temperature (WBGT) is a more effective measure of heat stress (National Weather Service, 2017), however it requires special instrumentation to measure its key component value, black globe temperature, which is not measured by ECCC weather stations. Though solar radiation is considered a key biometeorological variable (Eliasson, 2007), it is often inconsistently measured. The Universal Thermal Comfort Index (UTCI) (Bröde, Krüger, & Fiala, 2013) accounts for temperature, windspeed, relative humidity and radiation, but again requires specialized instrumentation and obtaining measurements at the human biometeorological standard height of 1.1m above ground level. Although it was not developed in the Canadian context, the Australian Apparent Temperature (AAT) index created by the Australia Bureau of Meteorology (Steadman, 1994; Steadman, 1984) provides a continuous scale of perceived temperature based on a model of heat balance in the human body. It accounts for temperature, windspeed and humidity, and optionally also accounts for solar radiation, hence it was selected as a usable measure of perceived temperature for this study. It is recognized that a measure that has been validated in our local climate, and in the context of local human response to that climate, would be preferable.

The purpose of walking has also been shown to play an important role (Lee & Moudon, 2006), and walking behaviours are associated to daily routines (Aultman-Hall, Lane, & Lambert, 2009). Part of the population in any given locale must walk to carry out their daily activities (utilitarian walking), and thus the influence of weather on that subset may be limited. Conversely, walking for the purpose of leisure may be moderated by weather conditions (Aultman-Hall, Lane, & Lambert, 2009). The 2011 Origin-Destination Survey completed in the National Capital Region indicates that trip purpose for all transport modes in Ottawa and neighbouring communities may be related to time of day (TRANS Committee, 2012). Indeed, there is evidence elsewhere that walking behaviour is associated with time of day and week (Shaaban, Muley, & Elnashar, 2017; Aultman-Hall, Lane, & Lambert, 2009; Schasberger, Raczkowski, Newman, & Polgar, 2012). A categorization of pedestrian volumes by time category (e.g. morning rush hour, mid-day, evening rush hour, weekends, holidays, etc.), could therefore potentially account to some degree for pedestrian purpose.

Research Approach

This research design was cross-sectional and relied on secondary data analysis. We studied a large dataset that is inherently nested. From an overall methodological standpoint, this investigation utilized MLM to mitigate overestimation of sample size in multiple regression. Multilevel models are a form of regression analysis that account for nested data (Raudenbush & Bryk, 2002). They are more appropriate than regular regression when analyzing data that are nested or grouped, where within-group observations tend to be more similar than between-group observations (e.g., pedestrian volumes from the same sampling location are likely to be more similar than pedestrian volumes from different sampling locations). Temporal autocorrelation of sequential observations at each sampling location in the MLM was inherently addressed conservatively by the multi-level approach, since the observations within a group do not contribute to the degrees of freedom available for analysis, hence the analysis assumes by default that the observations within a group are autocorrelated (Raudenbush & Bryk, 2002). Spatial autocorrelation between sampling locations was addressed by introducing a spatial lag to the models, similar to the approach used in Park (Park & Kim, 2014). As we expected the dependent variable to be influenced directly by its neighbouring values, a spatial lag autoregressive model was considered appropriate (Ward & Gleditsch, 2008).

The overall study hypothesis, based on findings in the literature, was that both weather and walkability account for significant portions of the variance in pedestrian volumes. It was also expected that a significant portion of the variance would not be explained by any of the available variables, given the inherent complexity of walking behaviour. However, since there is no known example of a similar MLM analysis of walkability and pedestrian volume, and given the secondary basis of the data, the exact specification of the model was subject to optimization, and the exact weather and time predictors and moderators, and their proportional contribution to the model, could not be predicted. Furthermore, the models developed herein were correlational; thus, it was not possible to determine whether the people who walk more chose to frequent more walkable areas, or if the built environment influenced the walking (Ewing, 2005). Study hypotheses included the following:

- Walk Score® is positively related to pedestrian volume
- Temperature increase will lead to pedestrian volume increase
- Precipitation increase will lead to pedestrian volume decrease
- Winter conditions will lead to pedestrian volume decrease
- Weather will have less effect on pedestrian volume during utilitarian walks, compared to leisure walks
- Winter conditions will moderate the effect of weather on pedestrian volume

It was our hope that the model, in addition to providing new tools for researchers, planners and engineers, would also provide directions for future research.

Chapter 2: Data

Study Area

The study area is defined by Ottawa Division county limits (Ottawa) as shown on the index map of Figure 1. A small number of sampling locations (0.6%) are located just outside of Ottawa in the neighbouring counties of Lanark (1 location) and Prescott-Russell (4 locations), as well as Gatineau, Quebec (13 locations). Ottawa occupies an area of approximately 2,760 km² and had an estimated population of 943,260 in 2013 (City of Ottawa, 2013).

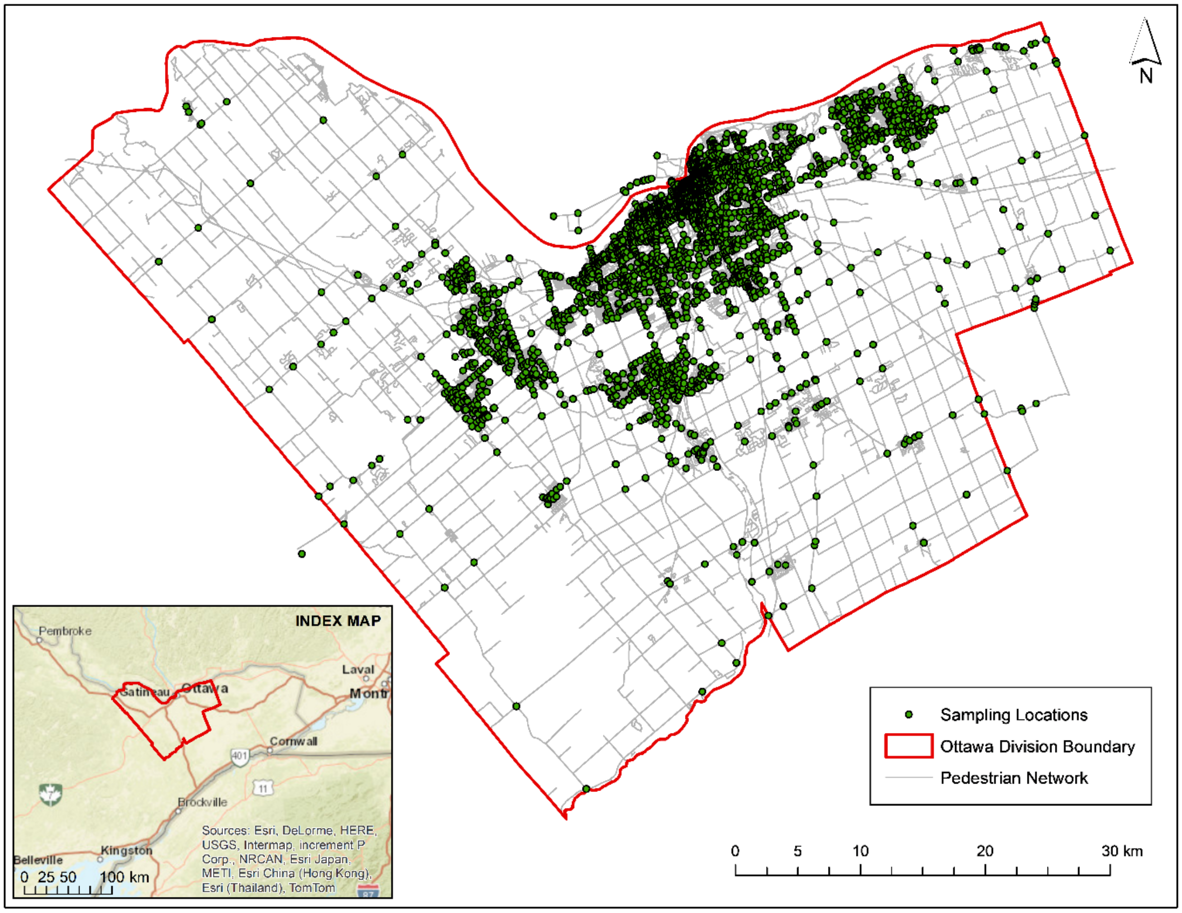


Figure 1: Study Area and Sampling Locations

The spatial dependence relationships for the sampling locations in this study were modeled based on distance, given that pedestrians are fundamentally travelling from an origin to a destination, and that the distance of these trips is typically limited. The spatial distribution scale of the sampling locations was characterized by a mean point density of 1 per km² and a maximum point density (in the downtown core) of 33 per km², with a mean pair-wise Euclidean point separation of 5,500 m within a search radius of 9,000 m[1]. Hence, there was a potentially high degree of connectivity between sampling locations. Since we considered that distance is a key factor in describing the walking spatial relationships, network distances (Euclidean distance constrained to roadways and paths) were considered, rather than the simpler unconstrained Euclidean

[1] Refer to methods section for determination of search radius.

distance, in developing a spatial weights matrix for the MLM modeling. ArcGIS Network Analyst was used to develop a functional pedestrian network, by combining the City of Ottawa Open Data Road Network with the Open Data Pathway Network (City of Ottawa, 2011). Ottawa is characterized by six major barriers to pedestrian traffic (structures or features that provide relatively few crossings at a pedestrian scale): Highways 174, 416 and 417, the Rideau River, the Rideau Canal and the Ottawa River. The pedestrian network accounted for these barriers. Key assumptions also included the presence of sidewalks along all roadways, and that the connectivity of the network is static through time over the course of the study period. During the timespan of pedestrian volume data acquisition (1999-2014), the Corktown footbridge was opened to cross the Rideau Canal at the University of Ottawa in 2006, linking this major institution more directly to major residential and commercial areas in Centretown Ottawa. Another significant footbridge, the Adàwe bridge, was opened over the Rideau River at Donald and Somerset in 2015, however this falls outside of the data acquisition period. The Corktown footbridge is therefore the only known significant violation of a static pedestrian network assumption for the study period. Its construction is likely to have had a sizeable impact on utilitarian and leisure-based pedestrian behaviour in this high pedestrian traffic area. It should be noted that crossings of the canal were possible on foot prior to construction of the bridge during an average of 50 (but highly variable) calendar days each winter, via the Rideau Canal ice. For overall model simplicity and given the relatively limited local effect of the change, the Corktown bridge was included in the network used to develop the spatial weights matrix. The canal ice itself and any other access points to the ice were not considered, but canal locks were included. In developing the pedestrian network, a topology was applied to ensure line segment endpoints were connected within a cluster tolerance of 5 m. Any remaining islands in the network were connected manually. Portions of the resulting network are illustrated in Figure 1.

Dependent Variable: Pedestrian Volume

The secondary pedestrian volume (dependent variable, or DV) data for the City of Ottawa originated from two different data sources, both of which were provided as final data products under license by the City of Ottawa Department of Public Works. Pedestrian volume is defined here as the number of pedestrians counted per unit time (i.e. count per hour) in all directions, in accordance with the MTO definition (MTO, 2010). However, the data do not distinguish between assisted and un-assisted pedestrians, contrary to the MTO definition. The temporal scale of the pedestrian volume data used for analysis is hourly. In cases of sub-hourly data collection, the pedestrian counts were aggregated by location on an hourly basis.

The primary source is sensor-acquired data via subcontractor Miovision (hereafter referred to as "Miovision"), while the secondary source consists of pedestrian counts manually obtained by City employees or contractors using manually operated clickers, turn movement counters, and/or data sheets ("Peds"). The two datasets were combined and treated equally in subsequent analyses, however the Peds data are subject to human error and/or bias given the collection methods used. The City of Ottawa could not provide an estimate of this error rate. The data provided were not geocoded, with location information limited to street intersections or place names represented as part of the file names or as string values within data row entries.

Once mapped to a common schema and collated, the data were geocoded. Addresses were generally created by parsing the file names (Miovision) or street name fields (Peds) and concatenating the information into standard address strings. Google Earth (GE) (Google, 2014) was used to automate geocoding of the majority of addresses given its capability to correct incorrect spellings, locate specific places by name, and adapt to incorrect street type suffixes. Manual geocoding was completed on unlocatable addresses (18.9%) using current (Google, 2014) and historical (City of Ottawa, 2015) aerial photos. The overall spatial accuracy of the geocoding is estimated at 9 metres (m). Due to slight position differences between the sampling locations geocoded with GE and those that were manually geocoded, sampling location geometry was simplified following the merging of the Miovision and Peds datasets using a point distance analysis with a buffer of 15 m, thereby aggregating sampling locations to a final count of 2,999, per Table 1, and calculating the mean pedestrian counts for each final location. The combined Miovision and Peds datasets yield 102,873 hourly observations of pedestrian volume. The timestamps for the pedestrian volume data appear to have been recorded while accounting for daylight savings time (DST), when applicable, based on a comparative bivariate correlation analysis of model variables using adjusted versus unadjusted timestamps. The timestamps are the primary key used to associate level 1 independent variables (IV) to the dependent variable (DV).

Table 1: Pedestrian volume descriptive statistics by data source

Descriptive	Miovision	Peds	Total
N hourly observations	12,471	90,402	102,873
Date Range	Sept. 22, 2010 to Sept. 17, 2014	Jan. 12, 1999 to Aug. 30, 2013	Jan. 12, 1999 to Sept. 17, 2014
Earliest Daily Observation	0:00	06:00	0:00
Latest Daily Observation	24:00	24:00	24:00
Sampling Locations[2]	402	2597	2,999
Hourly Observations per Location	3 - 173, mean 15.69	2 - 296, mean 37.18	2 - 296, mean 34.30

Table 2 summarizes the mean calculation of log-transformed pedestrian counts/volume (LTPC)[3] by sampling location.

Table 2: Summary of LTPC by sampling location

Metric	Value
N	2999
Minimum	2
Maximum	296
Mean number of hourly observations per location	34.3

[2] After geometric simplification

[3] Refer to Figure 2 for further information on pedestrian volume/count data transformation.

Standard Deviation	34.149
Sampling locations with 10 or more hourly observations	62%

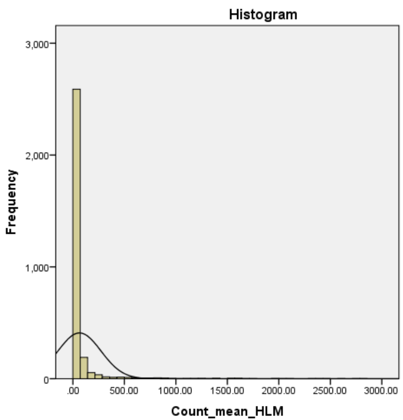
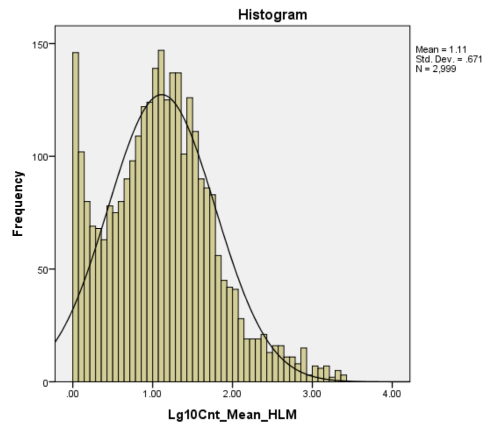
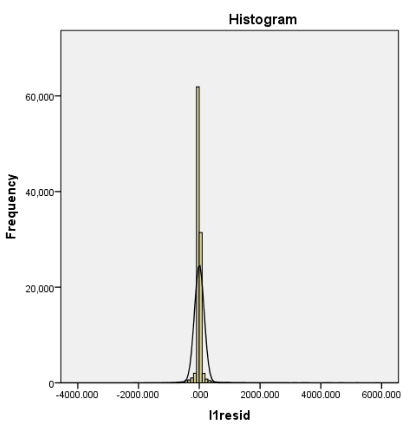
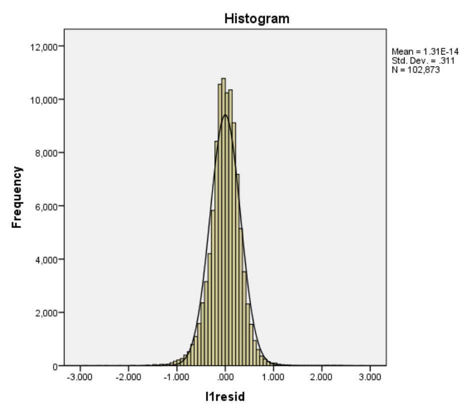
The between-sampling-location variance (level-2 variance) in LTPC calculated in the MLM was highly significant either without controlling for spatial dependency[4] (χ^2=586,424.91, p < 0.001), or with control (χ^2=148,511.18, p < 0.001). The level-2 variance decreased when accounting for spatial dependency because the spatial dependence variable accounted for some of the variance, so there was less variance remaining to be explained. Moreover, the proportion of total variance (ratio of level-2 variance to total variance) in LTPC between sampling locations (as opposed to within sampling locations), i.e., the Intra-Class Correlation (ICC), was 0.817 when not controlling for spatial dependency, or 0.598 when controlling for it. Thus, the majority of the variance was between sampling locations, indicating that HLM is needed to protect against inflated Type I error (indeed, an ICC as low as 0.05 is typically grounds for switching from regular regression to HLM).

As with any MLM, data screening was completed within and between groups, whereby in this case the groups were sampling locations. The aspects of the screening process displayed here are the distributions of the untransformed and transformed DV at the between and within sampling location levels, given that it was the only variable that required transformation, as determined by screening the data. Mean pedestrian volume at level 2 was not normally distributed (refer to Figure 2a), with skew 7.408 and kurtosis 67.239, so pedestrian volume was transformed using $\log_{10}$; (hereafter referred to as 'log-transformed pedestrian counts/volume' or LTPC) the distribution of the transformed variable was then found to be approximately normal (refer to Figure 2b), with skew .469 and kurtosis .175[5], as were its residuals at level 1, with skew -.084 and kurtosis of 3.440 (refer to Figure 2d). 97.6% of sampling locations exhibit variability with the LTPC variable (based on identifying sampling locations with mean of zero). Hence, a model based on the normal distribution function of the transformed DV was favoured over Poisson or negative binomial probability distribution functions typically used to model count data (Hilbe, 2014).

[4] Refer to Chapter 3 - Spatial Autocorrelation for spatial dependence methodology

[5] Given the large sample sizes at levels 1 and 2, standard error for skew and kurtosis were not reported

Figure 2: HLM DV Distributions and Transformations

	Untransformed	Log-Transformed
Mean (Level 2 – between sampling locations)		
	Figure 2a: Histogram, Level 2 Mean pedestrian volume counts	*Figure 2b: Histogram, Level 2 Mean LTPC*
	N = 2,999 Range = 0 to 2,790 Mean = 63.8379 Std Dev =208.63324 Skew = 7.408 Kurtosis = 67.239	N = 2,999 Range = 0 to 3.41 Mean = 1.1134 Std Dev =.67099 Skew = .469 Kurtosis =.175
Level 1 Residual (within sampling locations)		
	Figure 2c: Histogram, Level 1Pedestrian volume count residuals	*Figure 2d: Histogram, LTPC Level 1 Residuals*
	N = 102,873 Range = -2,129.452 to 5,857.684 Mean = 0.0000 Std Dev =158.015586 Skew = 6.728 Kurtosis = 154.036	N = 102,873 Range = -2.882 to 2.793 Mean = 0.0000 Std Dev =.311258 Skew = -.084 Kurtosis =3.440

Predictors and Moderators of Walking

Independent Variables

The descriptive level 1 statistics for all independent variables (IV) in this research are provided in Table 4 (the dependent variable, or DV, is also included). Statistics are given for the Level 1 data set as a whole, without taking nesting (within sampling locations) into account.

Walkability

WS (Walk Score Advisory Group, 2017) values were obtained for each of the 2,999 LTPC sampling locations indicated in Figure 1. The frequency distribution of the WS samples is provided in Figure 4. Skewness and kurtosis of the WS values were -0.456 and -0.724, respectively, indicating normality. Per Figure 3, the full range of WS was reasonably sampled, with some bias towards higher WS values.

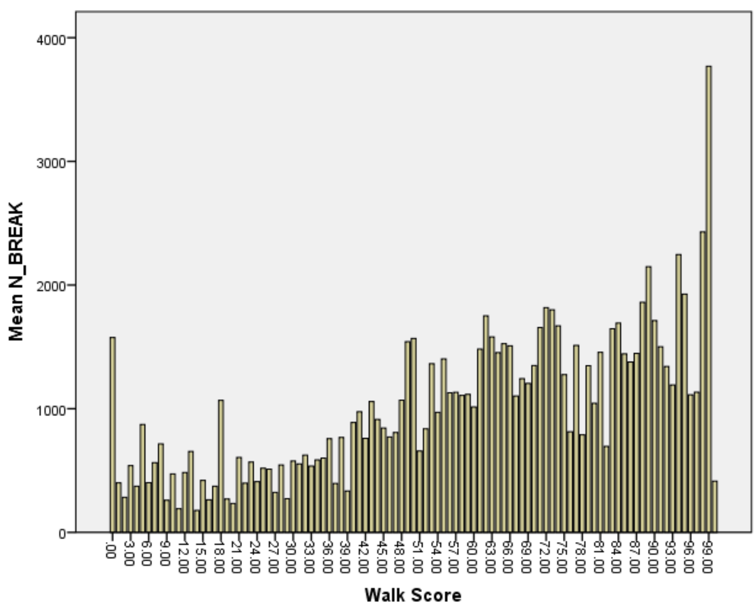


Figure 3: Number of LTPC samples (N_Break) per WS Value

Weather

Hourly weather data from 1999 to 2015 were purchased from Environment and Climate Change Canada (ECCC). They consist of hourly weather observations assembled from three Ottawa weather stations[6] to produce the most complete dataset possible for all variables. Sunshine and solar radiation datasets were found to be largely incomplete, with data entries ending in 2006-2007, and so these were discarded. ‘No data’ values and elements with data flags indicating poor integrity were treated equally as missing values. In cases where multiple hourly measurements were provided (up to 3), the mean of the measurements was taken. Measurements of total precipitation and hourly rainfall were consolidated. Total hourly precipitation was used when both hourly rainfall and precipitation were available, and rainfall was used when precipitation data was missing. Weather variables in the ECCC dataset that were sufficiently complete, and thereby used in subsequent model screening and analyses, are summarized in Table 4. Since data

[6] Station IDs and percent of data from each: 6105976/6105978 - Ottawa Central Experimental Farm (43.7%), 6106000/6106001 - Ottawa MacDonald-Cartier International Airport (56.1%), 7032682 - Gatineau Airport (0.2%)

was recorded in local standard time (LST), all timestamps were adjusted to DST where applicable. Sunset and sunrise times for each timestamp were calculated based on the NOAA solar calculator (National Oceanic and Atmospheric Administration, 2017).

The AAT (Steadman, 1994) is given by Equation 1, while the perceived temperature was calculated based on Humidex and Windchill[7] (Environment and Climate Change Canada, 2017)..

Equation 1: Australian Apparent Temperature, excluding effects of radiation

$$AAT = T_{Dry-Bulb} + 0.33 * \left[\frac{RH}{100} * 6.105 * e^{\left(\frac{17.27 \times T_{Dry-Bulb}}{(237.7 + T_{Dry-Bulb})}\right)}\right] - 0.70 \times \left(\frac{V}{3.6}\right) - 4.00$$

Where $T_{Dry-Bulb}$ (dry-bulb temperature) is in degrees Celsius (°C), V (wind speed) is in kilometers per hour (km/h) and RH (relative humidity) is given in %.

The temperature variables (dry bulb, wet bulb, dew point, perceived, AAT) all exhibited a high kurtosis, which may be indicative of outliers (specifically, potentially incorrect values not flagged by ECCC). Their maximum Z-scores were in the range of -5.12 (AAT) to -6.54 (Wet Bulb Temperature). A Z-score of -5.62 was calculated for Dry Bulb Temperature. The probability of getting a Z-score of +/- 5.12 is 1.528e-07. In a sample size of 102,873, .015 observations would exceed this z-score by chance. Therefore, even observations at Z-scores of 5.12 did not need to be treated as outliers with a sample size this large. Despite the high kurtosis, the distributions of temperature were considered normal and no transformations were applied. Precipitation and snow depth were dichotomized due to their non-normality (94.6% zero precipitation and 90.0% zero snow depth observations).

The snow depth variable was dichotomized given its high skew and kurtosis (4.503 and 21.18, respectively). The dichotomized variable was comprised of a large proportion of missing data (N = 52,248 out of 102,873 LTPC observations), and most valid observations (90%) represented no snow cover, which limited its use in subsequent models.

Season

The moderating effect of seasonality on the relationship between weather variables and walking (Aultman-Hall, Lane, & Lambert, 2009) was accounted for by a dichotomous variable representing 'winter' and 'non-winter' conditions based on calendar dates halfway between winter and summer solstices. Hence, the 'winter' season was defined as November 6 to May 6. As indicated in Table 4, most of the pedestrian volume data (88.4%) were collected in the 'non-winter' season, which indicated data collection bias (i.e. data preferentially collected during favourable weather conditions). A second definition of winter was examined for screening purposes using calendar months with minimum average dry bulb temperatures below zero

[7] The maximum dry bulb temperature at which perceived temperature can reach 0 °C under 100 km/h winds (maximum wind speed shown on wind chill charts) is approximately 7.3 °C. This was used as a cutoff for comparing wind chill temperature with dry bulb temperature, and choosing the lower of the two as the perceived temperature. With windspeeds of zero, the Windchill may be calculated to be greater than the dry bulb temperature. In those cases, the minimum of the two was taken as the perceived temperature.
If the dry bulb temperature is 21°C or greater (the temperature at which, with maximum humidity, discomfort begins to be felt), the greater of the Humidex or the dry bulb temperature was selected as the perceived temperature. Dry bulb temperatures within the range of 7.3 °C and 21°C were assigned the dry bulb temperature as the perceived temperature.

degrees Celsius, which was considered likely to correspond with the presence of snow on the ground. In Ottawa, that period was from November to March inclusive. In the latter definition of season, 93.1% of the pedestrian volume data was collected in the non-winter season, representing further loss of data in the 'winter' category, hence it was not retained for further analysis. The snow cover variable could be employed as a proxy for season, however that variable had fewer valid observations (N=52,248) and had to be dichotomized due to non-normality, rendering its use potentially problematic.

Time of day/week

As a proxy of pedestrian intent based on the association of walking purpose with week day and time of week, three time categories were developed based on the observed variations in LTPC within the dataset, illustrated in Figure 4 (Figure 4a through Figure 4w). Table 3 provides a breakdown of sample size by time category, as well as a more granular breakdown by weekday and time of day intervals drawn upon the observed variations in mean LTPC, and bivariate correlations of LTPC by weekday and day hour with WS. The final time categories based on this exploratory analysis aligned with the trends in walking trip frequency, by hour of day and for employed Canadians, from Statistics Canada's General Social Survey of 2010 (Michelson & Lachapelle, 2016).

Table 3: LTPC sample size by Time Category

Final Time Category	N	Initial Time Category	N
Work and errands	123,541	Saturday, 07:00 to 12:00	551
		Weekdays, 07:00-11:00 and 12:00 to 18:00	121,289
		Weekdays, 11:00 to 12:00	653
Leisure	16,122	Friday, 18:00 to 23:00	2,283
		Saturday and Sunday, 22:00 to 23:00	372
		Saturday, 12:00 to 22:00	10,232
		Sunday, 07:00 to 22:00	2,271
		Weekdays, 18:00 to 21:00	1,416
		Weekdays, 21:00 to 22:00	285
Night	2,114	Early morning, 05:00 to 07:00	156
		Friday, 23:00 to midnight	446
		Night, 00:00 to 05:00	1,185
		Saturday and Sunday, 23:00 to midnight	364
		Weekdays, 22:00 to midnight	196
Insufficient Data			98

The time categories subsequently used in the analyses were defined as follows:

Work and errands:	Time periods where pedestrian activity will reflect patterns of utilitarian / required travel (also sometimes referred to as destination walking, although this could be misconstrued as leisure). Includes measurements occurring in whole or in part within the period from 07:00 to 18:00, Monday to Friday, and from 07:00 to 12:00 on Saturdays.
Leisure:	Time periods where pedestrian activity will reflect patterns of voluntary travel. Includes measurements occurring in whole between 18:00 and 21:00, Monday to Thursday; between 18:00 and 23:00 on Fridays, between 12:00 and 23:00 on Saturdays, and between 07:00 and 23:00 on Sundays.
Night:	Time periods where pedestrian activity is low or nil due to night-time conditions, but likely related to leisure activities of a sub-population (young adults), if present. Measurements occurring in whole or in part between 21:00 and 07:00, Monday to Thursday; and between 23:00 and 07:00, Friday, Saturday and Sunday.

As defined above, 35.7% of time-of-day-and-week hours were categorized as "Work and Errands", while 64.3% were categorized as "Leisure" and "Night". The majority of sampling locations (73.6%) included only LTPC data in the "Work and Errands" time category, which, in consideration of its definition, was an indication of data collection bias (i.e. data mostly collected for traffic studies during rush hour periods). Since it was expected that interactions may occur between the time category and weather variables (e.g. people are less likely to walk in the rain during their leisure time than they are for utilitarian trips), the time category variable was nonetheless maintained for further analysis given its theoretical importance. The time category variable was dichotomized and added to level 1 to evaluate the moderating effect of the "Work and Errands" (WE) and the combined "Leisure" and "Night"[8] (LN) time categories on the relationship between LTPC and WS and other level 1 variables.

Table 4: Descriptive bulk statistics for all study variables at levels 1 and 2

HLM Level	Variable	N[9]	Min.	Max.	Mean	Std. Dev.	Skew	Kurtosis	Percentages[10]
1	Pedestrian volume (hourly count)	102,873	0	8626	107.34	325.20	6.99	70.76	NA
1	LTPC	102,873	0	3.94	1.29	0.797	0.237	-0.357	NA
1	Time Category	102,873 (95,345 + 6,948 + 580)	1	3					92.7 % WE 6.8 % Leisure 0.6 % Night
1	Time Category Dichotomous	102,873 (7,528 + 95,345)							7.3 % LN (0) 92.7 % WE (1)

[8] Given the small hourly sample size of the "Night" time category, these data were combined with "Leisure".
[9] Counts in parentheses match the sequence of categories in the percentages column.
[10] Categorical and dichotomous variables.

HLM Level	Variable	N[9]	Min.	Max.	Mean	Std. Dev.	Skew	Kurtosis	Percentages[10]
1	Dry-Bulb Temperature (°C)	102,838	-29.3	36.6	18.14	8.58	-1.416	3.062	
1	Wet-Bulb Temperature (°C)	97,925	-29.1	28.15	14.24	6.69	-1.556	3.722	
1	Dew point temperature (°C)	102,860	-33.27	26.75	10.98	8.157	-1.480	2.971	
1	Perceived Temperature (°C)	102,812	-40.18	48.39	19.64	11.463	-1.051	2.141	
1	Australian Apparent Temperature (°C)	102,860	-35.46	38.34	16.47	10.225	-1.310	2.355	
1	Precipitation (hourly total mm)	98,855	0	361.9	0.110	4.256	133.78	23516.8	
1	Precipitation dichotomous	98,855 (93,560 + 5,295)							94.6 % no precipitation (0) 5.4 % precipitation (1)
1	Total Cloud Cover (%)	89,515	0	100	62.74	34.64	-0.536	-1.129	
1	Daylight (dichotomous)	102,873 (2,001 +100,872)							1.9 % dark (0) 98.1 % light (1)
1	Atmospheric pressure (kPa)	102,860	96.74	102.89	100.21	0.659	-0.070	0.574	
1	Wind speed (km/h)	102,860	0	46.67	12.66	6.39	0.693	0.422	
1	Relative Humidity (%)	102,860	15.5	100	64.17	19.39	-0.105	-0.973	
1	Weather Indicator	102,860 (87,787 + 15,073)							85.3 % normal weather (0) 14.7 % severe weather (1)
1	Season (calendar)	102,873 (11,943 +90,930)							11.6 % winter (0) 88.4 % non-winter (1)
1	Season (temperature)	102,873 (7,118 +95,755)							6.9 % winter (0) 93.1 % non-winter (1)

HLM Level	Variable	N[9]	Min.	Max.	Mean	Std. Dev.	Skew	Kurtosis	Percentages[10]
1	Sum of Intensity (0-6 index)[11]	57,952	0	6	0.347	0.683	2.429	7.593	
1	Snow Depth (cm)[12]	52,248	-2	100	1.93	7.72	4.503	21.18	
	Snow Depth dichotomous	52,248 (47,040 + 5,208)							90.0 % no snow cover (0) 10.0 % snow cover (1)
2	WS (0-100 index)	2,999	0	100	57.90	27.28	-0.456	-0.724	
2	LTPC Spatial Lag (Wy or LTPC-lag)[13]	2,999	0	3.3976	1.1756	0.5464	0.437	0.609	
2	Walk Score Spatial Lag (Wx or WS-lag)[13]	2,999	0	98.999	60.819	23.170	-0.545	-0.267	

[11] Sum of intensities for: Thunderstorms, Rain, Rain Showers, Drizzle, Freezing Rain, Freezing Drizzle, Snow, Snow Grains, Ice Pellets, Ice Pellet Showers, Snow Showers, Snow Pellets, and Hail. Each variable has a scale from scale of 0 to 3 (except thunderstorm 2-3), yielding a total intensity range of 0 to 6

[12] Variable measured starting in 2004. From 2004, inconsistent measurements from year to year, with significant portion of missing data.

[13] Refer to methodology section for spatial lag variable development

Figure 4: Independent Variable Histograms

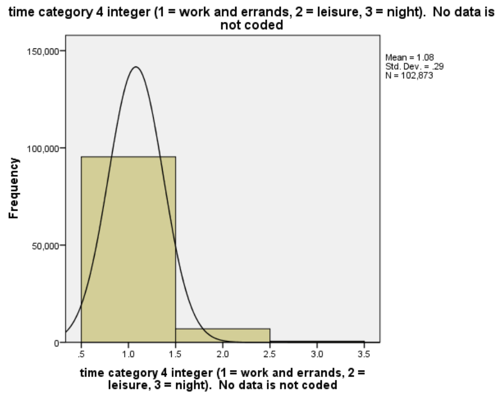
<table>
<tr><th>HLM Level</th><th colspan="2">Histograms</th></tr>
<tr><td>1</td><td>


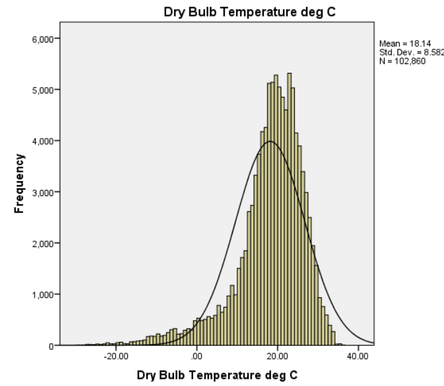
Figure 4a: Time Category histogram</td><td>


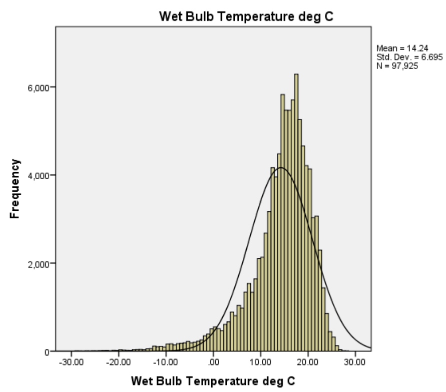
Figure 4b: Dry-Bulb Temperature (°C) histogram</td></tr>
<tr><td>1</td><td>


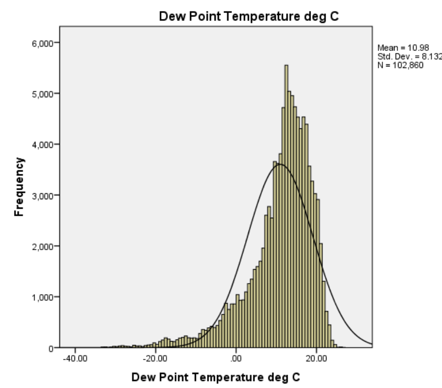
Figure 4c: Wet-Bulb Temperature (°C) histogram</td><td>


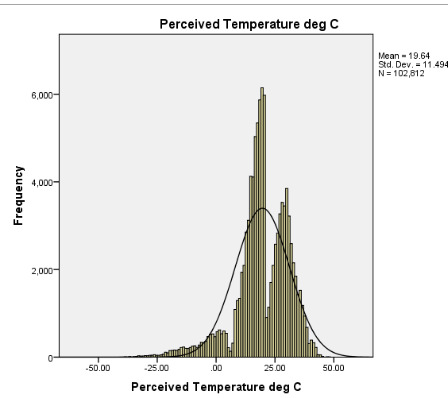
Figure 4d: Dew point temperature (°C) histogram</td></tr>
<tr><td>1</td><td>


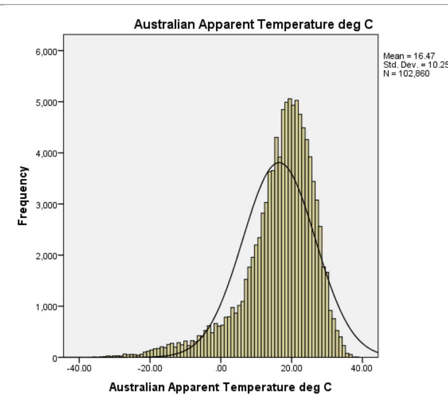
Figure 4e: Perceived Temperature (°C) histogram</td><td>


Figure 4f: Australian Apparent Temperature (°C) histogram</td></tr>
</table>

HLM Level	Histograms

1

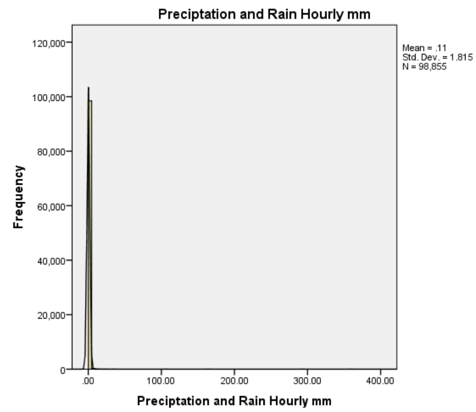


Figure 4g: Precipitation (hourly total mm) histogram

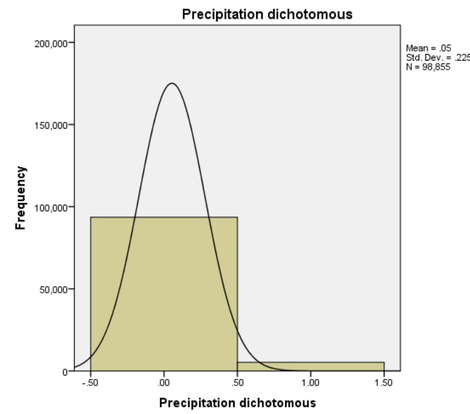


Figure 4h: Precipitation dichotomous histogram

1

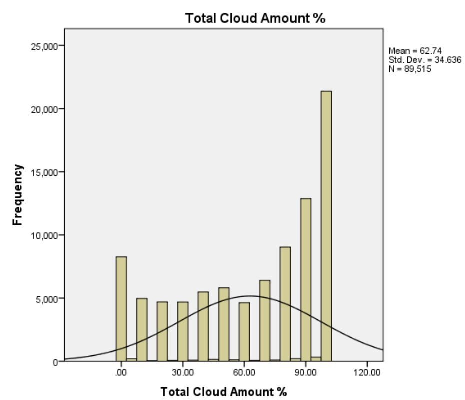


Figure 4i: Total Cloud Cover (%) histogram

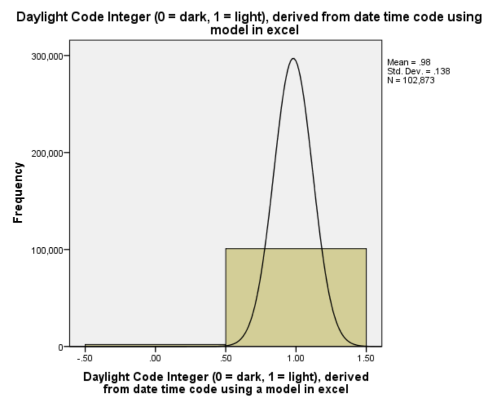


Figure 4j: Daylight (dichotomous) histogram

1

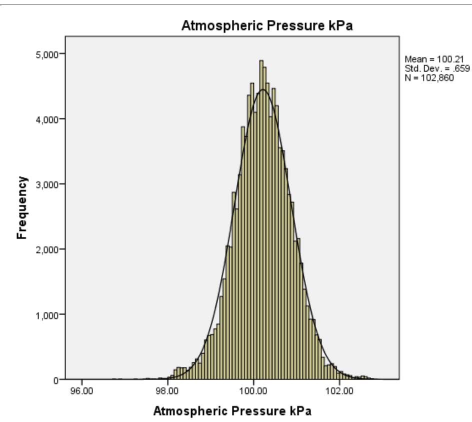


Figure 4k: Atmospheric pressure (kPa) histogram

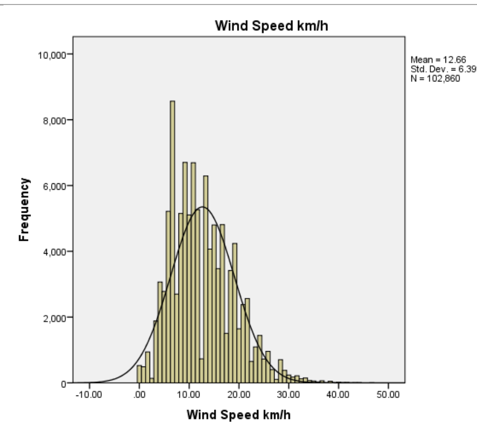


Figure 4l: Wind speed (km/h) histogram

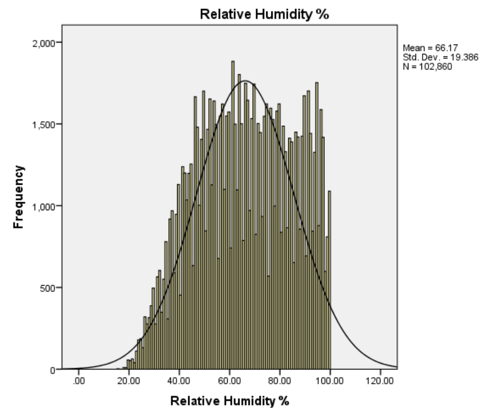
<table>
<tr><th>HLM Level</th><th>Histograms</th><th></th></tr>
<tr><td>1</td><td>

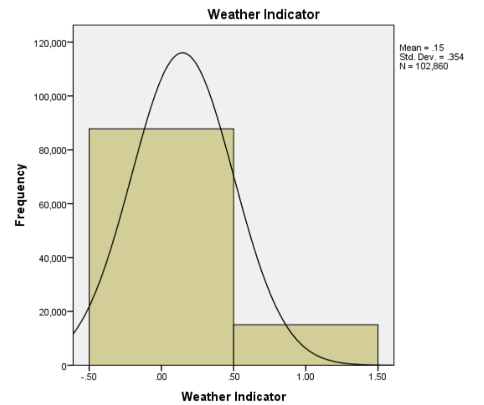
Figure 4m: Relative Humidity (%) histogram</td><td>

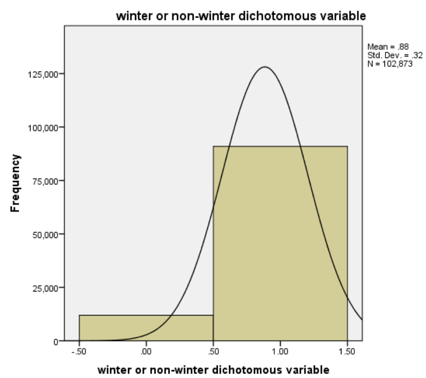
Figure 4n: Weather Indicator histogram[14]</td></tr>
<tr><td>1</td><td>

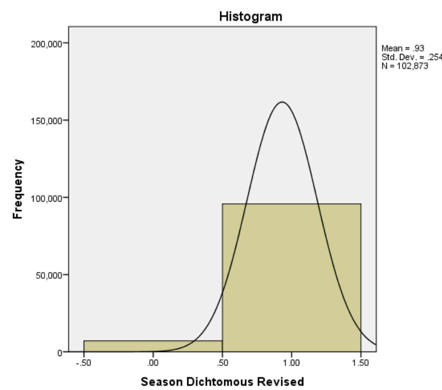
Figure 4o: Season (calendar) histogram</td><td>

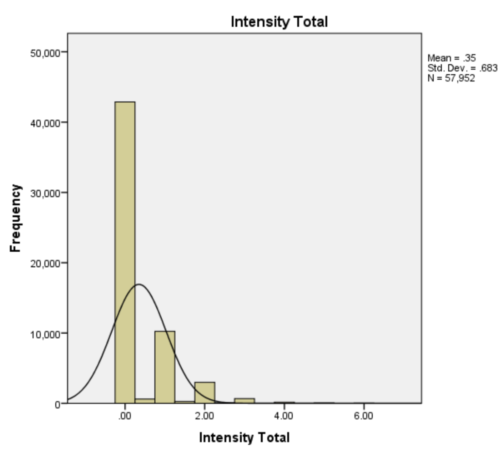
Figure 4p: Season (temperature) histogram</td></tr>
<tr><td>1</td><td>

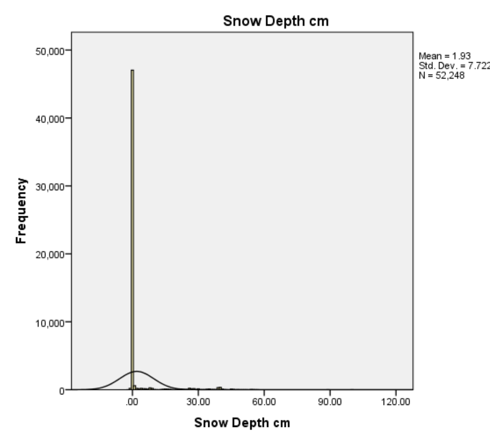
Figure 4q: Sum of Intensity (0-6 index) histogram</td><td>

Figure 4r: Snow Depth (cm) histogram</td></tr>
</table>

[14] The dichotomous weather indicator variable indicates presence or absence of severe weather per ECCC's definition (Environment and Climate Change Canada, 2015)

HLM Level	Histograms

1

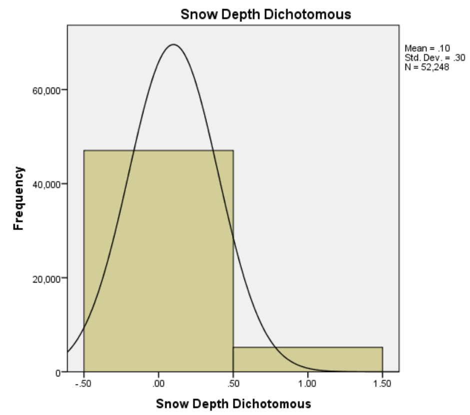


Figure 4s: Snow Depth dichotomous histogram

2

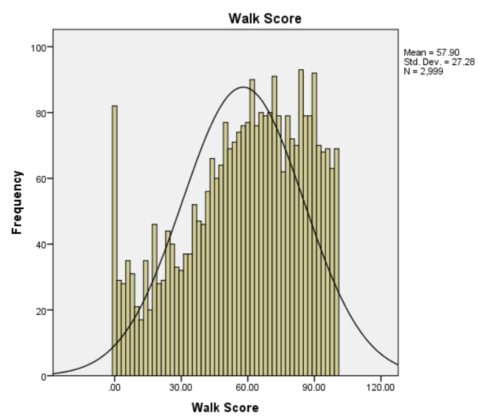


Figure 4t: WS (0-100 index) histogram

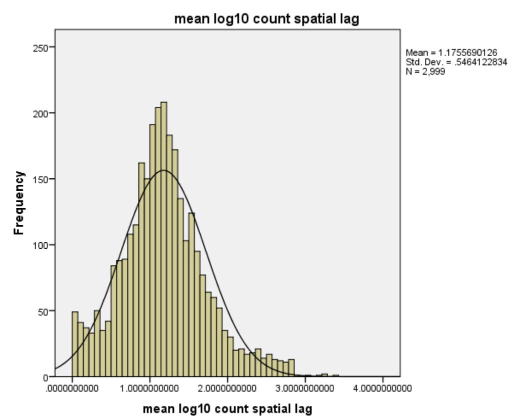


Figure 4u: LTPC-lag (Wy) histogram

2

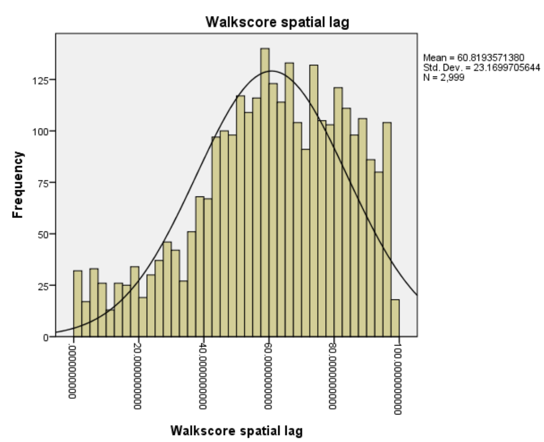


Figure 4v: Walk Score Spatial Lag (Wx) histogram

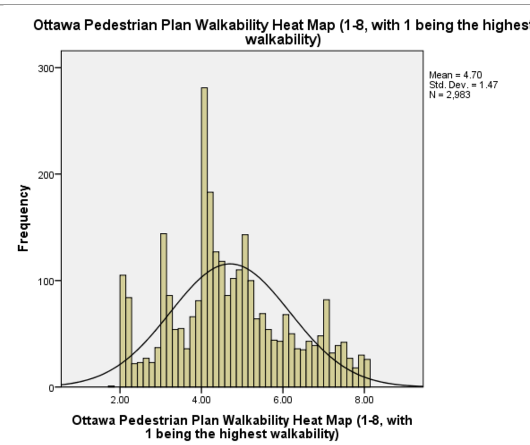


Figure 4w: Ottawa Pedestrian Plan (2013) Walkability Heat Map (1-8 index) histogram

Figure 5: Rose Histograms of Mean LTPC by week day and time of day

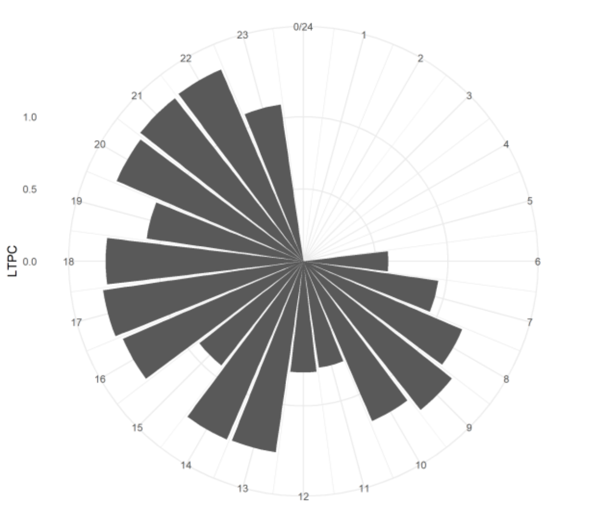


*Figure 5*a: LTPC Mondays[15]

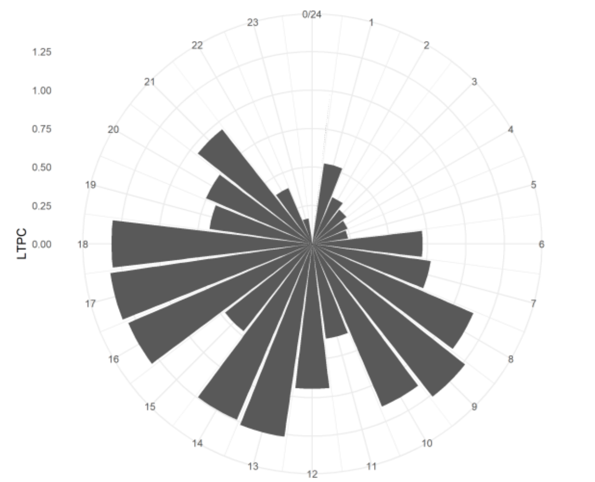


*Figure 5*b*: LTPC Tuesdays*[16]

[15] Mondays: Insufficient data (N<10) for times 01:00 to 07:00, and for 22:00 to 24:00

[16] Tuesdays: Insufficient data (N<10) for times 01:00 to 05:00, and for 22:00 to 24:00

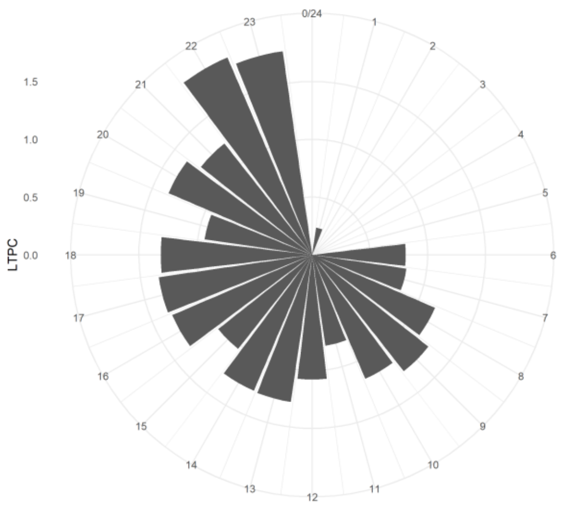


*Figure 5*c: LTPC Wednesdays[17]

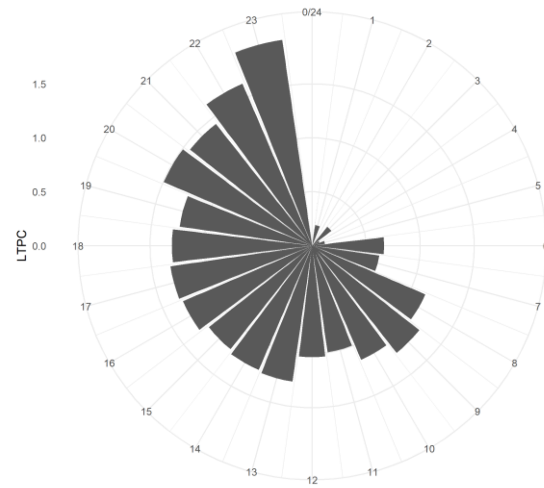


*Figure 5*d*: LTPC Thursdays*[18]

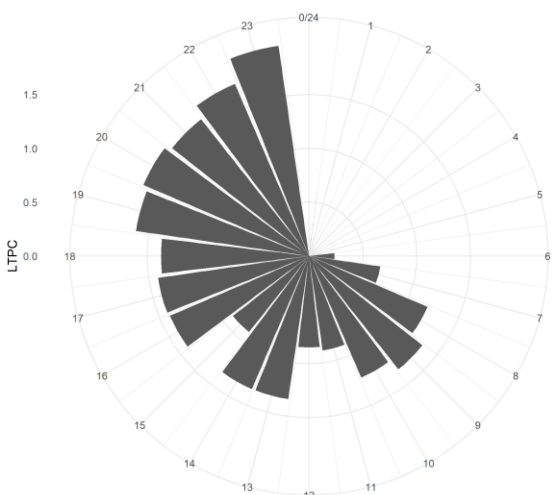


*Figure 5*e: LTPC Fridays[19]

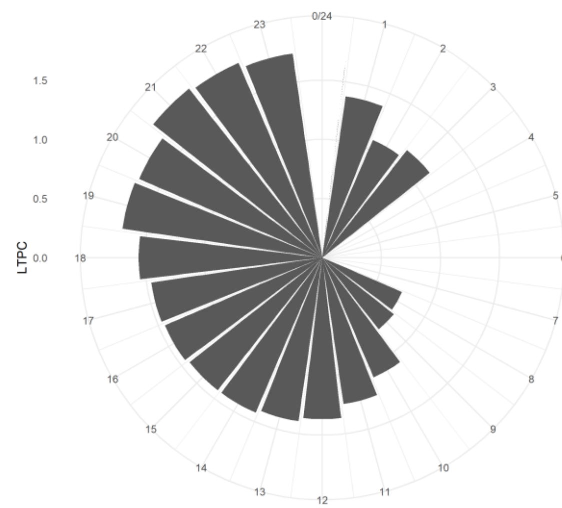


*Figure 5*f*: LTPC Saturdays*[20]

[17] Wednesdays: Insufficient data (N<10) for times 00:00 to 05:00, and for 23:00 to 24:00
[18] Thursdays: Insufficient data (N<10) for times 01:00 to 06:00, and for 24:00
[19] Fridays: Insufficient data (N<10) for times 01:00 to 07:00, and for 24:00
[20] Saturdays: Insufficient data (N<10) for times 04:00 to 07:00

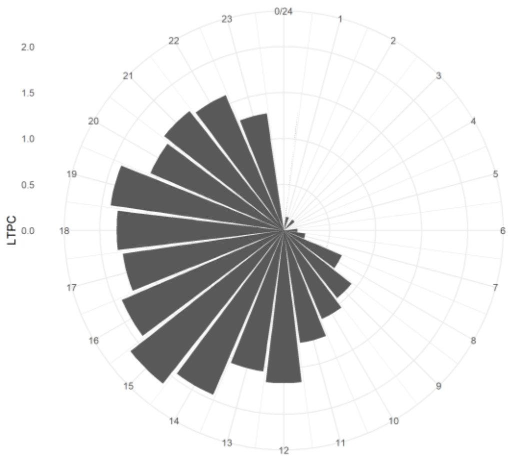


*Figure 5*g: LTPC Sundays[21]

Variability of Dichotomous Independent Variables

As indicated in Table 5, there was a significant lack of variability for the main dichotomous variables at the sampling location level (level 2), which forced the MLM to perform Bayesian slope estimation at several sampling locations, thus reducing model reliability. The season variable, for example, had only 821 intersections with observations in both seasons (defined by calendar day), which represented a significant loss of data. Time category was similarly problematic, with only 788 intersections having observations in both time categories.

Table 5: Percentage of sampling locations with any variability for dichotomous variables

Dichotomous Variable	Percent Sampling Locations with variability
Weather Indicator	65.2%
Precipitation	49.6%
Season (Calendar)	27.4%
Season (Temp)	18.0%[22]
Time category	26.3%
Snow Depth	22.5%
Daylight	19.3%

[21] Sundays: Insufficient data (N<10) for times 01:00 to 08:00
[22] Given this low proportion of intersections with variability, the calendar definition of winter/non-winter was used moving forward

Chapter 3: Methods

Approach

Relevant model independent variables (IVs) were iteratively screened and selected based on data completeness, bivariate correlations and possible interactions within the bulk data. A subset of theoretically relevant variables having substantial association with log-transformed pedestrian volume/counts (LTPC) was then selected for inclusion in a series of MLMs, considering the secondary nature of the data and associated risk of data loss when nesting. Relevant MLMs were retained for discussion based on explanatory variable significance to the model and overall data sufficiency.

Software

HLM 7's (a popular MLM software) spatial modeling feature could not be used for this study due to the large size of the dataset. Therefore, a hybrid approach was used whereby the two-level MLMs were solved using HLM 7 (Raudenbush, Bryk, Cheong, Congdon, Jr., & du Toit, 2011), while incorporating a spatial lag term derived in GeoDaSpace (Anselin & Rey, 2014; Anselin, Syabri, & Kho, 2006) in lieu of the HLM7's spatial dependence term, and using a spatial dependency matrix generated with ArcGIS 10.5.1 Network Analyst. Descriptive statistics and data transformations were completed with SPSS v24 and R v3.4.3.

Spatial Autocorrelation

To avoid misspecification of multi-level regression models due to spatial autocorrelation (resulting in an incorrect association between DV and IVs), a spatial auto-regressive (SAR) approach was applied at level 2. Spatial lag terms based on the mean of the DV (LTPC)[23] and the level-2 IV (WS) at each of the 2,999 sampling locations were included as explanatory variables at level 2, similar to existing spatial MLM approaches (Park & Kim, 2014; Chen & Wen, 2010; Morenoff, Sampson, & Raudenbush, 2001; Morenoff J. D., 2003).

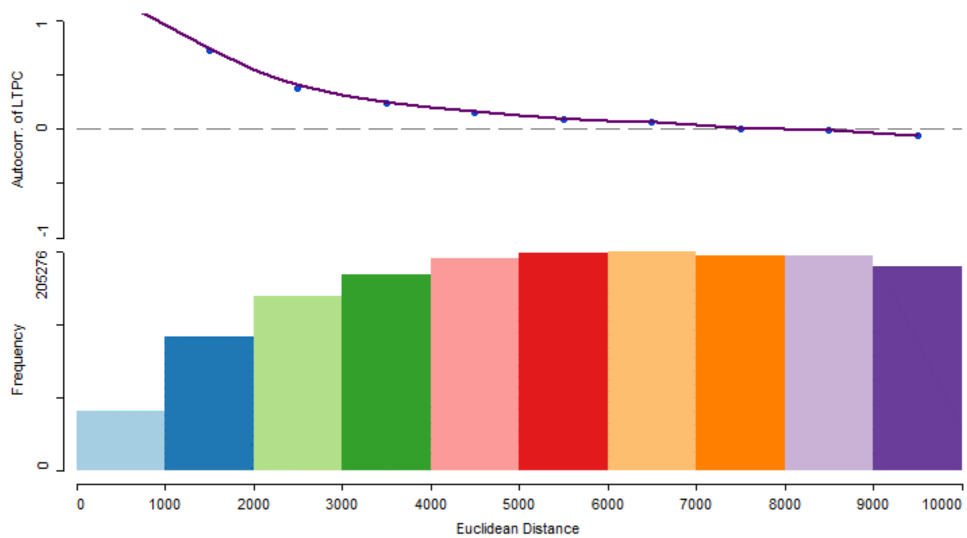
Global Moran's I for the mean DV and level 2 IV at each of the 2,999 sampling locations in the dataset (locations per Figure 1) were summarized in Table 6. The tests indicated that spatial autocorrelation was likely a significant factor for both the DV and level 2 IV.

An attempt was also made to eliminate the spatial autocorrelation by randomly sub-sampling the locations with a minimum Euclidean spacing of 2, 3 or 4km, resulting in at least 80 locations in the sub-sample (the rule-of-thumb for a 2-level MLM is 20 groups per IV). In all sub-sampled cases, the Global Moran's I[24] still indicated significant spatial autocorrelation with the mean LTPC outcome variable, therefore the concept of analyzing random sub-samples with minimum spacing was dropped

[23] This mirrors the operation in HLM, rather than calculating log10 of the mean pedestrian count.

Table 6: Global Moran's I of variables of interest

Variable	Moran's I[24]	z-score	p-value	Conclusion
Mean LTPC by sampling location	0.6434	54.8879	0.0000	Significantly clustered pattern
WS	0.7871	67.1325	0.0000	Significantly clustered pattern



Autocorr.	1.193	0.726	0.369	0.240	0.149	0.092	0.062	0.003	-0.016	-0.061
Min	0	1000	2000	3000	4000	5000	6000	7000	8000	9000
Max	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000
# Pairs	55737	125113	162305	182830	198788	203201	205276	201028	200737	190235

min: 0, max: 10000.000, total # pairs: 1.72525e+006, Autocorr. = 0 at ~7663.682 in range: [7000.000, 8000.000]

Figure 6: Non-parametric spatial correlogram of LTPC, based on Euclidean distance[25]

Figure 6 illustrates that beyond approximately 7,664 m, LTPC were no longer spatially auto-correlated. Furthermore, based on the travel survey results from 2005 and 2011 summarized in the Draft Ottawa Pedestrian Plan (OPP) (City of Ottawa, 2013), walking trip distances in Ottawa were rarely greater than 10,000 m. This corroborated with the non-parametric spatial correlogram of the LTPC data in Figure 6[26]. The survey results also supported, per Figure 7 and Figure 8, a conceptual model of spatial relationships for our sampling locations with an impedance, or distance decay, pattern. Based on the correlogram in Figure 6, and Figure 7, an impedance cutoff of 9,000 m was selected, which also allowed inclusion of at least one neighbour for every sampling location, since the greatest distance between two sampling locations in the dataset was 8,940 m. As the majority of walking trip distances were less than 1,000 m (median of 720 m) and declined with increasing distance, per Figure 7 and Figure 8, an

[24] Derived using a Network Spatial Weight Matrix with inverse distance exponent of 2.1.
[25] The colour coding has no significance and cannot be adjusted in the software.
[26] The spatial autocorrelations of LTPC were represented by the average correlations for all pairs of observations in a given distance bin (Bjornstad & Falck, 2001; Anselin, Syabri, & Kho, 2006)

inverse weighting scheme was assumed within the 9,000 metre impedance cutoff. An inverse exponent of 1.574 was surmised based on an exponential decay model relationship between the proportion of trips and trip distance indicated in the OPP in Figure 8. The above assumed that regional-scale spatial regimes[27] were absent, and subsequent analyses also assumed a continuous spatial relationship within a 9,000 m radius of each sampling location.

Since we know that walking distance was a critical factor in the degree of spatial autocorrelation of the pedestrian volume measurements across the entire spatial extent of this study, global standardization was considered most relevant to the problem. However, GeoDaSpace cannot handle global standardized spatial weights in its spatial lag engine using a network spatial weights matrix (NSWM), therefore the weights were row standardized.

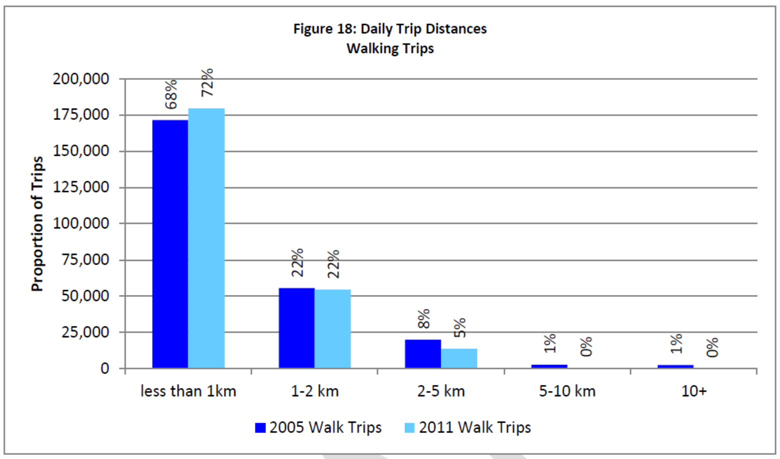


Figure 7: OPP Daily Trip Distances for Walking Trips

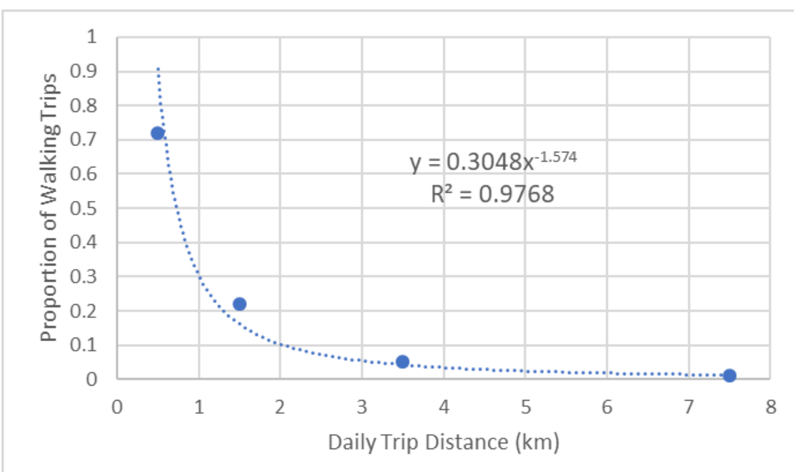


Figure 8: Proportion of Walking Trips vs Daily Trip Distance, Best Fit Equation and R-squared

An Ordinary Least Squares regression (OLS) was completed as an initial exploratory spatial autocorrelation analysis, comparing LTPC and WS. The OLS residuals were heteroscedastic, and spatially autocorrelated to a significant degree (using the NSWM with inverse exponent 2.1), per the Global Moran's I (MI) and Lagrange Multiplier (LM) metrics summarized in Table 7, and as

[27] Spatial regimes are defined here as structural differences across a geography, whereby sub-regions may be characterized by a statistically different population or pattern.

illustrated by the non-linear pattern on the right-hand side in Figure 9, thus a spatial lag model was considered, rather than a spatial error model (Anselin & Rey, 2014). Furthermore, a spatial lag model was conceptually appropriate given the self-explanatory diffusion process occurring whereby the value of LTPC at a given location was related to nearby LTPC, since it was assumed that people walk between multiple sampling locations within each hour. An attempt was made to account for spatial dependence using an eigenvector spatial filtering (ESF) approach (Griffith & Chun, 2014), with the R spdep package, however the processing time given the size of the dataset was found through empirical testing to likely be excessive (on the order of 10^{13} years). Reasonable processing times (hours) could likely be achieved with the spdep approach with sample sizes of approximately 600 or less.

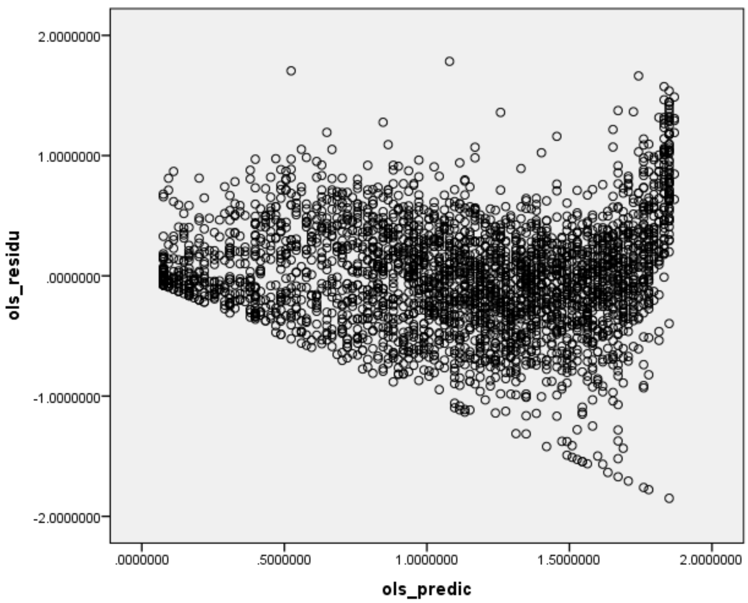


Figure 9: OLS Residuals – mean LTPC vs WS

Table 7: OLS Diagnostics

Parameter/Test	Test result (p-value)[28]	Conclusion
R-squared/Adjusted R-squared	0.5310/0.5308	
Log likelihood	-1923.045	n/a
Akaike info criterion (AIC)	3850.091	n/a
Schwarz criterion	3862.103	n/a
WS (IV) coefficient	0.0179230	n/a
WS (IV) t-statistic	58.2476 (0.0000*)	Linear relationship was statistically significant

[28] * statistically significant

Jarque-Bera	108.095 (0.0000*)	Errors were non-normal
Breusch-Pagan test	248.697 (0.0000*)	Heteroskedasticity present, BP and KB were different = errors were non-normal
Koenker-Bassett test	169.813 (0.0000*)	
White	173.945 (0.0000*)	Heteroskedasticity present
Moran's I	28.194 (0.0000*)	Significant spatial clustering/auto-correlation of the DV
Lagrange Multiplier (LM) (lag)	859.124 (0.0000*)	Spatial autocorrelation present. All LM Diagnostics and Robust LM Diagnostics significant, indicating relevance of Spatial Lag Model
Robust LM (lag)	78.608 (0.0000*)	
LM (error)	790.369 (0.0000*)	
Robust LM (error)	9.853 (0.0000*)	
LM (SARMA)	868.978 (0.0000*)	

A column vector of spatial lags was generated for the LTPC and WS variables using GeoDaSpace (Anselin & Rey, 2014) and the NSWM described previously. To select a single spatial lag vector, several NSWM were created and tested using a range of exponent values (1, 1.5, 1.75, 1.9, 2.0 and 2.1 – only 2.0 and 2.1 are shown). The NSWM producing 1^{st}-order Spatial Two Stage Least Squares (S2SLS) using Generalized Method of Moments (GMM) residuals for Equation 2 (DV = LTPC, IV = Walk Score®) that were no longer spatially autocorrelated per the Anselin-Kelejian Test (AKT), were retained for multi-level modeling. Characteristics of the selected NSWM are summarized in Table 9. The results of the exploratory spatial modeling analysis are summarized in Table 8. As indicated in Table 8, a NSWM created using an inverse exponent of 2.1 yielded a spatial lag model with non-spatially autocorrelated residuals (as illustrated by the more linear grouping of residuals and pattern on the RHS of Figure 10, compared to Figure 9), and its LTPC-lag vector Wy was thus selected for inclusion in the MLM at level 2. Although the bivariate S2SLS model included a Wx spatial lag *instrumental* variable for WS, it was included in the MLM in HLM 7 as an *explanatory* variable (WS-lag).

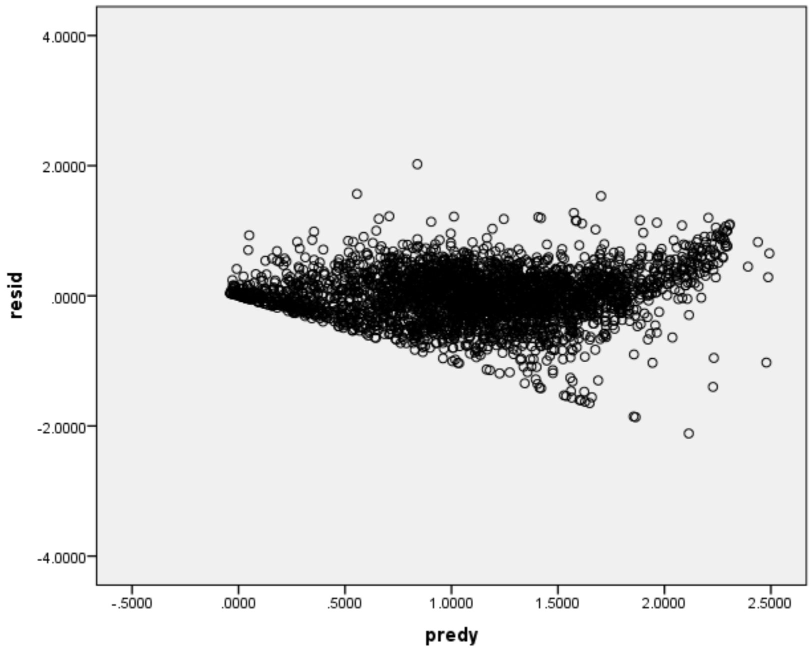


Figure 10: S2SLS residuals - mean LTPC vs WS

S2SLS used the WS-lag as an instrument to manage the endogeneity of the LTPC-lag, whereas this feature is absent from the HLM 7 spatial dependence model. Our approach was to emulate the HLM spatial dependence model by introducing spatial lag terms as explanatory variables. This raised concerns of collinearity problems in the model because of how closely related the spatial lag variables were to WS. The issue of overfitting was not likely to be a major concern due to the large sample size, while a workaround was proposed in the discussion to address the issue of variance overlap.

Table 8: Spatial Two Stage Least Squares Exploratory Analysis Results

Statistical Test	NSWM Inverse Exponent[28]	
DV	LTPC	LTPC
NSWM Inverse Exponent	2.0	2.1
Anselin-Kelejian Test[29] (AKT) score	4.042	2.461
AKT p-value	0.0444*	0.1167
Pseudo R-squared	0.6466	0.6489

[29] The Anselin-Kelejian test checks the null hypothesis that there is no spatial autocorrelation in the residuals of the model.

Spatial Pseudo R-squared	0.5457	0.5447
Spatial Lag (Wy) coefficient	0.4494	0.4471
LTPC Wy z-statistic	10.6868	10.2193
LTPC Wy p-value	0.0000	0.0000
WS coefficient	0.0109	0.0108
WS z-statistic	15.3417	14.5715
WS p-value	0.0000	0.0000
Conclusion	RSA[30]	RNSA[31]

Equation 2: Simplified S2SLS Model Equation

$$Y = \rho Wy + X\beta + \varepsilon$$

$$
\begin{aligned}
&\textit{where}: \\
&Y = DV\ (mean\ LTPC) \\
&X = IV\ (WS) \\
&Wy = Spatial\ Lag\ of\ mean\ LTPC \\
&\rho = \text{Spatial Lag coefficient} \\
&\beta = IV\ coefficient \\
&\varepsilon = error\ term
\end{aligned}
$$

Table 9: Selected Network Spatial Weights Matrix Specifications and Descriptives

Descriptive/Specification	Value
Number of origins	2999
Impedance type	Network inverse distance
Impedance cutoff distance	9,000 m
Inverse Distance Exponent	2.1
Minimum number of neighbours	1
Maximum number of neighbours	1,595
Mean number of neighbours	830.6
Maximum distance between neighbours	8,940 m

[30] RSA = AKT null hypothesis was rejected, residuals were spatially autocorrelated
[31] RNSA = AKT null hypothesis was accepted, residuals were not spatially autocorrelated

Multi-Level Modeling

The MLMs comprised two levels:

- Level 1: IVs = Hourly LTPC, weather, season and time IVs
- Level 2: DV = Mean LTPC, IVs = Mean LTPC-lag (Wy), WS, WS-lag (Wx)

Spatial auto-correlation was accounted for at level-2 of the MLM by including the selected DV and IV HLM Spatial Lag vector (Wy and Wx) generated in GeoDaSpace. The spatial lag term Wy effectively replaced the spatial dependence term "b_o" in the HLM7 spatial dependence model (Verbitsky-Savitz & Raudenbush, 2009). The built-in HLM7 spatial dependence model could not accept the large NSWM describing the spatial relationships of the sampling locations in this study, likely due to memory limitations. The hourly data at level 1 did not vary between sampling locations (the weather data was aggregated from multiple stations in the region but was in effect stationary). There was likely a combination of spatial and temporal autocorrelation, since pedestrians move between volume measurement locations. For the purposes of this study, the model was be simplified by accounting for spatial dependence only at level 2, accounting for temporal dependence only at level 1 within HLM, but not modeling for the potential interaction between them. Because the spatial lag terms were correlated with the error term in level-2, the regression had to be completed using either the Maximum likelihood (ML) or the S2SLS approach (Anselin, 1988). The specific concerns raised in Morenoff (Morenoff J. D., 2003) regarding this spatial regression approach in HLM were mitigated by the fact that HLM 7 uses either restricted maximum likelihood (REML) or full maximum likelihood (ML) regression models, rather than OLS. In addition, comparison of ML and OLS using the spatial lag vectors yielded nearly identical results, indicating that the large sample size allowed model solution despite the high correlation between the DV residual spatial lag and the DV itself. Given the dataset properties provided in Table 2, ML was not likely to bias the analysis. ML tends to underestimate within sampling location variance: as the number of IVs increases and/or group size decreases, sigma (within sampling location variance) is underestimated, thus inflating the Type 1 error rate. In addition, as group size decreases, the values in the variance/co-variance matrix become biased. The MLM is expressed in simplified form per Equation 3, and was run with both REML and ML in HLM.

Equation 3: Simplified general form of two-level MLM

$$Level\ 1\ (Hourly): Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + r_{ij}$$
$$Level\ 2\ (Sampling\ Location): \beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j},$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j}$$

$where$:
$Y_{ij} = \mathrm{i}^{th}\ hourly\ LTPC\ at\ sampling\ location\ j$
$\beta_{0j} = hourly\ intercept$
$\beta_{1j} = hourly\ coefficients\ for\ weather, season\ and\ time\ category\ IVs$
$X_{ij} = hourly\ weather, season\ and\ time\ category\ IVs$
$r_{ij} = hourly\ error\ term$
$\gamma_{00} =$ mean LTPC per sampling location
$\gamma_{01} = sampling\ location\ coefficients\ of\ Walk\ Score®\ and\ spatial\ lag$
$\gamma_{11} = sampling\ location\ coefficients\ of\ Walk\ Score®\ and\ spatial\ lag$
$\mathrm{Z}_j = Walk\ Score®\ and\ spatial\ lag\ (Wy\ and\ Wx)\ at\ sampling\ location\ j$
$\gamma_{10} =$ overall $coefficient\ at\ sampling\ location\ level$
$u_{0j} = error\ term\ at\ sampling\ location\ level$
$u_{1j} = error\ term\ at\ sampling\ location\ level$

Chapter 4: Results

Bivariate Correlations

Pearson and Spearman bulk (non-hierarchical) zero-order correlations for the screened and transformed variables are presented in Table 11 and Table 12. Observed positive and significant correlations between Log 10 Pedestrian Count (LTPC) and temperature (Australian Apparent Temperature or AAT, perceived or dry bulb) aligned with other studies that have shown that pedestrian volume increases with temperature (Shaaban, Muley, & Elnashar, 2017; Aultman-Hall, Lane, & Lambert, 2009; de Montigny, Ling, & Zacharias, 2012). Other substantial zero-order associations with LTPC included the following negative correlations: cloud cover, weather indicator, precipitation, snow depth, relative humidity; and positive correlations: season and atmospheric pressure. Furthermore, these correlations were all in the expected direction.

Although wind speed was not directly associated with LTPC, it was correlated with the other weather variables, most significantly with relative humidity (-.279) and season (-.102) and was thus thought to play a role in the overall model. Wind speed and relative humidity were both captured in the AAT variable along with dry bulb temperature. Atmospheric pressure was also indirectly captured by AAT through psychrometric relationships, via dry bulb temperature and relative humidity (Gatley, 2013). Although the bulk zero-order association of dry bulb temperature with LTPC was slightly greater than with AAT (.100 versus .090), AAT, being based on a published model of human thermal comfort (Steadman, 1994), was considered more robust as a single variable for use in multilevel modeling (MLM), rather than relying on an inferential multiple regression model based on several related variables and interaction terms. The perceived temperature scale based on humidex and windchill exhibited discontinuities in the frequency distribution at the temperatures where the humidex and windchill scales terminated (see Figure 4e), and though it was substantially associated with LTPC (.095), these structural problems with the scale were of concern. Furthermore, it accounted for temperature and humidity for only portions of its scale, so it was dropped from further consideration in favour of AAT. In summary, to reduce the likelihood of problems of co-linearity and loss of degrees of freedom, AAT was retained, while perceived temperature, dry bulb temperature, relative humidity, wind speed and atmospheric pressure as separate or alternative explanatory variables were dropped from further modeling.

Line graphs of the key level 1 explanatory variables retained following bivariate correlation screening are illustrated in Figure 11. The directions of the associations with LTPC were evident from these line graphs.

Figure 11: Level 1 Explanatory Variable Line Graphs vs LTPC

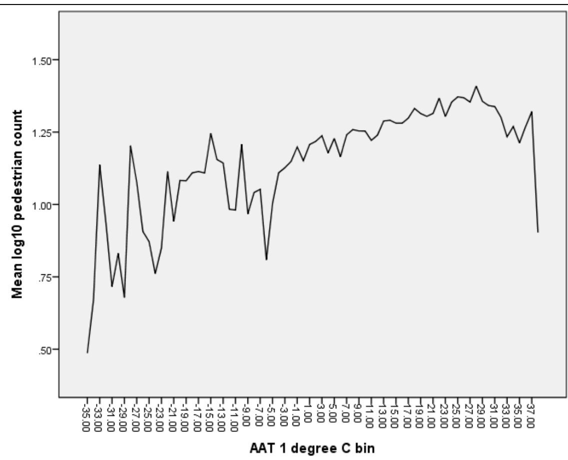


Figure 11a: Mean LTPC versus AAT (1deg bins)

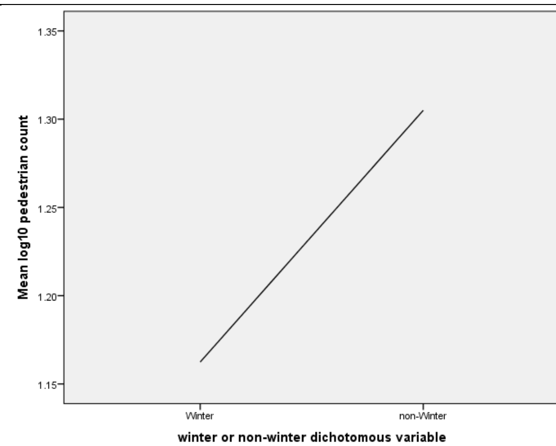


Figure 11b: Mean LTPC versus Season

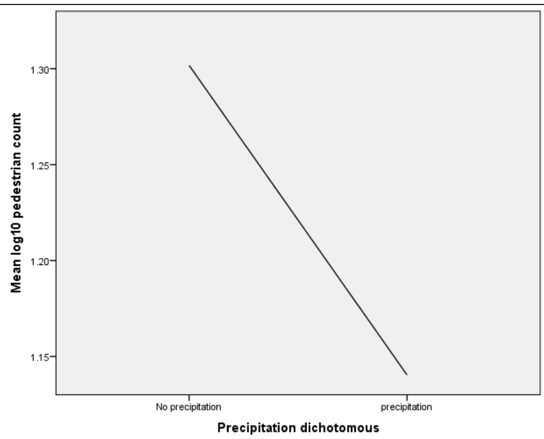


Figure 11c: Mean LTPC versus Dichotomous Precipitation

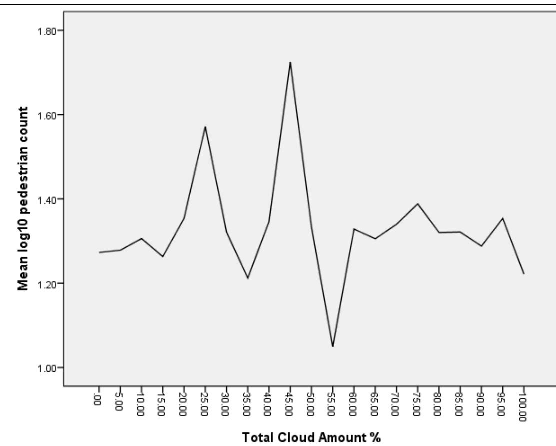


Figure 11d: Mean LTPC versus Total Cloud Amount

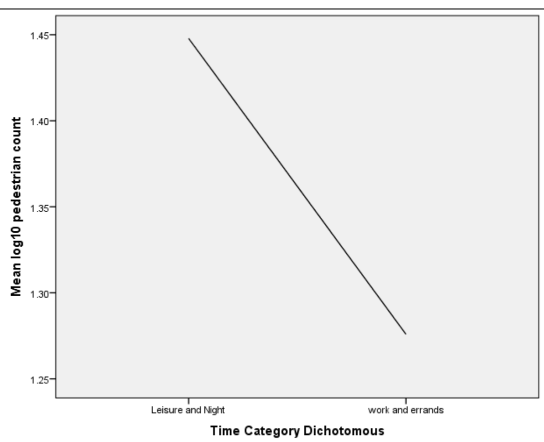


Figure 11e: Mean LTPC versus Time Category

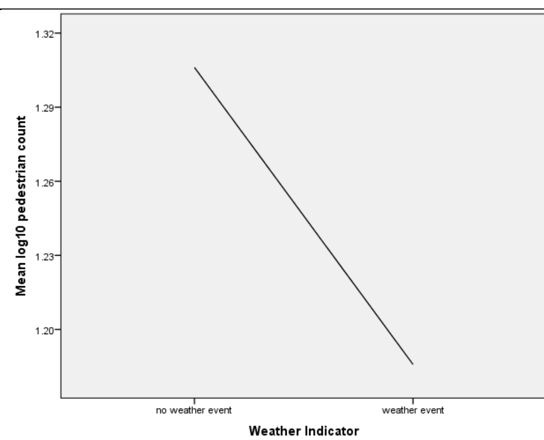


Figure 11f: Mean LTPC versus Weather Indicator

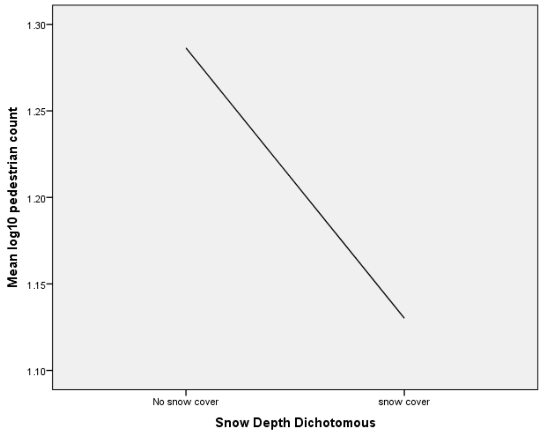


Figure 11g: Mean LTPC versus Dichotomous Snow Depth

Figure 11a provided visual evidence for a linear AAT/LTPC relationship. Increased noise associated with that sparser winter data (i.e. 0°C and below) was also evident in Figure 11a. Whether or not an interaction between season and AAT was present was difficult to ascertain given this noisiness. We would have expected to see an inflection point around 0°C (because of the hypothesized interaction between AAT and season), but that was not apparent in the data. In addition, per Figure 11a, there were too few data points at the high and low ends of the temperature range to justify a piece-wise linear model that could account for the apparent drop in walking at extreme high and low temperatures. The sample sizes below -31.8°C and above 35.9°C were relatively small (representing only 0.025% and 0.032% of the data, respectively). This was unlikely to result from a bias in the sampling of pedestrian volume, since from 1999 to 2015, hourly temperatures less than -31.8°C only occurred 0.182% of the time, and temperatures above 35.9 °C occurred only 0.014% of the time. Consequently, we opted for a simple linear model to describe the relationship of AAT with LTPC as this would be more robust in the MLM.

The relationship of LTPC with total cloud amount had a low signal to noise ratio based on Figure 11d. However, given the significant bulk zero-order correlation, the variable was maintained for further analysis. Excluding time category, line graphs of the other dichotomous variables with respect to LTPC all showed effects in the expected direction (Figure 11b: winter = less walking, Figure 11c: precipitation = less walking, Figure 11f: severe weather event = less walking, Figure 11g: snow cover = less walking).

The dichotomous daylight variable had a weak but significant correlation with LTPC (-.014), but the relationship was not in the expected direction (1 represents daylight, therefore one would expect LTPC to increase with daylight). The time category variable exhibited a similar unexpected direction, despite a significant correlation of -.056. With time category, the direction of the effect shown in Figure 11e and Table 11 indicated less bulk walking during the WE (work and errands) time category. One would expect more walking during rush hour periods included in the WE time category. Indeed, previous work (Aultman-Hall, Lane, & Lambert, 2009) found that the most walking occurs at times of day matching those of the present WE time category, which is opposite to the effect direction that was observed here. Their study also suggests that sampling location could have significantly impacted these results. This unexpected result was likely due to measurement bias in the secondary data set. For example, in the LN category, 8.8% of the LTPC measurements were zero, and 28.8 % were 1 or less. In contrast, with the WE time category, 11.2% of LPTC measurements were zero, while 37.0% were 1 or less. This discrepancy suggested a sampling location and sampling time bias, such that many more locations with little pedestrian traffic were surveyed during the WE time category compared to the LN time category.

Previous work also suggests that time moderates the effect of weather on walking, such that weather conditions would have less of an effect on walking during utilitarian trips. This appeared to be the case in our data (with the exception of total cloud amount) based on the bulk zero-order correlations by time category indicated in Table 10, suggesting that further examination of these potential interactions would be warranted. Table 13 summarizes significant correlations of IVs with LTPC.

Table 10: Zero-order Pearson correlations of AAT with LTPC by time category

Weather Variable	Test	WE	LN
AAT	Pearson Correlation	.083**	.174**
	N	95,336	7,524
Precipitation (dichotomous)	Pearson Correlation	-.042**	-.097**
	N	91,693	7,162
Total Cloud Amount	Pearson Correlation	-.043**	-.043**
	N	84,280	5,235
Season (dichotomous)	Pearson Correlation	.051**	.120**
	N	95,345	7,528
Weather Indicator	Pearson Correlation	-.052**	-.081**
	N	95,336	7,524
Snow Cover (dichotomous)	Pearson Correlation	-.055**	-.126**
	N	49,455	2,793

Table 11: Bivariate Pearson Correlations[32]

		lg10_cnt	AAT	Perc_T	Tot_Cloud	Season_dich	Precip_dich	Weather_Indic	Snow_Dich	Snow_Depth	DBTemp	RelHum	Wind_Spd	tcat_dich	DLight	Precip_ALL	Atm_Pres
lg10_cnt	PR	1	.090**	.095**	-.030**	.057**	-.046**	-.053**	-.059**	-.042**	.100**	-.087**	0.003	-.056**	-.014**	-.014**	.023**
	p2t		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.419	0.000	0.000	0.000	0.000
	N	102873	102860	102812	89515	102873	98855	102860	52248	52248	102860	102860	102860	102873	102873	98855	102860
AAT	PR	.090**	1	.979**	-.031**	.642**	-.072**	-.113**	-.698**	-.604**	.980**	-.167**	-.172**	.008*	.096**	-.008**	-.216**
	p2t	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.011	0.000	0.009	0.000
	N	102860	102860	102812	89515	102860	98842	102860	52248	52248	102860	102860	102860	102860	102860	98842	102860
Perc_T	PR	.095**	.979**	1	-.052**	.611**	-.087**	-.146**	-.673**	-.586**	.983**	-.254**	-.041**	0.004	.101**	-.014**	-.216**
	p2t	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.158	0.000	0.000	0.000
	N	102812	102812	102812	89474	102812	98795	102812	52237	52237	102812	102812	102812	102812	102812	98795	102812
Tot Cloud	PR	-.030**	-.031**	-.052**	1	-.032**	.211**	.332**	.015**	0.004	-.067**	.337**	.027**	-.024**	-0.001	.058**	-.340**
	p2t	0.000	0.000	0.000		0.000	0.000	0.000	0.001	0.354	0.000	0.000	0.000	0.000	0.837	0.000	0.000
	N	89515	89515	89474	89515	89515	86486	89515	49271	49271	89515	89515	89515	89515	89515	86486	89515
Season dich	PR	.057**	.642**	.611**	-.032**	1	-0.001	-.081**	-.698**	-.583**	.632**	-.011**	-.102**	.038**	.100**	0.004	-.156**
	p2t	0.000	0.000	0.000	0.000		0.757	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.175	0.000
	N	102873	102860	102812	89515	102873	98855	102860	52248	52248	102860	102860	102860	102873	102873	98855	102860
Precip dich	PR	-.046**	-.072**	-.087**	.211**	-0.001	1	.424**	-0.002	0.007	-.093**	.273**	.040**	.011**	0.003	.255**	-.196**
	p2t	0.000	0.000	0.000	0.000	0.757		0.000	0.658	0.122	0.000	0.000	0.000	0.001	0.377	0.000	0.000
	N	98855	98842	98795	86486	98855	98855	98842	52191	52191	98842	98842	98842	98855	98855	98855	98842
Weather Indic	PR	-.053**	-.113**	-.146**	.332**	-.081**	.424**	1	.117**	.093**	-.163**	.448**	-.024**	.006*	-.013**	.121**	-.214**
	p2t	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.037	0.000	0.000	0.000
	N	102860	102860	102812	89515	102860	98842	102860	52248	52248	102860	102860	102860	102860	102860	98842	102860
Snow Dich	PR	-.059**	-.698**	-.673**	.015**	-.698**	-0.002	.117**	1	.752**	-.701**	.052**	.121**	-.117**	-.212**	-0.004	.171**
	p2t	0.000	0.000	0.000	0.001	0.000	0.658	0.000		0.000	0.000	0.000	0.000	0.000	0.000	0.366	0.000
	N	52248	52248	52237	49271	52248	52191	52248	52248	52248	52248	52248	52248	52248	52248	52191	52248
Snow Depth	PR	-.042**	-.604**	-.586**	0.004	-.583**	0.007	.093**	.752**	1	-.610**	.022**	.117**	-.054**	-.095**	0.002	.150**
	p2t	0.000	0.000	0.000	0.354	0.000	0.122	0.000	0.000		0.000	0.000	0.000	0.000	0.000	0.709	0.000
	N	52248	52248	52237	49271	52248	52191	52248	52248	52248	52248	52248	52248	52248	52248	52191	52248
DBTemp	PR	.100**	.980**	.983**	-.067**	.632**	-.093**	-.163**	-.701**	-.610**	1	-.311**	-.022**	0	.106**	-.016**	-.207**
	p2t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.916	0.000	0.000	0.000
	N	102860	102860	102812	89515	102860	98842	102860	52248	52248	102860	102860	102860	102860	102860	98842	102860
RelHum	PR	-.087**	-.167**	-.254**	.337**	-.011**	.273**	.448**	.052**	.022**	-.311**	1	-.261**	.075**	-.040**	.079**	-.264**

[32] Lg10cnt = LTPC
Precip_ALL = Total Precipitation
Precip_dich = dichotomous precipitation
DBTemp = Dry Bulb Temperature
Snow_Dich = Snow depth dichotomous
Season_Dich = Season dichotomous
RelHum = relative humidity
AAT = Australian Apparent Temperature
Perc_T = perceived temperature

	p2t	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000		0.000	0.000	0.000	0.000	0.000
	N	102860	102860	102812	89515	102860	98842	102860	52248	52248	102860	102860	102860	102860	102860	98842	102860
Wind_Spd	PR	0.003	-.172**	-.041**	.027**	-.102**	.040**	-.024**	.121**	.117**	-.022**	-.261**	1	-0.005	.038**	.009**	-.197**
	p2t	0.419	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.132	0.000	0.003	0.000
	N	102860	102860	102812	89515	102860	98842	102860	52248	52248	102860	102860	102860	102860	102860	98842	102860
tcat_dich	PR	-.056**	.008*	0.004	-.024**	.038**	.011**	.006*	-.117**	-.054**	0	.075**	-0.005	1	.398**	0.004	-.030**
	p2t	0.000	0.011	0.158	0.000	0.000	0.001	0.037	0.000	0.000	0.916	0.000	0.132		0.000	0.203	0.000
	N	102873	102860	102812	89515	102873	98855	102860	52248	52248	102860	102860	102860	102873	102873	98855	102860
DLight	PR	-.014**	.096**	.101**	-0.001	.100**	0.003	-.013**	-.212**	-.095**	.106**	-.040**	.038**	.398**	1	0.004	-.036**
	p2t	0.000	0.000	0.000	0.837	0.000	0.377	0.000	0.000	0.000	0.000	0.000	0.000	0.000		0.191	0.000
	N	102873	102860	102812	89515	102873	98855	102860	52248	52248	102860	102860	102860	102873	102873	98855	102860
Precip ALL	PR	-.014**	-.008**	-.014**	.058**	0.004	.255**	.121**	-0.004	0.002	-.016**	.079**	.009**	0.004	0.004	1	-.056**
	p2t	0.000	0.009	0.000	0.000	0.175	0.000	0.000	0.366	0.709	0.000	0.000	0.003	0.203	0.191		0.000
	N	98855	98842	98795	86486	98855	98855	98842	52191	52191	98842	98842	98842	98855	98855	98855	98842
Atm_Pres	PR	.023**	-.216**	-.216**	-.340**	-.156**	-.196**	-.214**	.171**	.150**	-.207**	-.264**	-.197**	-.030**	-.036**	-.056**	1
	p2t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	
	N	102860	102860	102812	89515	102860	98842	102860	52248	52248	102860	102860	102860	102860	102860	98842	102860

Table 12: Bivariate Spearman Correlations

		lg10_cnt	AAT	Perc_T	Tot_Cloud	Season_dich	Precip_dich	Weather_Indic	Snow_Dich	Snow_Depth	DBTemp	RelHum	Wind_Spd	tcat_dich	DLight	Precip_ALL	Atm_Pres
lg10_cnt	SR	1.000	.076**	.087**	-.042**	.055**	-.046**	-.055**	-.061**	-.062**	.090**	-.089**	.008**	-.058**	-.014**	-.046**	.024**
	p2t	.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.008	0.000	0.000	0.000	0.000
	N	102873	102860	102812	89515	102873	98855	102860	52248	52248	102860	102860	102860	102873	102873	98855	102860
AAT	SR	.076**	1.000	.974**	-.093**	.476**	-.100**	-.110**	-.482**	-.449**	.963**	-.211**	-.146**	-.022**	.065**	-.099**	-.192**
	p2t	0.000	.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	N	102860	102860	102812	89515	102860	98842	102860	52248	52248	102860	102860	102860	102860	102860	98842	102860
Perc_T	SR	.087**	.974**	1.000	-.119**	.464**	-.118**	-.157**	-.481**	-.448**	.985**	-.321**	.012**	-.025**	.072**	-.117**	-.193**
	p2t	0.000	0.000	.	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	N	102812	102812	102812	89474	102812	98795	102812	52237	52237	102812	102812	102812	102812	102812	98795	102812
Tot Cloud	SR	-.042**	-.093**	-.119**	1.000	-.053**	.252**	.392**	.032**	.010*	-.151**	.390**	.015**	-.022**	-0.004	.253**	-.335**
	p2t	0.000	0.000	0.000	.	0.000	0.000	0.000	0.000	0.029	0.000	0.000	0.000	0.000	0.274	0.000	0.000
	N	89515	89515	89474	89515	89515	86486	89515	49271	49271	89515	89515	89515	89515	89515	86486	89515
Season dich	SR	.055**	.476**	.464**	-.053**	1.000	-0.001	-.081**	-.698**	-.649**	.459**	-.014**	-.089**	.038**	.100**	0.000	-.144**
	p2t	0.000	0.000	0.000	0.000	.	0.757	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.895	0.000
	N	102873	102860	102812	89515	102873	98855	102860	52248	52248	102860	102860	102860	102873	102873	98855	102860
Precip dich	SR	-.046**	-.100**	-.118**	.252**	-0.001	1.000	.424**	-0.002	0.004	-.133**	.274**	.036**	.011**	0.003	.999**	-.190**
	p2t	0.000	0.000	0.000	0.000	0.757	.	0.000	0.658	0.385	0.000	0.000	0.000	0.001	0.377	0.000	0.000
	N	98855	98842	98795	86486	98855	98855	98842	52191	52191	98842	98842	98842	98855	98855	98855	98842

Weather Indic	SR	-.055**	-.110**	-.157**	.392**	-.081**	.424**	1.000	.117**	.107**	-.187**	.452**	-.035**	.006*	-.013**	.426**	-.205**
	p2t	0.000	0.000	0.000	0.000	0.000	0.000	.	0.000	0.000	0.000	0.000	0.000	0.037	0.000	0.000	0.000
	N	102860	102860	102812	89515	102860	98842	102860	52248	52248	102860	102860	102860	102860	102860	98842	102860
Snow Dich	SR	-.061**	-.482**	-.481**	.032**	-.698**	-0.002	.117**	1.000	.921**	-.479**	.049**	.100**	-.117**	-.212**	-0.004	.165**
	p2t	0.000	0.000	0.000	0.000	0.000	0.658	0.000	.	0.000	0.000	0.000	0.000	0.000	0.000	0.409	0.000
	N	52248	52248	52237	49271	52248	52191	52248	52248	52248	52248	52248	52248	52248	52248	52191	52248
Snow Depth	SR	-.062**	-.449**	-.448**	.010*	-.649**	0.004	.107**	.921**	1.000	-.447**	.048**	.092**	-.077**	-.144**	0.002	.152**
	p2t	0.000	0.000	0.000	0.029	0.000	0.385	0.000	0.000	.	0.000	0.000	0.000	0.000	0.000	0.627	0.000
	N	52248	52248	52237	49271	52248	52191	52248	52248	52248	52248	52248	52248	52248	52248	52191	52248
DBTemp	SR	.090**	.963**	.985**	-.151**	.459**	-.133**	-.187**	-.479**	-.447**	1.000	-.408**	.031**	-.035**	.076**	-.133**	-.165**
	p2t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.	0.000	0.000	0.000	0.000	0.000	0.000
	N	102860	102860	102812	89515	102860	98842	102860	52248	52248	102860	102860	102860	102860	102860	98842	102860
RelHum	SR	-.089**	-.211**	-.321**	.390**	-.014**	.274**	.452**	.049**	.048**	-.408**	1.000	-.279**	.076**	-.040**	.275**	-.263**
	p2t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.	0.000	0.000	0.000	0.000	0.000
	N	102860	102860	102812	89515	102860	98842	102860	52248	52248	102860	102860	102860	102860	102860	98842	102860
Wind_Spd	SR	.008**	-.146**	.012**	.015**	-.089**	.036**	-.035**	.100**	.092**	.031**	-.279**	1.000	-0.001	.045**	.035**	-.173**
	p2t	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.	0.832	0.000	0.000	0.000
	N	102860	102860	102812	89515	102860	98842	102860	52248	52248	102860	102860	102860	102860	102860	98842	102860
tcat_dich	SR	-.058**	-.022**	-.025**	-.022**	.038**	.011**	.006*	-.117**	-.077**	-.035**	.076**	-0.001	1.000	.398**	.011**	-.035**
	p2t	0.000	0.000	0.000	0.000	0.000	0.001	0.037	0.000	0.000	0.000	0.000	0.832	.	0.000	0.001	0.000
	N	102873	102860	102812	89515	102873	98855	102860	52248	52248	102860	102860	102860	102873	102873	98855	102860
DLight	SR	-.014**	.065**	.072**	-0.004	.100**	0.003	-.013**	-.212**	-.144**	.076**	-.040**	.045**	.398**	1.000	0.003	-.036**
	p2t	0.000	0.000	0.000	0.274	0.000	0.377	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.	0.332	0.000
	N	102873	102860	102812	89515	102873	98855	102860	52248	52248	102860	102860	102860	102873	102873	98855	102860
Precip_ALL	SR	-.046**	-.099**	-.117**	.253**	0.000	.999**	.426**	-0.004	0.002	-.133**	.275**	.035**	.011**	0.003	1.000	-.191**
	p2t	0.000	0.000	0.000	0.000	0.895	0.000	0.000	0.409	0.627	0.000	0.000	0.000	0.001	0.332	.	0.000
	N	98855	98842	98795	86486	98855	98855	98842	52191	52191	98842	98842	98842	98855	98855	98855	98842
Atm_Pres	SR	.024**	-.192**	-.193**	-.335**	-.144**	-.190**	-.205**	.165**	.152**	-.165**	-.263**	-.173**	-.035**	-.036**	-.191**	1.000
	p2t	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	.
	N	102860	102860	102812	89515	102860	98842	102860	52248	52248	102860	102860	102860	102860	102860	98842	102860

Table 13: Significant IV correlations with LTPC, ranked in descending order

Variable	**Pearson**	**Spearman Rho**
DBTemp	0.100	0.090
Perc_T	0.095	0.087
AAT	0.090	0.076
RelHum	0.087	0.089
Snow_Dich	0.059	0.061
Season_dich	0.057	0.055
tcat_dich	0.056	0.058
Weather_Indic	0.053	0.055
Precip_dich	0.046	0.046
Snow_Depth	0.042	0.062
Tot_Cloud	0.030	0.042
Atm_Pres	0.023	0.024
Precip_ALL	0.014	0.046
DLight	0.014	0.014

Interactions

To evaluate the potential role of pair-wise interactions among the independent variables in the MLM, F-values were calculated for each IV pair, (Table 14). Very high F-values (>50) were generally regarded as suspect, in particular the daylight variable pairs with season, AAT, dry bulb temperature, perceived temperature and relative humidity. There was nonetheless evidence of several potential valid interactions between time category, season and weather IVs.

Table 14: LTPC as DV, GLM[33] Interactions between IVs (F value)[34]

Variable	AAT	Perc. Temp	Tot Cloud	Precip	Precip. Dich.	Weath. Indic.	Snow Depth	Snow Dich.	Time Cat	Day light	Rel Hum	Wind Spd	Dry Bulb Temp	Atm Press.
Season	**10.381 (.001)**	**6.608 (.010)**	.471 (.492)	5.555 (.018)	.003 (.955)	.215 (.643)	.477 (.490)	**7.541 (.006)**	**24.371 (.000)**	**136.194 (.000)**	**10.435 (.001)**	**19.241 (.000)**	.100 (.752)	**24.769 (.000)**
AAT		**51.591 (.000)**	**5.970 (.015)**	1.909 (.167)	1.914 (.167)	**8.169 (.004)**	**5.107 (.024)**	.258 (.611)	**42.791 (.000)**	**157.240 (.000)**	**53.473 (.000)**	.947 (.331)	.753 (.386)	1.666 (.197)
Perc Temp.			**12.046 (.001)**	.810 (.368)	**8.600 (.003)**	**11.794 (.001)**	**6.092 (.014)**	.539 (.463)	**33.952 (.000)**	**147.121 (.000)**	**58.693 (.000)**	.082 (.775)	.294 (.587)	2.506 (.113)
Tot Cloud				.073 (.787)	**36.260 (.000)**	**5.996 (.014)**	.710 (.400)	4.071 (.044)	.002 (.969)	1.056 (.304)	**25.581 (.000)**	1.549 (.213)	**15.007 (.000)**	**67.887 (.000)**
Precip.					N/A	.811 (.368)	3.561 (.059)	2.292 (.130)	**11.055 (.001)**	**13.820 (.000)**	.198 (.656)	0.646 (.421)	2.050 (.152)	0.025 (.874)
Precip Dich						**19.047 (.000)**	0.549 (.459)	0.205 (.651)	**22.921 (.000)**	3.181 (.075)	**36.302 (.000)**	0.908 (.341)	**6.449 (.011)**	**6.506 (.011)**
Weather Indic.							.002 (.969)	0.279 (.598)	**6.574 (.010)**	1.102 (.294)	**7.950 (.005)**	**22.873 (.000)**	**7.912 (.005)**	**33.507 (.000)**
Snow Depth								N/A	**22.810 (.000)**	**22.404 (.000)**	**14.299 (.000)**	4.990 (.025)	**11.503 (.001)**	1.140 (.286)
Snow Dich.									2.343 (.126)	2.728 (.099)	**20.557 (.000)**	1.845 (.174)	2.838 (.092)	0.556 (.456)
Time Cat										**49.813 (.000)**	**9.180 (.002)**	**38.659 (.000)**	**33.486 (.000)**	**91.317 (.000)**
Day light											**5.854 (.016)**	**31.508 (.000)**	**136.907 (.000)**	1.091 (.296)
Rel Hum												.167 (.683)	**71.359 (.000)**	3.038 (.081)
Wind Spd													0.032 (.858)	**17.030 (.000)**
Dry Bulb Temp														0.007 (.932)

Table 15: Mean LTPC by Season and Daylight Cells

Mean LTPC	Dark	Light
Winter	.9902 (N=686)	1.1729 (N=11,257)
Non-Winter	1.5653 (N=1,315)	1.3012 (N=89,615)

Per Table 15, the substantial interaction between Season and daylight with respect to LTPC may have been due to a change in slope direction, whereby walking at night in non-winter was greater than walking during the day in non-winter – an unexpected outcome. This could be due to a

[33] GLM = General Linear Model

[34] Highlighted values have F>5.03, two-tailed significance for 2,999 intersections.

measurement bias, also implied by the small number of data points for non-winter/dark, and a lack of variability within sampling locations, per Table 5. Looking at the same cells, but with cell mean Walk Score as the DV as a coarse measure of geographic bias, we can see in Table 16 that measurements of pedestrian volume were completed on average in more walkable areas at night in the non-winter compared to measurements during the day in non-winter, supporting the geographic bias implied in Table 15. Therefore, any interactions based on the daylight variable were considered suspicious, and in fact the daylight dichotomous variable was dropped to avoid this bias.

Table 16: Mean Walk Score by Season and Daylight Cells

Mean Walk Score	Dark	Light
Winter	56.66 (N=686)	58.1921 (N=11,257)
Non-Winter	74.1163 (N=1315)	62.7913 (N=89,615)

As indicated previously, AAT encompassed interactions of dry bulb temperature, relative humidity and wind speed with respect to human comfort. Since the relationships between dry bulb temperature, wet bulb temperature, dew point temperature, relative humidity, and atmospheric pressure can be mathematically described based on psychrometric relationships, the interactions observed between some of these variables and other variables with respect to LTPC, and direct correlations with LTPC, were considered to be encompassed by AAT based on biometeorological theory. Therefore, relative humidity, wind speed and dry bulb temperature were dropped from further investigation.

Aultman-Hall reported an interaction between precipitation and season, whereby increased reduction in pedestrian volume was observed due to precipitation from 13% in non-winter to 16% in winter (Aultman-Hall, Lane, & Lambert, 2009). This interaction was not observed in our data, however an interaction between precipitation and time category was observed in the expected direction (greater reduction in walking when it rains when walking for leisure compared to utilitarian trips), per Table 14 and Table 17.

Table 17: Mean LTPC by Precipitation and Time Category

Mean LTPC	No Precip	Precip
Work and Errands	1.2887 (N=86,467)	1.1576 (N=5,186)
Leisure and Night	1.4581 (N=6805)	1.1673 (N=357)

The weather indicator variable, though related to LTPC per Table 13, introduced further loss of variability in the dataset due to the small number of observations of extreme weather (per Table 4). Furthermore, this variable encompassed a variety of severe weather event types characterized by limiting visibility, per the Environment Canada definition (Environment and Climate Change Canada, 2015). Several stations in Canada do not even have the capability to record this information, and its availability in other countries and cities is unlikely. Hence, due to the poor quality of the data, and to increase the reproducibility of this model in other areas, this variable was dropped from further study.

Although dichotomous snow cover related to LTPC, it did not exhibit the expected interaction with temperature, whereby snow cover and low temperatures would be expected to interact to yield a greater decrease in walking. This could be due to more complex time and sequence-dependent pedestrian responses to winter weather phenomena (e.g. low temperatures often co-occur with high pressure and low snow fall, which could promote walking), or snow clearing response times, which were not studied here. Winter conditions appeared to have the expected effect on the slope of AAT with LTPC, whereby in winter a decrease in AAT more negatively impacted LTPC than in non-winter. Since dichotomous snow cover and season were highly correlated (-.698), one of these could be considered redundant. Since season exhibited the expected interactions with temperature, but dichotomous snow cover did not, the latter was dropped. Table 18 provides an initial summary of variables retained for multi-level modeling on the basis of both their interactions and correlations with LTPC

Table 18: Summary of selected Level-1 IV interactions, significant F values, LTPC as DV, with significant LTPC correlations and between variable correlations

Variable	AAT (.090)	Season (.057)	Total Cloud (.030)	Precip. Dich (.046)	Time Cat (.056)
AAT (.090)					
Season (.057)	**F=10.4**, R=.642				
Total Cloud (.030)	**F=5.97**, R=-.031	R=-.032			
Precip. Dich. (.046)	R=-.072		**F=36.26**, R=.211		
Time Cat (.056)	**F=42.79**, R=.008	**F = 24.4**, R =.038	R=-.024	**F=22.9**, R=.011	

A final selection of variables and interaction terms for multi-level modeling was summarized in Table 19. The interaction terms were calculated per Equation 4 using the HLM grand means listed in Table 20. The selected bulk main effects did exhibit some collinearity, as summarized in Table 18. Most significant was the collinearity of Season with AAT, where R=.642. Otherwise, within IV correlation were below the typical rule of thumb threshold of .6 to consider combining variables. Given the observed interaction between AAT and Season, they were carried forward as separate variables for the MLMs. Within HLM, the importance of collinearity problems was subsequently re-examined.

Table 19: Final summary of selected Level-1 IVs and interaction terms

Var Type	Variable
IV	AAT
	Season
	Precip. Dich
	Total Cloud
	Time Cat
Interaction Term	Season * AAT
	Season * Time Cat
	Total Cloud * AAT
	Total Cloud * Precip. Dich
	Time Cat * Precip. Dich.
	Time Cat * AAT

Table 20: HLM Grand Means for Level-1 Interaction Terms

Variable	HLM Grand Mean
AAT	13.524885
Season	0.809956
Precip. Dich	0.056602
Total Cloud	62.657065
Time Cat	0.943201

Equation 4: Level-1 Interaction Terms

$$d(x_1 \times x_2) = \left(x_1 - \overline{x_1}_{HLM}\right) \times \left(x_2 - \overline{x_2}_{HLM}\right)$$

Multi-Level Model Results and Diagnostics

Given the secondary nature of the dataset and highly variable distribution of the data across sampling locations, or level-2 groups, multiple MLM models were screened in a forward step-wise fashion to identify relevant main effects and interactions while accounting for the nested structure of the data. Interaction terms were individually tested in the MLMs to account for situations where the main effects might be insignificant, but where the interactions were significant. This included inter-level interactions on a model that included only the final subset of level-1 main effects and interactions. Each of the screening models was examined for appropriate variable coefficient correlations, level-1 coefficient reliability, coefficient statistical significance, and fraction of total variance explained, as summarized in Table 21. Each model within that table was referenced with an alphanumeric ID (A through AN). Only models with sufficient reliability while including all level-1 error terms were carried forward.

Models A through H represented the different permutations of the three level-2 explanatory variables LTPC-lag (Wy or W_log10count), WS and WS-lag (Wx or W_WS). Models I through AI represented the different permutations of the selected 5 level-1 main effects and the 6 level-1 interactions. From Models I-AI, a single significant model having the greatest fraction of variance explained using the fewest significant level-1 explanatory variables, and maintaining a number of degrees of freedom for error (DFE) of at least 1,000 (maximum possible of 2,998 with only the DV), was selected and carried forward to be combined with the level-2 variables. Models AJ through AN included all of the selected level-1 and level-2 variables, but varied in the application of cross-level interactions to the level-1 variables with the main level-2 variable of interest, WS. A single, final model was selected from models AJ-AN that maximized the fraction of variance explained using significant level-1 main effects and interactions, level-2 main effects, and cross-level interactions. The final model was then diagnosed in more detail.

Table 21: MLM Outcome Summary

MLM ID	Variable Applied: DV	Variable Applied: Level-1 IV											Variable Applied: Level-2 IV			Variance Explained			Degrees of Freedom for Error	Fraction of Total Variance Explained	Variable Coefficient Correlation Check (<.9)	Level 1 Coefficient Reliability Estimate Check (>0.05)	Notes (all sig. = p-values all <0.001,unless otherwise noted)	Preliminary Model Quality
	LTPC	AAT (RH, Wspd, DBTemp)	Season	Precip. Dich	Total Cloud	Time Cat	Season * AAT	Season * Time Cat	Total Cloud * AAT	Total Cloud * Precip. Dich	Time Cat * Precip. Dich.	Time Cat * AAT	WS	W_log10count	W_WS	Itrcpt1	level-1	Total						
A	■															0.4451	0.0997	0.5448	2998	0.00%	ok	ok	all sig.	GOOD
B	■												■			0.2017	0.0998	0.2434	2997	44.67%	ok	ok	all sig.	GOOD
C	■														■	0.2042	0.0998	0.2409	2997	44.21%	ok	ok	all sig.	GOOD
D	■													■		0.1483	0.0998	0.2967	2997	54.46%	ok	ok	all sig.	GOOD
E	■													■	■	0.1450	0.0998	0.3000	2996	55.07%	ok	ok	all sig.	GOOD
F	■												■	■	■	0.1340	0.0998	0.3110	2995	57.09%	ok	ok	all sig. W_WS has small neg. slope	GOOD
G	■												■	■		0.1373	0.0998	0.3077	2995	56.48%	ok	ok	all sig.	GOOD
H	■												■		■	0.1961	0.0998	0.2490	2996	45.71%	ok	ok	all sig.	GOOD
I	■	■														0.4415	0.0939	0.0095	2943	1.74%	ok	ok	all sig.	GOOD
J	■		■													0.4401	0.0964	0.0084	820	1.53%	ok	ok	non-sig: season	POOR
K	■						■									0.4440	0.0954	0.0054	2973	0.99%	ok	ok	non-sig: season*AAT	POOR
L	■	■	■													0.4392	0.0916	0.0140	816	2.58%	ok	ok	all sig., but season effect in wrong direction	POOR
M	■	■	■				■									0.4390	0.0909	0.0149	741	2.73%	ok	ok	all sig., but season effect in wrong direction	POOR
N	■					■										0.4412	0.0937	0.0100	786	1.83%	ok	ok	non-sig: tcat. also more walking during leisure	POOR
O	■							■								0.4431	0.0933	0.0084	1256	1.55%	ok	ok	non-sig, negative slope	POOR
P	■		■			■										0.4366	0.0908	0.0174	349	3.19%	ok	ok	all non-sig.	POOR
Q	■		■			■		■								0.4368	0.0905	0.0176	117	3.22%	ok	ok	all non-sig.	POOR
R	■				■											0.4421	0.0941	0.0086	2792	1.58%	ok	ok	all sig., expected effect direction (less cloud, more walking)	GOOD
S	■								■							0.4419	0.0943	0.0086	2607	1.58%	ok	ok	non-sig.	POOR
T	■	■			■											0.4377	0.0894	0.0177	2584	3.25%	ok	ok	all sig.	GOOD
U	■	■			■				■							0.4372	0.0881	0.0195	1718	3.58%	ok	ok	non-sig: x_cloaat	POOR
V	■			■												0.4495	0.0973	-0.0020	1434	-0.37%	ok	ok	all sig., expected effect direction (more precip, less walk). Negative variance explained. Indicates a very small effect	POOR
W	■									■						0.4474	0.0946	0.0028	2772	0.52%	ok	ok	all sig., negative slope	GOOD
X	■			■	■											0.4471	0.0922	0.0055	1325	1.02%	ok	ok	all sig., precip and cloud expected direction. P value cloud is .005	GOOD
Y	■			■	■					■						0.4470	0.0921	0.0057	453	1.06%	ok	ok	all sig., x_clopre negative slope (ok?). Final estimation slopes p values are .014 (precip) and .005 (x_clopre)	GOOD
Z	■										■					0.4496	0.0951	0.0001	1662	0.03%	ok	ok	all sig.	GOOD
AA	■			■		■										0.4457	0.0913	0.0079	529	1.45%	ok	ok	non-sig.: tcat_dich	POOR
AB	■			■		■					■					0.4457	0.0912	0.0080	142	1.46%	ok	ok	non-sig.: tcat_dich & x_timpre	POOR
AC	■			■							■					0.4486	0.0932	0.0030	529	0.56%	*violation*	ok	non-sig.: x_timpre. Coefficient correlation >0.9	POOR
AD	■											■				0.4391	0.0925	0.0132	2996	2.43%	ok	ok	x-timaat p-value = .029 or moderately significant	FAIR
AE	■	■				■										0.4382	0.0882	0.0185	768	3.39%	ok	ok	non-sig.: tcat_dich	POOR
AF	■	■				■						■				0.4371	0.0874	0.0204	595	3.74%	ok	ok	non-sig.: tcat_dich and x_timaat	POOR
AG	■	■		■												0.4453	0.0917	0.0078	1423	1.43%	ok	ok	all sig., expected effect directions	GOOD
AH	■	■		■	■											0.4419	0.0877	0.0152	1308	2.79%	ok	ok	all sig., expected effect directions	GOOD
AI	■	■		■	■					■						0.4419	0.0875	0.0154	449	2.83%	ok	ok	all sig., expected effect directions. x_clopre has negative slope. Marginal improvement over model with no interaction terms	GOOD
AJ	■	■		■	■								■	■	■	0.1257	0.0878	0.3313	1305	60.81%	ok	ok	all sig. Tot_cloud p- value is .007. W_WS has small neg. slope. Final Model with no cross-level interaction terms	GOOD
AK	■	■*		■	■								■	■	■	0.1255	0.0878	0.3315	1305	60.85%	ok	ok	WSxAAT interaction term significant	GOOD
AL	■	■		■*	■								■	■	■	0.1257	0.0878	0.3313	1305	60.81%	ok	ok	WSxPrecip interaction term non-sig.	POOR
AM	■	■		■	■*								■	■	■	0.1256	0.0878	0.3314	1305	60.82%	ok	ok	all sig	GOOD
AN	■	■*		■	■*								■	■	■	0.1255	0.0878	0.3316	1305	60.86%	ok	ok	all sig. Final model with cross-level interactions	GOOD

sig. = significant
■ = Variable applied to model
■* = includes cross-level Walk Score interaction term with that level-1 IV

The impact of having no variance at several sampling locations became apparent and problematic with the dichotomous variables for season and time category. Of the 10 trial models that include either season or time category (Models J, L, M, N, P, Q, AA, AB, AE and AF), the number of DFE ranged from 117 to 820, or 4 to 27% of the maximum possible of 2,998. Eight of those 10 models had at least one non-significant main effect, and the two that indicated all main effects as being significant (Models L and M) had an effect for season that was not in the expected direction (i.e. more walking in winter). This suspect result was likely associated with the estimation process needed to compensate for the low DFE in those models, that stemmed from the data collection bias previously discussed.

At level-1, the remaining main effects (AAT, precipitation dichotomous and total cloud cover), independently accounted for 1.74%, -0.35% (or 0%) and 1.58% of the total variance, respectively (Models I, V and R). The only significant level-1 interaction when including the main effects was total cloud*precipitation dichotomous (Model Y), and its impact was minor as it increased the variance explained by only 0.04% above and beyond what was explained by the main effects, hence it was dropped from further consideration. A model combining the three level-1 main effects accounted for 2.79% of the variance. In summary, our selected model indicated that weather explained roughly 3% of the variance in LTPC. At any given sampling location (via group-mean centering of AAT, precipitation dichotomous and total cloud), weather accounted for approximately 12% of the variance explained in LTPC, which aligned somewhat better with the findings in Aultman-Hall (Aultman-Hall, Lane, & Lambert, 2009) whereby the impact of weather on walking was assessed longitudinally only at a single location.

Finally, cross-level interactions were examined. A small synergistic effect was observed for WS on the main effect of AAT with LTPC in the expected direction, whereby increased AAT led to a greater increase in walking at locations with higher WS values (Figure 12a, model AK). There was also a small synergistic effect of WS in the expected direction with Total Cloud and LTPC, whereby an increase in cloud cover had less impact on walking at locations with higher WS values (Figure 10b, model AM). A final model was selected that included these two cross-level interaction effects (model AN).

Figure 12: Cross-Level Interaction Plots

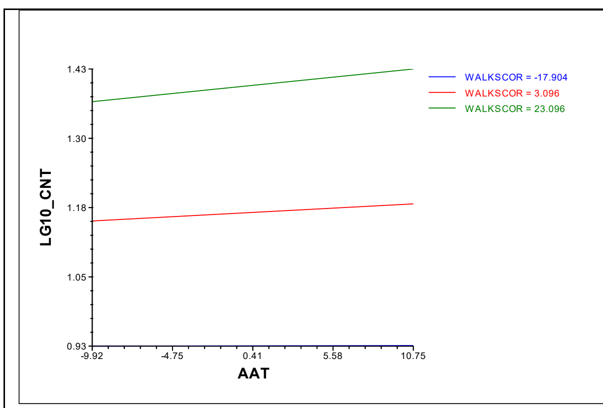


Figure 12a: Cross-Level Interactions - Walk Score and AAT[35]

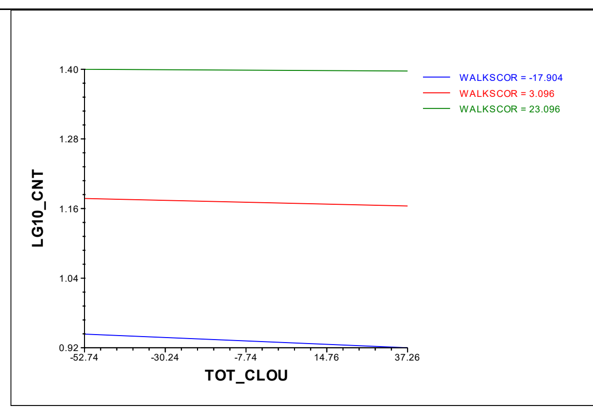


Figure 12b: Cross-Level Interactions - Walk Score and Total Cloud[36]

The final MLM (Model ID 'AN' in Table 21) can be summarized as follows in equation form:

[35] AAT was centered
[36] Total Cloud was centered

Equation 5: Final MLM Equations

Level-1 Model[37]

$$LG10\_CNT_{ij} = \beta_{0j} + \beta_{1j}*(AAT_{ij}) + \beta_{2j}*(PRECIP\_D_{ij}) + \beta_{3j}*(TOT\_CLOU_{ij}) + r_{ij}$$

Level-2 Model[38]

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(WALKSCOR_j) + \gamma_{02}*(W\_WALKSC_j) + \gamma_{03}*(W\_LG10\_C_j) + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11}*(WALKSCOR_j) + u_{1j}$$
$$\beta_{2j} = \gamma_{20} + u_{2j}$$
$$\beta_{3j} = \gamma_{30} + \gamma_{31}*(WALKSCOR_j) + u_{3j}$$

Mixed Model

$$\begin{aligned} LG10\_CNT_{ij} &= \gamma_{00} + \gamma_{01}*WALKSCOR_j + \gamma_{02}*W\_WALKSC_j + \gamma_{03}*W\_LG10\_C_j \\ &+ \gamma_{10}*AAT_{ij} + \gamma_{11}*WALKSCOR_j*AAT_{ij} \\ &+ \gamma_{20}*PRECIP\_D_{ij} \\ &+ \gamma_{30}*TOT\_CLOU_{ij} + \gamma_{31}*WALKSCOR_j*TOT\_CLOU_{ij} \\ &+ u_{0j} + u_{1j}*AAT_{ij} + u_{2j}*PRECIP\_D_{ij} + u_{3j}*TOT\_CLOU_{ij} + r_{ij} \end{aligned}$$

where β are level-1 coefficients, γ are level-2 coefficients, and u and r are the error terms.

Run-time deletion in HLM reduced the number of level-1 records to 86,486, and the number of level-2 groups to 2,891. Model solution was reached with 15 iterations. The value of the log-likelihood function at iteration 42 equaled -2.420743E+004, indicating good model convergence.

Table 22: Final MLM Level-1 IV Coefficient Correlations (τ)

Coefficient	Correlation			
INTRCPT1,β_0	1.000	0.013	-0.094	-0.011
AAT,β_1	0.013	1.000	0.231	-0.072
PRECIP_D,β_2	-0.094	0.231	1.000	0.090
TOT_CLOU,β_3	-0.011	-0.072	0.090	1.000

Table 23: Final MLM Reliability Estimates

Random level-1 coefficient	Reliability estimate[39]
INTRCPT1, β_0	0.885
AAT, β_1	0.588
PRECIP_D, β_2	0.243
TOT_CLOU, β_3	0.450

[37] AAT, PRECIP_D and TOT_CLOU have been centered around the grand mean.
[38] WALKSCOR, W_WALKSC and W_L10C_C have been centered around the grand mean.
[39] The reliability estimates were based on 1,309 of 2,891 sampling locations that had sufficient data.

Table 24: Final estimation of fixed effects (with robust standard errors)[40]

Fixed Effect	Coefficient	Standard error	*t*-ratio	Approx. *d.f.*	*p*-value
For INTRCPT1, β_0					
INTRCPT2, γ_{00}	1.131033	0.006973	162.209	2887	<0.001
WALKSCOR, γ_{01}	0.011445	0.000802	14.263	2887	<0.001
W_WALKSC, γ_{02}	-0.008188	0.001170	-6.997	2887	<0.001
W_LG10_C, γ_{03}	0.837833	0.032555	25.736	2887	<0.001
For AAT slope, β_1					
INTRCPT2, γ_{10}	0.001276	0.000294	4.342	2889	<0.001
WALKSCOR, γ_{11}	0.000068	0.000010	6.935	2889	<0.001
For PRECIP_D slope, β_2					
INTRCPT2, γ_{20}	-0.097400	0.005932	-16.421	2890	<0.001
For TOT_CLOU slope, β_3					
INTRCPT2, γ_{30}	-0.000162	0.000053	-3.087	2889	0.002
WALKSCOR, γ_{31}	0.000006	0.000002	3.131	2889	0.002

Table 25: Final estimation of variance components[41]

Random Effect	Standard Deviation	Variance Component	*DFE*	χ^2	*p*-value
INTRCPT1, u_0	0.35422	0.12547	1305	62550.06485	<0.001
AAT slope, u_1	0.01045	0.00011	1307	4346.64865	<0.001
PRECIP_D slope, u_2	0.11084	0.01229	1308	1814.12703	<0.001
TOT_CLOU slope, u_3	0.00150	0.00000	1307	2670.26532	<0.001
level-1, r	0.29629	0.08779			

Figure 13c indicated that there was evidence of heterogenous variance of the residuals at level-1 (fan shape). A Breusch-Pagan test of the level-1 fitted values against the residuals was found to be significant (i.e. violation of the null hypothesis, whereby BP = 572.61, df = 1, p-value < 2.2e-16). The rule-of-thumb for an acceptable level of heteroscedasticity is a ratio of less than 1:3 in the range of variance in the residuals across the range of fitted values (Tabachnick & Fidell,

[40] Fixed effects and variance components were based on all data.
[41] Note: The chi-square statistics were based on 1,309 of 2,891 sampling locations that had sufficient data.

2007). Based on this rule-of-thumb, the observed heterogeneity in the residuals would be considered acceptable.

Figure 13: MLM Residuals

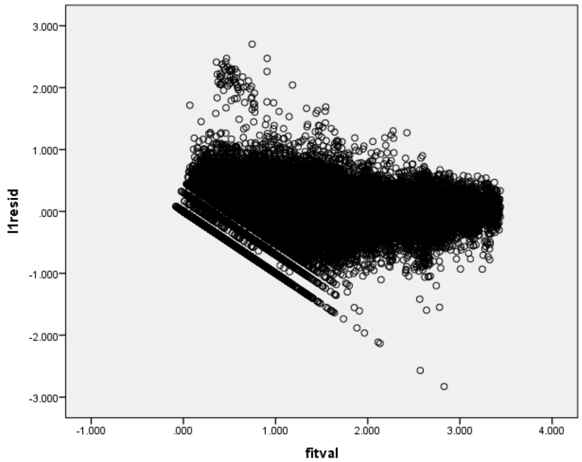


Figure 13c: Level-1 Fitted Value vs Empirical Bayes residual

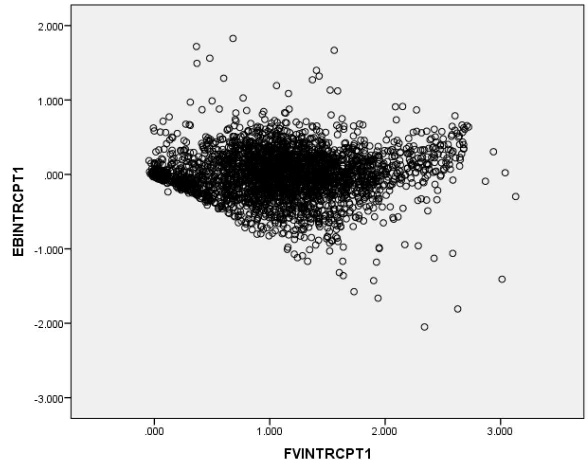


Figure 13b: Level-2 fitted value vs Empirical Bayes residual

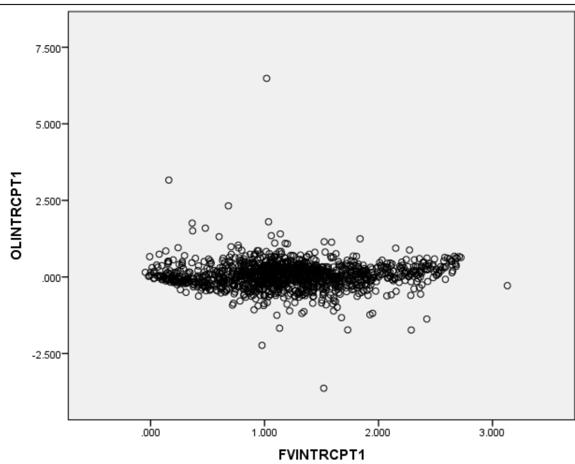


Figure 13c: Level-2 fitted value vs OLS residual

OLS level-2 residuals were somewhat non-linear, as shown in Figure 13c. Level-1 and level-2 residuals of the model were normally distributed per Figure 14a and Figure 14b) and Table 26. An outlier analysis was completed by calculating the z-scores on the level-2 OLS residual values. These potential outliers were evident on the residual scatter plot in Figure 13c, and to a lesser degree on the empirical Bayes version of that plot in Figure 13b. Any z-scores with magnitude exceeding 3.29[42] were treated as potential outliers. Ten sampling locations[43] were selected in this fashion, representing 140 level-1 observations. The LTPC fitted values were then compared, and the frequency distribution of those differences is shown in Figure 15. The

[42] Corresponds to .1% of a z-distribution, which is a recommended cutoff for large datasets
[43] Locations were geographically spread out in East Downtown, Westboro / Hintonburg, Orleans and Barrhaven suburbs.

maximum difference in fitted value observed was 0.0145, which led us to conclude that removal of the outliers had a minimal impact on the model. Furthermore, we had no conceptual explanation for their removal, therefore the outlier values were kept in the model. Based on the above diagnostics, the selected model was considered acceptable.

Figure 14: Residual Frequency Distributions

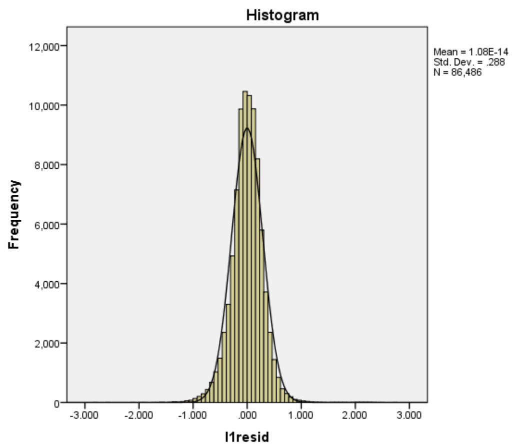


Figure 14a: Frequency Distribution of Level-1 Residuals

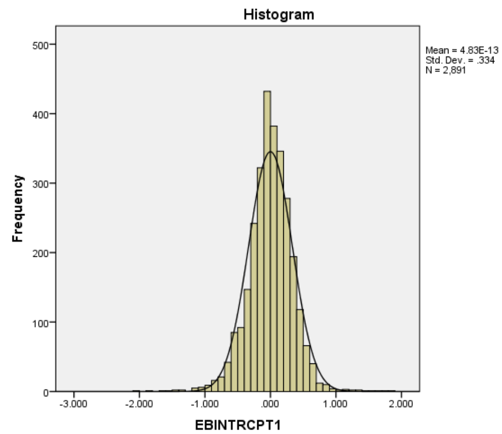


Figure 14b: Frequency Distribution of Level-2 Residuals (Empirical Bayes Residuals)

Table 26: Final MLM Residual Properties

Parameter	Level-1 Residuals	Level-2 Residuals
N	86,486	2,891
Std. Dev	.287534	.333946
Variance	.083	.112
Skew	.060	-.192
Kurtosis	3.887	2.857

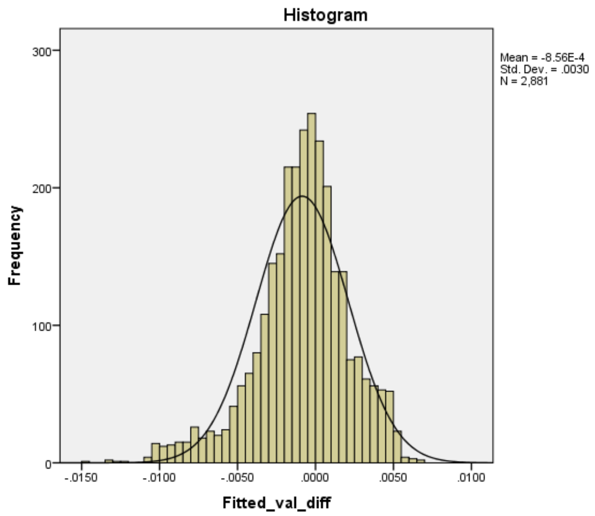


Figure 15: Frequency Distribution of Difference Between Fitted Values of Final Model, With or Without Outliers

The mixed model's final form, inserting coefficient values, is provided as follows:

Equation 6: Final Mixed Model[44]

$$
\begin{aligned}
&LG10\_CNT_{ij} = 1.131033 + 0.011445*WALKSCOR_j - \\
&0.008188*W\_WALKSC_j + 0.837833*W\_LG10\_C_j \\
&\quad + 0.001276*AAT_{ij} + 0.000068*WALKSCOR_j*AAT_{ij} \\
&\quad - 0.097400*PRECIP\_D_{ij} \\
&\quad - 0.000162*TOT\_CLOU_{ij} + 0.000006*WALKSCOR_j*TOT\_CLOU_{ij}
\end{aligned}
$$

We can observe from Table 24, that the WS-lag coefficient was unexpectedly negative, with a small variance explained of 0.6% per Table 25. This may have been the result of a suppression effect that was idiosyncratic to the data given the significant influence of nearby LTPC values. Table 27 below summarizes the quantitative effect on hourly pedestrian volume at a sampling location (10 raised to the power of LTPC) associated with unit changes for each of the individual explanatory variables based on the mixed model presented in Equation 6, excluding cross-level interactions, when all variables are grand-mean centered (i.e. when examining the slope of a given IV at the grand means of all the other IVs). For example, for every 10 units increase in WS, at a given sampling location, there was an increase in hourly pedestrian volume (not log10) of 1.31.

From Table 27, nearby LTPC (LTPC-lag) values had the most influence on LTPC at any given sampling location. An LTPC-lag of 1 (equivalent to 10 pedestrians per hour) yielded an increase in hourly pedestrian volume of 6.89.

[44] The model constant and coefficients are shown with the maximum precision provided by HLM

Table 27: Final MLM Pedestrian Volume Change by individual IV Unit Change

Explanatory Variable	Unit Change	Unit (Range)	Hourly Pedestrian Volume Change at Sampling Location
Walk Score	10	Walk Score (0 to 100)	1.31
WS-lag	10	WS (0 to 100)	-0.83
LTPC-lag	1[45]	LTPC (0 to 3.3976)	6.89
AAT	10	°C (-35.46 to 38.34)	1.03
Precipitation Dichotomous	1	1 = precip., 0 = none	-0.80
Total Cloud	10	Percent (0 to 100)	-1.00

Reliance on the coefficients of WS, WS-lag, and LTPC-lag in the final model can be misleading, however. As in any multiple regression, the variance explained by a given predictor in this MLM only represented the *incremental* contribution of that predictor *after controlling for* all the other predictors, and it excluded all variance that the predictor shared with other predictors. Thus, if there was a lot of multi-collinearity between predictors (as was the case between WS, WS-lag, and LTPC-lag), then a large percentage of the explained variance was ignored, except when looking at the variance explained by all the predictors as a set, which *did* include overlapping variance among predictors.

The high multicollinearity in this case was not an "error" that should be eliminated, but rather represented the true nature of the data: pedestrian traffic at a given sampling location (LTPC) *should have* related substantially to pedestrian traffic at nearby sampling locations (LTPC-lag) – this is exactly what is meant by the spillover/diffusion effect; pedestrian traffic at a given sampling location (LTPC) *should have* related substantially to the walkability of that sampling location (WS) – this was the core hypothesis we were testing; and by extension, pedestrian traffic at nearby sampling locations (LTPC-lag) *should have* similarly related substantially to the walkability at nearby sampling locations (WS-lag), assuming the core hypothesis was correct.

Because of the confusion that can be created by examining only the coefficients of the Level 2 predictors, it was helpful to view the relationship of these three predictors and LTPC as a 4-set Venn Diagram, as shown in Figure 16. This diagram allowed us to view not only the unique variance explained by the predictors, but also the shared variance explained by them. Where possible, the percentage of a variable's variance represented by a given region of the Venn Diagram (e.g., A, AB, AC) was obtained by subtraction of percent of LTPC variance explained by all predictors in models B through H in Table 21. The percentage of variance represented by the remaining regions was obtained by subtraction of squared zero-order correlations between the Level 2 predictors (i.e. bivariate correlations or linear regression).

[45] Equivalent to $10^1 = 10$ pedestrians

Examining the Venn diagram, we see that WS, WS-lag, and LTPC-lag represented *increments in LTPC variance explained* of only 2.0% (region AB), .6% (region AD), and 11.4% (region AC), respectively. This is what was potentially misleading about the coefficients in the final model – they only represented the *incremental* contributions of the predictors. For example, once we knew how many pedestrians were passing through nearby sampling locations (LTPC-lag) and how walkable those nearby sampling locations were (WS-lag), the walkability features of a target sampling location (WS, e.g., proximity of a grocery store at *that* sampling location) only explained an additional 2.0% of the variance in pedestrian volume at that sampling location – from this perspective, we were assigning the "credit" for regions ABC, ABD, and ABCD to the predictors WS-lag and/or LTPC-lag. However, it was equally accurate to say that, once we knew how walkable a target sampling location was (WS) and how walkable nearby sampling locations were (WS-lag), then the sheer volume of nearby pedestrian traffic (LTPC-lag) explained only an additional 11.4% of the variance in pedestrian volume at that sampling location – from this perspective, we were assigning the "credit" for regions ABC, ACD, and ABCD to the predictors WS and/or WS-lag. This was because the portions of explained variance in LTPC where predictors overlap with each other represented an indeterminate blend of influences – most notably, the large percentage of LTPC represented by region ABCD, 43.2%, could be attributed to WS, WS-lag, and/or LTPC-lag.

What was especially problematic in this spatial lag model was that the control variable (LTPC-lag) was likely influenced by the two predictors of direct interest (WS-lag and WS). Thus, a portion of the variance in LTPC-lag was not merely "error" variance to be ignored, but rather "signal" variance that was of direct interest. In other words, LTPC-lag partly served as the vehicle by which the predictor of interest (represented by WS and WS-lag) influenced LTPC.

Thus, perhaps a good way to estimate the strength of relationship between Walk Score and pedestrian volume would be as follows: We could introduce a variable called 'Combined-LTPC', which is the disjunction (i.e., combined area) of LTPC (pedestrian volume at a given sampling location) and LTPC-lag (pedestrian volume at nearby sampling locations), represented by outline 'E' in Figure 16; similarly, 'Combined-WS' , represented by outline 'F', could be conceptualized as the disjunction of WS and WS-lag. We could then compute the percentage of variance in Combined-LTPC (region 'E') that was explained by Combined-WS (region 'F'), which was 51.45% ('F' divided by 'E'). Following this approach, and assuming combined-WS caused combined-LTPC, but not vice versa, we could use causal language and say that combined-WS explained 51.45% of the variance in combined-LTPC. Hence, from this perspective, despite the relatively small incremental contribution of combined-WS on LTPC in our model, it was apparent that the variance explained by LTPC-lag was co-mingled with combined-WS, and thus LTPC-lag could not be treated as entirely error variance. This circularity cannot be overcome with the current modeling approach.

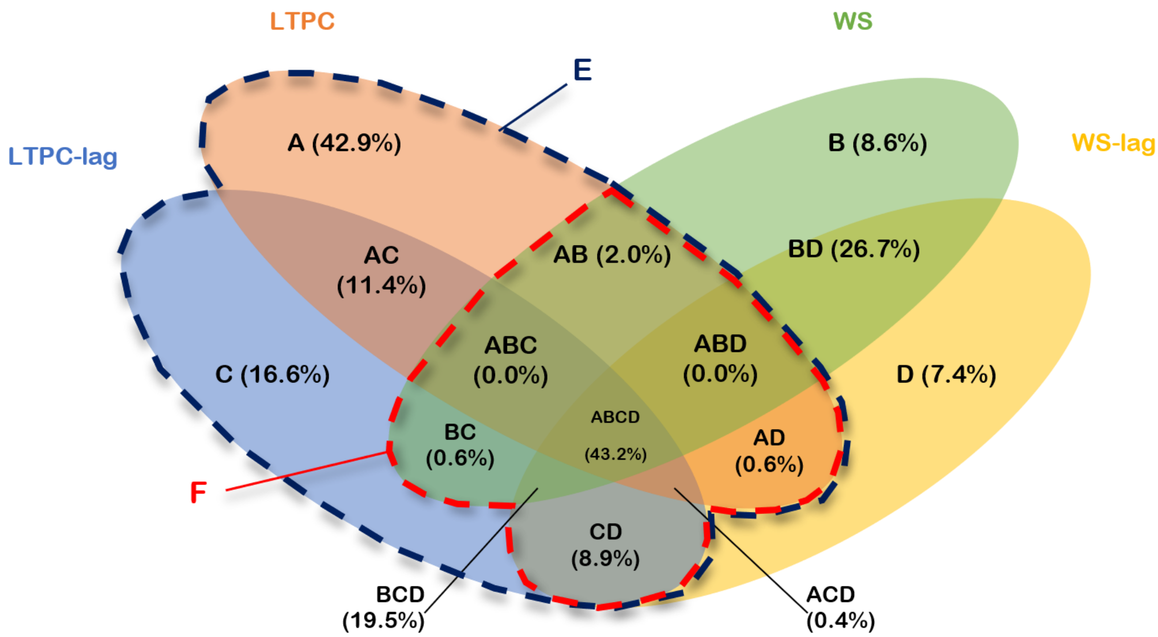


Figure 16: Venn Diagram of Variance Explained Overlap

Chapter 5: Conclusion

Discussion

The major findings of this study can be summarized as follows:

- City-wide, approximately 3% of the variance in pedestrian volume can be explained by weather (AAT[46], precipitation and cloud cover). At any given sampling location, weather accounted for approximately 12% of the variance explained (group-mean centering). Walking increased with AAT, decreased with precipitation, and decreased with cloud cover.
- Zero-order correlations suggest that both season and time (time of day and week) may play an important explanatory role, but that the structure and quality of the secondary data used in this study rendered [47]MLM analysis of these variables unfeasible.
- Pedestrian volume in Ottawa, and in likely most, if not all, urban centres, was characterized by an inherently large spill-over/diffusion effect. Our model showed that LTPC[48]-lag[49], WS[50] and WS-lag together explained approximately 57% of the variance in pedestrian volume, while 43% of the variance explaining pedestrian volume was an inextricable amalgam of all three variables. The 'combined-WS' explained approximately 51% of the variance in the 'combined-LTPC'. Consequently, we can conclude that 43% of the variance in pedestrian volume at any given location can be predicted with either LTPC-lag, WS or WS-lag, and approximately 46% of the variance in pedestrian volume at a given location (LTPC) can be explained by the 'combined-WS'.

In terms of practical application, Equation 7 will explain approximately 48% (45% by WS + 3% by weather) of the variance in pedestrian volume. In Equation 7, the LTPC-lag and WS-lag terms are removed, and the right-hand-side of the equation becomes the exponential of 10 to express pedestrian volume in hourly counts.

Equation 7: Simplified Mixed Model

$$HOURLY\ PEDESTRIAN\ VOLUME_{ij} = 10\text{^}(1.131033 + 0.011445*WALKSCOR_j + 0.001276*AAT_{ij} + 0.000068*WALKSCOR_j*AAT_{ij} - 0.097400*PRECIP\_D_{ij} - 0.000162*TOT\_CLOU_{ij} + 0.000006*WALKSCOR_j*TOT\_CLOU_{ij})$$

Where *HOURLY PEDESTRIAN VOLUME* is in pedestrian per hour (ped/h), WS is in %, AAT is in ºC, precipitation is 1 or 0, and total cloud cover is in %.

Because of the inherent uncertainty associated with this simplified mixed model's outcomes and the constraints on the variance it can explain, it is recommended that the more conservative estimates of pedestrian volume generated via Equation 7, under conditions of maximum Walk Score (100), zero precipitation, zero cloud cover and maximum AAT (+38.34ºC[51]), be treated as *minimum* values to estimate pedestrian LOS (level of service) requirements for design and planning purposes. In areas with moderate Walk Score values[52], someone seeking to estimate a

[46] AAT = Australian Apparent Temperature
[47] MLM = Multi-Level Model
[48] LTPC = Log-Transformed Pedestrian Count/Volume
[49] -lag = Spatial Lag
[50] WS = Walk Score®
[51] Although pedestrian volume appears to peak at approximately 36ºC
[52] In the case of our sampling locations in Ottawa, mean WS is approximately 60.

less conservative LOS could use the peak WS value within 1km of the location of interest in lieu of the WS value at the location of interest, since 72% of walking trips are within 1 km (City of Ottawa, 2013). One must also keep in mind that the volumes represent all flow directions at a given location. Peak volumes (>384 ped/h) as well as minimum volumes (<9 ped/h) are not captured by the model. A potential initial application of the simplified mixed model would be to estimate a minimum LOS for all pedestrian facilities in Ottawa to screen for major gaps in pedestrian LOS. More detailed validation field assessments could then be completed more cost effectively where specifically warranted.

A number of methodological improvements have been identified such that future study of pedestrian volume and walkability indices could yield even more representative models of their relationships. There was some evidence via random sub-sampling of locations using a minimum distance criterion that WS may have accounted for a greater individual share of the variance on pedestrian volume than is represented by the present model. This observation indicates that an improved specification of spatial dependencies could yield less co-mingled results. As discussed earlier, eigenvector spatial filtering (ESF) is not computationally feasible with large spatial weights datasets. One possibility to overcome this would be to conduct subsampling analysis (a resampling technique) of the data and perform ESF as part of the MLM on each subset that would contain less than 600 or so locations by design. A drawback of this design would be a loss of data variability given the apparent biases in the dataset. Another option could be to employ a form of ESF based on the promising approach developed by Murakami (Murakami & Griffith, 2017) to address the problem of extremely long computation time for large n with conventional ESF.

Other walkability indices that incorporate additional street-level factors into their measures (e.g. lighting, safety, aesthetics) could relate better to pedestrian volume. A comparative analysis of multiple walkability indices using the current study's methodology could help to identify those indices that explain the most variance in walking and serve as validation tools for those walkability indices. Moreover, variations on walkability indices could also be developed to specifically represent different pedestrian intents, such that they are more strongly associated with specific times of the day or week in accordance with this study's original hypothesis.

Apparent biases in the pedestrian volume dataset reduced our ability to perform MLM on all variables that may have been relevant based on bulk zero-order correlations, including season and time category. By having access to these variables for multi-level modeling, additional signal in the pedestrian volume data may be revealed. To realize this potential, a more complete longitudinal record of pedestrian volume that is unbiased in time, and for a sufficiently high number of spatially distributed sampling locations (>120), would be required. Acquiring this data would involve significant funding and/or active collaboration with the municipality to implement continuous and automatic counting systems (e.g. Miovision), for a time span of approximately 1 to 2 years to capture seasonal variation, at approximately 120 or more locations. Spatially and temporally concurrent, and more frequent, dedicated weather data (e.g. 0.25-hour frequency), including measures of WBGT[53], could be acquired to more accurately quantify weather as it relates to human comfort. Significantly more data is required than what was available in the present dataset to evaluate the impact of extreme conditions, such as very high or very low apparent temperatures, on walking behaviour. The larger level-1 dataset that was

[53] WBGT = Wet Bulb Globe Temperature

proposed could allow piece-wise linear models, for instance, to be generated to represent those extremes. The higher sampling frequency may also allow the study of temporal patterns in the response of pedestrians to weather changes (e.g. habituation). Such a study could also include person-level field survey data on intent or origin/destination and reported comfort response to weather conditions. Whether or not such a study is realistic is considered to be a matter of funding and buy-in from municipalities seeking facilitate an increase in their pedestrian modal share by having a thorough understanding of their pedestrian network and its utilization. A low-cost automated pedestrian traffic counting system alternative to Miovision could also significantly increase the probability of such a project coming to fruition.

Modeling of the pedestrian volume dataset in this study appears to push the limits of current and proven statistical methods due to its large size, apparent collection biases, and inherent spatial autocorrelation. An ideal methodological improvement may be to develop a spatially integrated MLM using a form of ESF having reasonable processing times, which may or may not be feasible in the short term. An altogether different approach could be to use deep learning, or neural networks, to develop a working model. However, by their very nature and current state of development, artificial-intelligence (AI)-based models do not provide an explanation of their inner workings, parameters and assumptions, and do not provide a general analytical solution that can be independently applied. Any subsequent predictions must be completed with the very same machine learning model, without the user understanding exactly how the solution is derived. This would nonetheless be an interesting approach to validate the predictability or generalizability of statistical models, or vice-versa.

Conclusion

Our primary hypothesis was that Walk Score® is positively related to pedestrian volume. We have found through MLM that this positive relationship exists and is significant, whereby 'combined-WS' explained approximately 51% of the variance in the 'combined-LTPC'.

As hypothesized, a temperature increase (expressed as 'felt' temperature, so accounting for wind speed and humidity) was indeed significantly and positively related to pedestrian volume, while a precipitation increase was negatively correlated with pedestrian volume. We have also found that cloud cover was negatively correlated with pedestrian volume. Together these three weather variables explained approximately 3% of the variance in pedestrian volume.

As hypothesized, preliminary evidence (as statistically significant bulk correlations) suggested that winter conditions (i.e. season) led to a decrease in pedestrian volume ($R=.057$, $p<.001$)). However, biases in the dataset precluded inclusion of this variable in the MLM. We also observed potential interactions between winter conditions and 'felt' temperature, subject to the same limitations.

Preliminary evidence also suggested that time of day and week, as a proxy of utilitarian and leisure walking trips, had a positive moderating effect on the relationship of 'felt' temperature, precipitation, and winter conditions with pedestrian volume. This aligns with our original hypothesis that adverse weather conditions will have more impact on walking when pedestrians are there by choice, though further research is needed to confirm these preliminary results.

This study demonstrates quantitatively that walkability and walking are significantly linked concepts. The modeling approach described herein can likely be successfully applied to other municipal settings, and ultimately other geographic/cultural/physical factors may be identified that permit creation of a general model of the relationship between walkability and walking within nations, or across borders. Yet, there is potential for significant improvement in our methodology and quantitative understanding of this relationship.

References

AASHTO. (2004). *AASHTO Guide for the Planning, Design, and Operation of Pedestrian Facilities.* Washington, DC: AASHTO.

Adams M.A., R. S. (2009). Validation of the neighborhood environment walkability scale (NEWS) items using geographic information systems. *Journal of Physical Activity and Health*, 6 (SUPPL. 1), pp. S113-S123.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Boston: Kluwer Academic Publishers.

Anselin, L., & Rey, S. J. (2014). *Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL.* Chicago: GeoDa Press LLC.

Anselin, L., Syabri, I., & Kho, Y. (2006). GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis*, 38 (1), 5-22.

Aultman-Hall, L., Lane, D., & Lambert, R. R. (2009). Assessing Impact of Weather and Season on Pedestrian Traffic Volumes. *Transportation Research Record: Journal of the Transportation Research Board*, pp. 35–43.

Badland H., W. M.-C. (2013). Using simple agent-based modeling to inform and enhance neighborhood walkability. *International Journal of Health Geographics*, 12 (58).

Bjornstad, O. N., & Falck, W. (2001). Nonparametric Spatial Covariance Functions: Estimation and Testing. *Environmental and Ecological Statistics*, 8:53–70.

Błażejczyk, K., Epstein, Y., Jendritzky, G., Staiger, H., & Tinz, B. (2012). Comparison of UTCI to selected thermal indices. *Int. J. Biometeor.*, 56, 515–535.

Bröde, P., Krüger, E. L., & Fiala, D. (2013). UTCI: Validation and Practical Application to the Assessment of Urban Outdoor Thermal Comfort. *Geographia Polonica*, 86, 1, pp. 11-20.

Castillo, J. M. (2010). *Relative Humidity.* New York: Nova Science Publishers.

Chapman, D., Nilsson, K., Larsson, A., & Rizzo, A. (2017). Climatic barriers to soft-mobility in winter: Luleå, Sweden as case study. *Sustainable Cities and Society*, 35, 574–580.

Chen, D. R., & Wen, T. H. (2010). Elucidating the changing socio-spatial dynamics of neighborhood effects on adult obesity risk in Taiwan from 2001 to 2005. *Health & Place*, 16:1248–1258.

City of New York. (2006). *New York City Pedestrian Level of Service Study, Phase I.* New York, NY: NYC Department of City Planning.

City of Ottawa. (2011, December). *City of Ottawa Open Data Beta (SHP)*. Retrieved from City of Ottawa: http://www.ottawa.ca/online_services/opendata/info/index_en.html

City of Ottawa. (2013). *City of Ottawa Annual Report 2013.* Ottawa: City of Ottawa.

City of Ottawa. (2013). *Ottawa Pedestrian Plan [DRAFT], October.* Ottawa, ON: City of Ottawa.

City of Ottawa. (2015, 08 30). *geoOttawa*. Retrieved from Ottawa: http://maps.ottawa.ca/geoOttawa/

City of Ottawa. (2015). *Open Data Ottawa*. Retrieved 02 28, 2015, from Ottawa: http://ottawa.ca/en/mobile-apps-and-open-data/open-data-ottawa

City of Toronto. (2009). *Toronto Walking Strategy.* Toronto: City of Toronto.

Clarke, P., Jana, H. A., Melendez, R., Winters, M., Gould, J. S., Ashe, M., . . . McKay, H. (2017). Snow and Rain Modify Neighbourhood Walkability for Older Adults. *Canadian Journal on Aging*, 36 (2) : 159–169.

Clifton, K. J., Smith, L. A., & Rodriguez, D. (2007). The development and testing of an audit for the pedestrian environment. *Landscape Urban Planning*, 80(1–2):95–110.

Craig, E. K., & Tofighi, D. (2007). Centering Predictor Variables in Cross-Sectional Multilevel Models: A New Look at an Old Issue. *Psychological Methods*, Vol. 12, No. 2, 121–138.

Crooks A, C. C. (2008). Key challenges in agent-based modelling for geospatial simulation. *Comp Enviro Urban Syst*, 32:417–430.

de Montigny, L., Ling, R., & Zacharias, J. (2012). The Effects of Weather on Walking Rates in Nine Cities. *Environment and Behavior*, 44(6) 821–840.

Dhanani, A., Tarkhanyan, L., & Vaughan, L. (2017). Estimating pedestrian demand for active transport evaluation and planning. *Transportation Research Part A: Policy and Practice*, 103, 54–69.

Downtown Halifax Business Commission. (2015). *Downtown Core Pedestrian Foot Traffic Counts*. Retrieved 02 17, 2015, from Downtown Halifax: http://downtownhalifax.ca/index.php/doingbusiness/downtown-core-pedestrian-foot-traffic-counts1/

Dredger S.M., K. A. (2007). Using participatory design to develop (public) health decision support systems through GIS. *International Journal of Health Geographics*, 6 (53).

Duncan, D. T., Aldstadt, J., Whalen, J., & Melly, S. J. (2013). Validation of Walk Scores and Transit Scores for estimating neighborhood walkability and transit availability: a small-area analysis. *GeoJournal*, 78:407–416.

Eco-counter. (2014). *Solutions: General Specifications*. Retrieved 02 17, 2015, from Eco-counter: http://www.eco-compteur.com/Solutions.html?wpid=15031

Eliasson, I. K. (2007). Climate and behavior in a Nordic city. *Landscape and Urban Planning*, 82, 72-84.

Environment and Climate Change Canada. (2015, 02 11). *Engineering Climate Datasets*. Retrieved 02 23, 2015, from Climate: http://climate.weather.gc.ca/prods_servs/engineering_e.html

Environment and Climate Change Canada. (2017, 04 28). *Environment and natural resources - Weather, Climate and Hazard - Past weather and climate - About the Data - Glossary - Humidex, Wind Chill*. Retrieved from Government of Canada: http://climate.weather.gc.ca/glossary_e.html#w

Environment and Climate Change Canada. (2017, 04 28). *Environment and natural resources - Weather, Climate and Hazard - Past weather and climate - About the Data - Glossary - Wind Chill*. Retrieved from Government of Canada: http://climate.weather.gc.ca/glossary_e.html#w

Environment and Climate Change Canada. (2017, 04 28). *Government of Canada*. Retrieved from Environment and natural resources - Weather, Climate and Hazard - Past weather and climate - About the Data - Glossary - Humidex: http://climate.weather.gc.ca/glossary_e.html#h

European Commission. (2015). *Cities (Urban Audit)\Data\Database*. Retrieved 03 31, 2015, from Eurostat: http://ec.europa.eu/eurostat/web/cities/data/database

Ewing, R. (2005). Can the Physical Environment Determine Physical Activity Levels? *Exercise and Sport Sciences Reviews*, Vol. 33, No. 2, pp.69-75.

Frank, L. D., Sallis, J. F., Saelens, B. E., Leary, L., Cain, K., Conway, T. L., & Hess, P. M. (2010). The development of a walkability index: application to the Neighborhood Quality of Life Study. *British Journal of Sports Medicine*, 44:924–933.

Furuhashi, R., & Yamada, K. (2013). Document Estimation of street crossing intention from a pedestrian's posture using multiple image frames. *IEEJ Transactions on Electronics, Information and Systems*, 133 (5), pp. 1069-1075+22.

Gatley, D. P. (2013). *Understanding Psychrometrics, Third Edition.* Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers.

Gehl, J. G. (1996). *Public spaces and public life.* Copenhagen: Danish Architectural Press.

Geocoder.ca. (2015). *Home Page*. Retrieved 02 25, 2015, from Geocoder.ca: http://geocoder.ca/

Google. (2014, June 1). Google Earth. Ottawa, Ontario, Canada.

Greenberg, J. A., Rueda, C., Hestir, E. L., Santos, M. J., & Ustin, S. L. (2010). Least cost distance analysis for spatial interpolation. *Computers & Geosciences*, 272–276.

Griffith, D. A. (2000). A linear regression solution to the spatial autocorrelation. *J Geogr Syst*, 2:141–156.

Griffith, D., & Chun, Y. (2014). Spatial Autocorrelation and Spatial Filtering. In M. M. Fischer, & P. Nijkamp, *Handbook of Regional Science* (pp. Chapter 75, pp 1477-1507). Vienna: Springer Reference.

Hajna S., D. K. (2013). Neighborhood walkability: Field validation of geographic information system measures. *American Journal of Preventive Medicine*, 44 (6), pp. e51-e55.

Hajrasouliha, A., & Yin, L. (2015). The impact of street network connectivity on pedestrian volume. *Urban Studies*, Vol. 52(13) 2483–2497.

Hilbe, J. M. (2014). *Modeling Count Data.* New York: Cambridge University Press.

Hillier, B., & Iida, S. (2005). Network and psychological effects in urban movement. *Lecture Notes in Computer Science 3693 (international conference on spatial information theory, COSIT 2005) (eds Cohn AG and Mark DM), Ellicottville, NY, USA, 14–18 September* (pp. pp. 475–490). Berlin: Springer-Verlag.

Ipsos Reid. (2013). *Walking Habits and Attitudes Report, City of Toronto.* Toronto: Ipsos Reid.

ITE. (2010). *Design of Walkable Urban Thoroughfares: A Context Sensitive Approach.* Washington DC: Institute of Transportation Engineers.

Klügl F, R. G. (2007). Large-scale agent-based pedestrian simulation. *Proceedings of Multi-agent systems technologies, Leipzig, Springer*.

Knez, I., Thorsson, S., Eliasson, I., & Lindberg, F. (2009). Psychological mechanisms in outdoor place and weather assessment: towards a conceptual model. *International Journal of Biometeorology*, Volume 53 (Issue1) pp.101-111.

Lafontaine, S. J., Sawada, M., & Kristjansson, E. (2017). A direct observation method for auditing large urban centers using stratified sampling, mobile GIS technology and virtual environments. *International Journal of Health Geographics*, 16(1), 6.

Lee, C., & Moudon, A. V. (2006). Correlates of Walking for Transportation or Recreation Purposes. *Journal of Physical Activity and Health*, 3, Suppl 1, S77-S98.

Lee, S., & Talen, E. (2014). Measuring Walkability: A Note on Auditing Methods. *Journal of Urban Design*, Vol. 19, No. 3, 368–388.

Lindelöw, D., Svensson, Å., Sternudd, C., & Johansson, M. (2014). What limits the pedestrian? Exploring perceptions of walking in the built environment and in the context of every-day life. *Journal of Transport & Health*, 1, 223–231.

Maghelal, P. K., & Capp, C. J. (2011). Walkability: A Review of Existing Pedestrian Indices. *Journal of the Urban and Regional Information System Association*, 23 (2): 5–19.

Maponics LLC. (2015). *Context Walkability™*. Retrieved 02 18, 2015, from Maponics: http://www.maponics.com/products/context/walkability

Maras, I., Schmidt, T., Paas, B., Ziefle, M., & Schneider, C. (2016). The impact of human-biometeorological factors on perceived thermal comfort in urban public places. *Meteorologische Zeitschrift*, 25.

Michelson, W., & Lachapelle, U. (2016). Patterns of Walking Among Employed, Urban Canadians: Variations by Commuting Mode, Time of Day, and Days of the Week. *Applied Research Quality Life*, 11:1321–1340.

Miovision Technologies Inc. (2014). *We turn video into traffic data*. Retrieved 02 26, 2015, from Miovision - rethink traffic: http://miovision.com/

Morenoff, J. D. (2003). Neighborhood mechanisms and the spatial dynamics of birth weight. *Am J Sociol*, 108:976–1017.

Morenoff, J. D. (2003). Neighborhood Mechanisms and the Spatial Dynamics of Birth Weight. *American Journal of Sociology*, v.108, n.5, p.976-1017.

Morenoff, J. D., Sampson, R. J., & Raudenbush, S. W. (2001). Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. *Criminology*, 39 (3), 517–558.

Moudon, A. V., Lee, C., Cheadle, A. D., Garvin, C., Johnson, D., Schmid, T. L., . . . Lin, L. (2006). Operational Definitions of Walkable Neighborhood: Theoretical and Empirical Insights. *Journal of Physical Activity and Health*, 3, Suppl 1, S99-S117.

MTO. (2010). *Ontario Traffic Manual - Book 15: Pedestrian Crossing Facilities.* Toronto: Queen's Printer for Ontario.

MTO. (2012). *Ontario Traffic Manual - Book 12: Traffic Signals.* Toronto: Queen's Printer for Ontario.

Murakami, D., & Griffith, D. A. (2017). Eigenvector spatial filtering for large data sets: fixed and random effects approaches. *Geographical Analysis*, doi: 10.1111/gean.12156.

National Oceanic and Atmospheric Administration. (2017, 2 19). *Solar Calculation Details*. Retrieved from U.S. Department of Commerce, National Oceanic & Atmospheric Administration, NOAA Research, Earth System Research Laboratory, Global Monitoring Division, Global Radiation Group: https://www.esrl.noaa.gov/gmd/grad/solcalc/calcdetails.html

National Weather Service. (2017, 11 16). *WetBulb Globe Temperature*. Retrieved from US Dept of Commerce, National Oceanic and Atmospheric Administration, National Weather Service: http://www.weather.gov/tsa/wbgt

Natural Resources Canada. (2009). *Annual Mean Total Precipitation.* Ottawa: Atlas of Canada 6th Edition.

NHTSA. (2004). *Walkability Checklist: How walkable is your community?* Washington DC: National Highway Traffic Safety Administration.

Ozbil, A., Peponis, J., & Stone, B. (2011). Understanding the link between street connectivity, land use and pedestrian flows. *Urban Design International*, v16,i2,125-141.

Parenteau M.-P., S. M. (2008). Development of neighborhoods to measure spatial indicators of health. *URISA Journal*, 20 (2), pp. 43-55.

Park, Y., & Kim, Y. (2014). A spatially filtered multilevel model to account for spatial dependency: application to self-rated health status in South Korea. *International Journal of Health Geographics*, 13:6.

Prince S.A., K. E.-M. (2011). A multilevel analysis of neighbourhood built and social environments and adult self-reported physical activity and body mass index in Ottawa, Canada. *International Journal of Environmental Research and Public Health*, 8 (10), pp. 3953-3978.

Prince S.A., K. E.-M. (2012). Relationships between neighborhoods, physical activity, and obesity: A multilevel analysis of a large Canadian city. *Obesity*, 20 (10), pp. 2093-2100.

PTV Group. (2015). *PTV Viswalk*. Retrieved 02 25, 2015, from PTV Group - the mind of movement: http://vision-traffic.ptvgroup.com/en-us/products/ptv-viswalk/

Pucher, J., Buehler, R., & Seinen, M. (2011). Bicycling renaissance in North America? An update and re-appraisal of cycling trends and policies. *Transportation Research Part A*, 45 (6), pp. 451-475.

Raford, N., & Ragland, D. R. (2003). Space Syntax: An Innovative Pedestrian Volume Modeling Tool for Pedestrian Safety. *Safe Transportation Research & Education Center*.

Rainham D., K. D. (2008). Development of a wearable global positioning system for place and health research. *International Journal of Health Geographics*, 7 (59).

Rainham D., M. I. (2010). Conceptualizing the healthscape: Contributions of time geography, location technologies and spatial ecology to place and health research. *Social Science and Medicine*, 70 (5), pp. 668-676.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods, Second Edition.* Thousand Oaks, CA: Sage Publications.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., Congdon, Jr., R. T., & du Toit, M. (2011). *HLM 7 Hierarchical Linear and Nonlinear Modeling.* Lincolnwood: Scientific Software International.

Riley D.L., M. A. (2013). Neighbourhood walkability and physical activity among family members of people with heart disease who participated in a randomized controlled trial of a behavioural risk reduction intervention. *Health and Place*, 21, pp. 148-155.

Rodriguez, D. B. (2009). The relationship between segment-level built environment attributes and pedestrian activity around Bogota's BRT stations. *Transportation Research Part D*, 15(1), 470–478.

Rotton, J. S. (1990). Temperature and pedestrian tempo: Walking without awareness. *Environment and Behavior*, 22, 650-674.

Sacramento Area Council of Governments. (2011). *Application of New Pedestrian Level of Service Measures.* Sacramento: Sacramento Area Council of Governments.

Savitz, N. V., & Raudenbush, S. W. (2009). Exploiting Spatial Dependence to Improve Measurement of Neighborhood Social Processes. *Sociological Methodology*, Volume 39, Issue 1, Pages 151–183.

Schasberger, M. G., Raczkowski, J., Newman, L., & Polgar, M. F. (2012). Using a Bicycle–Pedestrian Count to Assess Active Living in Downtown Wilkes-Barre. *American Journal of Preventive Medicine*, 43(5S4):S399–S402.

Shaaban, K., Muley, D., & Elnashar, D. (2017). Temporal variation in walking behavior: An empirical study. *Case Studies on Transport Policy*, 5, 671–680.

Space Syntax. (2015). *Home page*. Retrieved 02 25, 2015, from Space Syntax: http://www.spacesyntax.com/

Statistics Canada. (2013). *Commuting to work, National Household Survey (2011), NHS in Brief, Catalogue no. 99-012-2011003.* Ottawa, ON: Minister of Industry.

Statistics Canada. (2014, 06 12). *Physical activity during leisure time, 2013*. Retrieved 02 24, 2015, from Statistics Canada: http://www.statcan.gc.ca/pub/82-625-x/2014001/article/14024-eng.htm

Steadman, R. G. (1984). A Universal Scale of Apparent Temperature. *Journal of Applied Meteorology*, Vol. 23, No. 12, pp. 1674.

Steadman, R. G. (1994). Norms of Apparent Temperature in Australia. *Aust. Met. Mag.*, Vol 43, 1-16.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics, 5th Edition.* Boston: Pearson Education Inc.

TES Information Technology Ltd. (2013). *TES Information Technology Ltd.* Retrieved 03 01, 2015, from TES Traffic Engineering Software: http://www.tes.ca/overview.html

Thayn, J. B., & Simanis, J. M. (2013). Accounting for Spatial Autocorrelation in Linear Regression Models Using Spatial Filtering with Eigenvectors. *Annals of the Association of American Geographers*, 103:1, 47-66.

Tiefelsdorf, M., & Griffith, D. A. (2007). Semiparametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning A*, 39:1193–1221.

TRANS Committee. (2012). *2011 National Capital Region Travel Survey – Key Findings.* Ottawa.

TRANS Committee. (2015). *O-D Survey*. Retrieved 02 24, 2015, from TRANS: A Joint Transportation Planning Committee Serving the National Capital Region: http://www.ncr-trans-rcn.ca/surveys/o-d-survey/

Transport for London. (2010). *Pedestrian Comfort Guidance for London, Guidance Document, First Edition 2010.* London, UK: Atkins.

Transportation Research Board. (2000). *Highway capacity manual: HCM2000.* Washington D.C.: National Research Council.

USDOT. (2006). *FHWA Course on Bicycle and Pedestrian Transportation.* McLean, VA: Federal Highway Administration.

USDOT FHA. (1998). *Recommended Procedures Chapter 13, "Pedestrians," of the Highway Capacity Manual, Capacity Analysis of Pedestrian and Bicycle Facilities Task Order 8: Pedestrian-Bicycle Research Program.* McLean, VA: Office of Safety & Traffic Operations Research & Development.

USDOT FHA. (2004). *Signalized Intersections: Informational Guide, Publication No.FHWA-HRT-04-091.* McLean, VA: Turner-Fairbank Highway Research Center.

USDOT FHA. (2006). *Shared-Use Path Level of Service Calculator - A user's guide.* McLean, VA: Turner-Fairbank Highway Research Center.

USDOT FHA. (2014). *Statewide Pedestrian and Bicycle Planning Handbook (DOT-VNTSC-FHWA-14-04, FHWA-HEP-14-035).* Cambridge, MA: John A Volpe National Transportation Systems Center.

Verbitsky-Savitz, N., & Raudenbush, S. W. (2009). Exploiting spatial dependence to improve measurement of neighborhood social processes. *Sociological Methodology*, Vol. 39, No. 1, pp 151-183.

Walk Score. (2015). *Walk Score API*. Retrieved 02 19, 2015, from Walk Score Professional: https://www.walkscore.com/professional/api.php

Walk Score Advisory Group. (2017, 12 13). *Walk Score Methodology*. Retrieved from Walk Score: https://www.walkscore.com/methodology.shtml

Walk21. (2006, 10). *International Charter for Walking.* Retrieved 02 24, 2015, from Walk21: http://www.walk21.com/papers/International%20Charter%20for%20Walking.pdf

Walkonomics. (2014). *How Walkable is your Street?* Retrieved 02 25, 2015, from Walkonomics: http://www.walkonomics.com/

Wang, Y., & Kockelman, K. M. (2013). A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. *Accident Analysis and Prevention* , 60, 71-84.

Wang, Y., Kockelman, K. M., & Wang, X. (2013). Understanding Spatial Filtering for Analysis of Land Use Data Sets. *Journal of Transport Geography*, No. 31, 123-131.

Ward, M. D., & Gleditsch, K. S. (2008). Spatially Lagged Dependent Variables. In M. D. Ward, *Spatial Regression Models* (pp. pp 35-64). Los Angeles: Sage Publications.

Yang Y, D. R. (2013). Using an agent-based model to simulate children's active travel to school. *Int J Behav Nutr Phys Act*, 10: 10.1186/1479-5868-10-67.

Yang Y, e. a. (2011). A spatial agent-based model for the simulation of adults' daily walking within a city. *Am J Prev Med*, 40:353–361.

Yin, L. (2013). Assessing Walkability in the City of Buffalo: Application of Agent-Based Simulation. *Journal of Urban Planning and Development*, 139, pp. 166-175.

Zacharias, J., Stathopoulos, T., & Wu, H. (2001). Microclimate and downtown open space activity. *Environment and Behavior*, 33, 296-315.

Zacharias, J., Stathopoulos, T., & Wu, H. (2004). Spatial behavior in San Francisco's plazas: The effects of microclimate, other people and environmental design. *Environment and Behavior*, 36, 638-658.

Zhou, Y. (2008). Agent-based modeling and simulation for pedestrian movement behaviors in space: A review of applications and GIS issues. *Proceedings of SPIE - The International Society for Optical Engineering*, 7143.

Zohreh, A.-S., Moeinaddini, M., & Zaly Shah, M. (2013). Non-motorised Level of Service: Addressing Challenges in Pedestrian and Bicycle Level of Service. *Transport Reviews: A Transnational Transdisciplinary Journal*, 33:2, 166-194.