Journal of Transport Geography 93 (2021) 103067

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Varying influences of the built environment on daily and hourly pedestrian crossing volumes at signalized intersections estimated from traffic signal controller event data

Patrick A. Singleton [a,*], Keunhyun Park [b], Doo Hong Lee [b]

[a] Utah State University, Department of Civil and Environmental Engineering, 4110 Old Main Hill, Logan, UT 84322–4110, United States of America
[b] Utah State University, Department of Landscape Architecture and Environmental Planning, 4005 Old Main Hill, Logan, UT 84322–4005, United States of America

ARTICLE INFO

Keywords:
Pedestrian traffic
Pedestrian counts
Built environment
Land use
Pedestrian detectors

ABSTRACT

Direct-demand models of pedestrian volumes (identifying relationships with built environment characteristics) require pedestrian data, typically from short-duration manual counts at a limited number of locations. We overcome these limitations using a novel source of pedestrian data: estimated pedestrian crossing volumes based on push-button event data recorded in traffic signal controller logs. These continuous data allow us to study more sites (1494 signalized intersections throughout Utah, US) over a much longer time period (one year) than in previous models, including the ability to detect variations across days-of-week and times-of-day. Specifically, we develop direct demand (log-linear regression) models that represent relationships between built environment variables (calculated at ¼- and ½-mile network buffers) and annual average daily and hourly pedestrian metrics. We control spatial autocorrelation through the use of spatial error models. All results confirm theorized relationships: There is more pedestrian activity at intersections with greater population and employment densities, a larger proportion of commercial and residential land uses, more connected street networks, more nearby services and amenities, and in lower-income neighborhoods with larger households. Notably, we also find relevant day-of-week and time-of-day differences. For example, schools attract pedestrian activity, but only on weekdays during daytime hours, and the coefficient for places of worship is higher in the weekend model. K-fold cross-validation results show the predictive power of our models. Results demonstrate the value of these novel pedestrian signal data for planning purposes and offer support for built environment interventions and land use policies to encourage walkable communities.

1. Introduction

Quantifying pedestrian volumes and levels of walking activity is critical for many transportation planning, engineering, and management tasks. Traffic safety analyses require estimates of pedestrian exposure to risk, and durations and distances of physically active transportation are inputs to transportation health impact assessments. Information on walking is also useful for analyzing pedestrian level/quality of service, designing pedestrian infrastructure, and prioritizing pedestrian investments. Furthermore, there is a growing interest in creating active living and walk friendly communities in order to improve health, reduce automobile dependence, and strengthen local economies.

Pedestrian volume data can be collected. Nevertheless, traditional data collection methods for monitoring pedestrian traffic have limitations: They involve short durations, few locations, or samples of the population. Manual intersection or street segment counts are time consuming and often infeasible to conduct over long periods of time. Instruments such as infrared counters can record continuous data on trail users, but they are costly to deploy across multiple sites (Ryus et al., 2014). The passive collection of crowdsourced pedestrian data from mobile devices shows promise, but data may be non-representative and require calibration and factoring methods (StreetLight InSight, 2018). Methods have been developed to adjust short-duration counts to average pedestrian volumes using factors developed from permanent counters (FHWA, 2016), but they still usually require manual counts and are sensitive to count duration, seasonality, and factor group selection.

Alternatively, pedestrian volume data be modeled. Conventional methods of modeling roadway volumes are inappropriate for

* Corresponding author.
E-mail addresses: patrick.singleton@usu.edu (P.A. Singleton), keunhyun.park@usu.edu (K. Park), doohong.lee@usu.edu (D.H. Lee).

https://doi.org/10.1016/j.jtrangeo.2021.103067
Received 9 January 2021; Received in revised form 13 April 2021; Accepted 17 April 2021

pedestrians, due to data and scale challenges with including pedestrians in regional travel demand forecasting models (Singleton et al., 2018). Instead, planners interested in facility-specific information have turned to using direct demand models (Kuzmyak et al., 2014; Munira, 2017). Direct demand models predict pedestrian volumes using observed counts and measures of the surrounding streetscape, land uses, built environment, and street network. Such models help to understand how environmental features affect pedestrian volumes and inform transportation and land-use planning and urban design strategies to promote walkable communities. Still, direct demand models require large quantities of (pedestrian) estimation data in order to be generalizable beyond the few locations where they were developed, and they are often insensitive to temporal variations in walking activity.

The objective of this research is to examine relationships between the built environment and pedestrian activity through the development of direct demand models of pedestrian volumes, taking advantage of a novel and relatively ubiquitous (in both time and space) source of pedestrian data. Specifically, we utilize estimates of pedestrian crossing volumes—taken from pedestrian push-button activity data from high-resolution traffic signal controller logs—and apply log-linear regression models for different time periods to study nearly 1500 signalized intersections throughout Utah, US. Our study's primary contribution is the use of continuously-collected pedestrian activity data from traffic signals (measured over the course of one year, and averaged per day and per hour) for direct demand pedestrian volume modeling. Notably, this allows us to uncover some theoretically-consistent built environment relationships with walking that many other similar studies have not found, and to identify day-of-week and time-of-day variations in those relationships. In the following sections, we first summarize existing literature on built environment relationships and direct-demand models of pedestrian counts/volumes. We then describe our data and methods, present our results, and finally discuss key findings.

2. Literature review

Two general threads of research have investigated built environment correlates of pedestrian counts or volumes. One research path is motivated by developing models to predict pedestrian demand, for use in various transportation engineering, planning, and safety analysis tasks. For example, Schneider et al. (2009) describe several applications of such models: to "quantify pedestrian exposure in safety analysis," prioritize pedestrian projects, design pedestrian infrastructure, predict pedestrian volumes in the future, analyze crossings warrants, and evaluate commercial visibility (p. 13). In these studies, built environment characteristics predict pedestrian counts and are used to estimate pedestrian volumes in areas where data have not been collected. The other strand of research focuses on understanding relationships between urban design characteristics and walking activity, to inform planning and design for walkable, healthy cities. These studies often focus on measuring more detailed and complex attributes of urban form and the built environment, including the so-called "D" variables (e.g., development density, land use diversity, street network design, destination accessibility, and distance to transit) (Ewing and Cervero, 2010), urban design qualities of the streetscape (Ewing and Handy, 2009), and/or street network connectivity elements derived from Space Syntax (Hillier, 2007). A simplified characterization is that studies of the first kind focus primarily on pedestrian volumes and secondarily on built environment measures, while studies of the second kind do the opposite. Of course, some research straddles the boundaries of the two kinds (Raford and Ragland, 2006; Raford and Ragland, 2004).

Table 1 and Table 2 summarize the methods, outcomes, and predictors used in studies modeling pedestrian volumes as a function of built environment measures. Similar summaries have been presented in recent publications (Munira, 2017; Schneider et al., 2021). In this summary, we focus on studies with models of pedestrian counts or volumes, not on literature using individual- or household-based measures of walking behavior. We also exclude studies that group walk and bicycle traffic together into one non-motorized mode.

In pedestrian volume models, some built environment measures (see Table 2) are consistently related to walking in expected directions, while results for other variables are more equivocal. More often than not, studies find positive associations with residential and employment density. Walking is also closely linked to public transit: locations closer to transit stops/stations and with more transit stops nearby tend to see greater pedestrian volumes. Diversity measures like land use mix and entropy are sometimes positively related to pedestrian volumes, but studies also find insignificant or even negative relationships. More studies find null or unexpectedly negative results than positive results for traditional street network design variables like intersection density and percentage of 4-way intersections. Studies of street network configurations tend to find positive associations with space syntax measures like integration. Studies of urban design and streetscape qualities tend to find positive associations with imageability (the quality of a place that makes it distinct, recognizable and memorable) and transparency (the degree to which people can see or perceive human activity beyond the edge of a street; Park et al., 2019). A few studies have found that pedestrian volumes are significantly explained by socioeconomic and environmental variables like household size, household incomes, parks, and slope.

As shown in Table 1, most pedestrian volume direct demand models utilize manually-collected, short-duration counts of the number of people walking along street segments or crossing at intersections. Sometimes these counts are as short as 30 or even 10 min (or multiple 5-min counts), but rarely do they exceed 12 h. These short durations are not surprising, given the cost and effort of conducting manual pedestrian counts at multiple locations (Ryus et al., 2014). One exception is the one week of automated pedestrian counts conducted in Blacksburg, Virginia (Hankey et al., 2017; Lu et al., 2018). For models relating pedestrian volumes to the built environment, studying many sites is critical for both the power of the analysis (to detect statistically-significant associations) and the generalizability of results (across varied locations). Most research builds models using data from between several dozen and several hundred locations. Three exceptions are the 1018 signals in Montréal (Miranda-Moreno and Fernandes, 2011), the 1270 intersections throughout California (Griswold et al., 2019), and the nearly 10,000 street segments with pedestrian counts in Seoul, South Korea (e.g., Kim et al., 2019).

The data collection methods used to obtain pedestrian volumes for most previous research led to some limitations in the accuracy, generalizability, and sensitivity of model results. First, the use of short-duration counts to represent average or typical volumes—even when adjusted for time-of-day and weather using a smaller number of longer-duration automated counts—adds measurement error to the dependent variable. This potentially affects the value and significance of estimated associations. Second, the short time periods typically studied—often weekdays during daytime or morning/midday/evening peak hours—limits the ability of models to consider temporal variations in relationships between the built environment and pedestrian volumes. There may be interesting and policy-relevant variations by time-of-day, day-of-week (weekdays vs. weekends), and season. Third, the number of locations studied—usually less than 1000 and sometimes less than 100—can limit both the generalizability of findings as well as the statistical power to detect significant associations.

In this study, we mitigate some of these limitations by utilizing a new source of pedestrian data: estimated pedestrian crossing volumes at signalized intersections, taken from pedestrian push-button events recorded in archived high-resolution traffic signal controller logs (Sturdevant et al., 2012). Assuming a traffic signal includes walk indications and pedestrian detection (usually push-buttons), at least two relevant pedestrian events can be recorded. Event code 90 ("pedestrian detector on") occurs whenever a pedestrian push-button is activated (pressed), which could happen multiple times per cycle. Event code 45

Table 1
Summary of pedestrian volume modeling studies.

	Information			Pedestrian			Model	
Study	*Geography*	*Locations*	*Time*	*Outcome*	*Method*	*Details*	*Type*	*Fit*
Pushkarev and Zupan (1971)	Manhattan, New York City, New York, US	≤605 block faces	1969 Apr–Jun	Volume, instant	AP	Twice, WD, MD & PM	L	0.23–0.61
Behnam and Patel (1977)	Downtown Milwaukee, Wisconsin, US	? street segments	1971–1973 Sum	Volume, 1 h	MC	Multiple times 6 min, WD, DT	LL	0.58
Hillier et al. (1993)	Central London, England, UK	≤239 street segments	??	Volume	MC	20–30 times, AM & MD & PM	LL	0.29–0.57
Penn et al. (1998)	Central London, England, UK	7 street segments	??	Volume, 50 min	MC	Ten times 5 min, AM & MD & PM	CR	0.98
Qin and Ivan (2001)	Rural Connecticut, US	32 crossings	1999 May, Jun, Oct, Nov	Crossing volume	MC	Twice 9.5 h, WD & WE, DT	LL	0.81–0.91
Desyllas et al. (2003)	Central London, England, UK	231 street segments	1999 Aug, 2000 Mar, 2001 Jul	Volume, 1 h	MC	Multiple 5 min, DT	LL	0.82
Raford and Ragland (2004)	Oakland, California, US	42 intersections	??	Volume, 1 year (extrapolated)	MC	Multiple 2 h, WD & WE, AM & PM	??	0.77
Liu and Griswold (2009)	San Francisco, California, US	63 intersections	2002 May, Jun, Aug, Sep	Crossing volume	MC	Once 4 h, WD, PM	L, SA	0.75
Miranda-Moreno et al. (2011)	Montréal, Quebec, CA	519 signalized intersections	2003 Spr–Sum	Volume	MC	Three times 1 h, WD, AM & MD & PM	LL	0.55
Raford and Ragland (2006)	Boston, Massachusetts, US	82 locations	2004 Aug	Volume	MC	24 times 5 min, WD & WE, DT	??	0.79–0.86
Pulugurtha and Repaka (2013, 2008)	Charlotte, North Carolina, US	176 signalized intersections	2005	Volume, 12 h	MC	Once 12 h, DT	L	0.15–0.86
Rodríguez et al. (2009)	Bogotá, Distrito Capital, CO	338 street segments	2005 Jun–Aug	Volume, 10 min	MC	Once 10 min, WD, AM	NB	0.03
Ewing et al. (2016), Ewing and Clemente (2013)	New York City, New York, US	588 block faces	2006 Sum	Volume	MC	Four times, WD, DT	NB, SA	??
Arnold et al. (2010)	San Diego County, California, US	80 locations	2007 Jul–Aug, 2008	Volume, 2 h (adjusted)	MC	Twice 2 h, WD & WE, AM or MD or PM	LL	0.52
Hajrasouliha and Yin (2015)	Buffalo, New York, US	302 street segments	2007–2010	Volume	MC	Twice, WD, DT	L	??
Hankey et al. (2012)	Minneapolis, Minnesota, US	259 street/path segments	2007–2010 Sep	Volume, 12 h (extrapolated)	MC	2 h or 12 h, WD, PM or DT	NB	0.42
Hankey and Lindsey (2016)	Minneapolis, Minnesota, US	471 street/trail segments	2007–2014 Sep	Volume, 1 h	MC	Various 2 h, PM	LL	0.50–0.53
Tabeshian and Kattan (2014)	Calgary, Alberta, CA	34 intersections	2007–2012	Volume, 2 h	MC	Three times 2 h, AM & MD & PM	L, P	0.79–0.92
Schneider et al. (2009)	Alameda County, California, US	50 intersections	2008 Apr–Jun	Crossing volume, 1 week (extrapolated)	MC	Twice 2 h, WD & WE, AM or MD or PM	L	0.89
Miranda-Moreno and Fernandes (2011)	Montréal, Quebec, CA	1018 signalized intersections	2008–2009	Crossing volume	MC	Once 8 h, WD, AM & MD & PM	LL	0.58
Ozbil et al. (2011)	Atlanta, Georgia, US	157 locations	??	Volume	MC	20 times (or ten times 20 min), DT & PM	LL	0.82–0.84
Kang (2018, 2017, 2015), Kim et al. (2019); Kang (2017), Sung et al. (2013, 2015)	Seoul, KR	≤9850 street segments	2009 Aug–Nov	Volume	MC	Six times 14 h, WD & WE, DT	LL, SA	0.24–0.81
Schneider et al. (2012)	San Francisco, California, US	50 intersections	2009 Sep, 2010 Jul–Aug	Crossing volume, 1 year (extrapolated)	MC	Once 2 h, WD, AM or PM	LL	0.80
Ameli et al. (2015)	Downtown Salt Lake City, Utah, US	179 block faces	2012 Sep–Oct	Volume	MC	Twice 30 min, WD, MD & PM	NB	??
Maxwell (2016)	Glasgow, Scotland, UK	693 street segments	2014–2015 Sum	Volume	MC	Four times, WD, DT	NB, SA	??
Sanders et al. (2017)	Seattle, Washington, US	49 intersections	??	Volume, 1 year (extrapolated)	MC	??, PM	P	0.76
Hankey et al. (2017), Lu et al. (2018)	Blacksburg, Virginia, US	72 locations	2015 Apr–Oct	Volume, 1 day & 1 h (averaged)	AC	Once 1 wk	LL	0.71, 0.00–0.78
Park et al. (2019)	Salt Lake County, Utah, US	881 block faces	2015	Volume	MC	Four times, WD, DT	NB, SA	??
Hamidi and Moazzeni (2019)	Downtown Dallas, Texas, US	402 block faces	2016 Spr–Sum	Volume, 30 min	MC	Once 30 min, WD, PM	NB, SA	??
Le et al. (2020)	Dallas, Texas, US	196 intersections	2016	Volume 1 day (extrapolated)	MC	Once 2 h or 8 h	NB	??
Griswold et al. (2019)	California, US	1270 intersections	2006–2016	Crossing volume, 1 year (extrapolated)	MC	Various 1–86 h, most two times 2 h, AM & PM	LL	0.71
Schneider et al. (2021)	Milwaukee, Wisconsin, US	260 intersections	2013–2018	Crossing volume, 1 year (extrapolated)	MC	Various, many 13 h, AM & MD & PM	NB	??

(*continued on next page*)

Table 1 (*continued*)

	Information			*Pedestrian*			*Model*	
Study	*Geography*	*Locations*	*Time*	*Outcome*	*Method*	*Details*	*Type*	*Fit*
This study	Utah, US	1494 signalized intersections	2017 Jun – 2018 Jul	Estimated volume, 1 day & 1 h (averaged)	AC	Continuous	LL, SA	

Notes:?? = unknown.
Method: AC = automated counts, AP = aerial photos, MC = manual counts.
Details: WD = weekday, WE = weekend, AM = morning peak, MD = midday, PM = evening peak, DT = daytime.
Type: L = linear, LL = log-linear (linear with natural log transformation), CR = linear with cube-root transformation, P = Poisson, NB = negative binomial, SA = checked or corrected for spatial autocorrelation.
Fit: R^2 or pseudo-R^2.

("pedestrian call registered") occurs when a call to service a walk phase is registered, which usually happens just once per cycle for a particular phase or crossing (upon the first pedestrian detection event). In recent years, several studies have investigated the use of pedestrian signal data for different purposes, including for pedestrian volume estimation (Blanc et al., 2015; Day et al., 2011; Kothuri et al., 2017; Li and Wu, 2021; Noyce and Bentzen, 2005; Singleton and Runa, 2021). More generally, high-resolution traffic signal event data are beginning to be used in a variety of other research and operational contexts (Wu and Lui, 2014), including through Automated Traffic Signal Performance Measures (ATSPM) systems (Day et al., 2016).

To our knowledge, this is the first study to relate traffic signal-based measures of pedestrian activity with built environment characteristics. Recall the three limitations of the short-duration manual count pedestrian volume data typically used in prior built environment direct demand models: measurement error due to factoring, an inability to model temporal variations, and the small number of locations studied. Since traffic signal data are recorded continuously (24 h a day, 365 days a year), they can overcome the second limitation. The third limitation is constrained only by the number of signalized intersections with such data in an area. Regarding the first limitation, we replace the measurement error associated with factoring short-duration counts with the error due to the fact that pedestrian push-button data may not be a perfect measure of pedestrian crossing volumes. One person may press the push-button multiple times (although, only one pedestrian call would be registered), or a group of pedestrians may not press the button at all. Nevertheless, prior research looking at a couple days of data at one intersection in Oregon found correlations of around 0.80 or greater between pedestrian actuations and crossing volumes (Blanc et al., 2015; Kothuri et al., 2017). Another study looked at two mid-block crossings in Arizona over several days and estimated pedestrian crossing volumes from push-button data with a mean error of around ±2 pedestrians per hour (Li and Wu, 2021).

A recent large-scale research effort in Utah investigating the feasibility of pedestrian traffic signal data for pedestrian volume estimation found similar levels of accuracy. Singleton et al. (2020; Singleton and Runa, 2021) collected traffic signal data as well as video recordings of pedestrian crossing events at 90 randomly selected signalized intersections across Utah in 2019. Almost 175,000 pedestrians were manually counted during more than 10,000 h of video, covering different months, weekdays, and hours. The authors then developed simple non-linear (quadratic and piecewise linear) regression models predicting hourly pedestrian crossing volumes as a function of constructed measures of pedestrian signal data (pedestrian actuations, and unique pedestrian detections (removing those within 15 s of another detection)). For ease of application, the models did not include traffic volumes or neighborhood socioeconomic/environmental characteristics, although they did account for non-linear relationships between push-button use and pedestrian volumes (high vs. low pedestrian activity signal) and different traffic signal operations (phase on pedestrian recall or not, short vs. long average cycle length; HAWK signal vs. traditional signal). Over more than 22,500 crossing-hours of observations, the correlation between observed and model-predicted hourly pedestrian crossing volumes was 0.84; most models had correlations close to 0.90, and the mean error was ±3 pedestrians per hour (Singleton et al., 2020b; Singleton and Runa, 2021). Thus, these results along with other recent research (Blanc et al., 2015; Kothuri et al., 2017; Li and Wu, 2021) suggest that pedestrian signal data can be used to estimate pedestrian crossing volumes with reasonable accuracy. Based on these prior research findings, we think the tradeoff in the sources of error in the dependent variable (factoring short-duration counts vs. adjusting pedestrian push-button data) is reasonable.

3. Data and methods

3.1. Estimated pedestrian volumes from traffic signal data

The study area includes the six most populous counties in Utah, US: Salt Lake, Utah, Davis, Weber, Washington, and Cache. Cumulatively, these six counties comprise 84% of Utah's population and contain most of the roughly 2100 traffic signals in the state. Fig. 1 shows a map of the traffic signals located within the six study counties in Utah. The Utah Department of Transportation (UDOT) has helped lead the development and deployment of the ATSPM system (Day et al., 2016) through which archived traffic signal controller event logs can be accessed. As of fall 2018, UDOT was actively archiving data from more than 1900 state- and locally-owned signals in a central database (Taylor and Mackey, 2018).

Our pedestrian volume data are estimates of annual average daily pedestrian (AADP) crossing volumes at signalized intersections, derived from pedestrian activity events recorded in high-resolution traffic signal controller event logs. For this study, we obtained one year—01 July 2017 through 30 June 2018—of pedestrian data from all traffic signals in our study area. After cleaning the data to remove missing observations, we applied the pedestrian volume estimation methods developed by Singleton et al. (2020; Singleton and Runa, 2021) to the pedestrian signal data. Next, we aggregated (over hours in a day and crossings at an intersection) and averaged (over days in the year) those estimates to calculate AADP at each signal. We then removed 143 locations with effectively no pedestrian activity (less than 1 per day); the vast majority of these were signals with no pedestrian push-buttons, either in dense downtowns (where signals operated on pedestrian recall) or in isolated locations (such as highway off-ramps and industrial areas). After this process, we were left with 1494 signals for our models. AADP ranged from 1 to nearly 6700, with a median of about 110 and a mean of about 270. The distribution of AADP was positively skewed and leptokurtic. Since our data are available continuously throughout the year, we also calculated AADP for weekdays vs. weekends. In addition, we calculated the annual average hourly pedestrian (AAHP) crossing volumes for various times of day. As noted in the literature review, most studies do not collect enough data to analyze time-of-day variations, so we think our ability to model both average daily and average hourly pedestrian volumes is a relatively unique contribution. Descriptive statistics for the pedestrian volume dependent variables are shown in Table 3.

Table 2
Summary of built environment predictors of pedestrian volumes.

Variable	Dir.[a]	Studies
Density		
Floor area ratio or building density	+	(Ameli et al., 2015; Ewing et al., 2016; Ewing and Clemente, 2013; Hamidi and Moazzeni, 2019; Maxwell, 2016; Ozbil et al., 2011; Park et al., 2019; Sung et al., 2013)
	ns / –	(Ameli et al., 2015; Kim et al., 2017; Park et al., 2019; Sung et al., 2013)
Population density, household density, or residential space density	+	(Ameli et al., 2015; Arnold et al., 2010; Behnam and Patel, 1977; Ewing et al., 2016; Ewing and Clemente, 2013; Griswold et al., 2019; Hankey and Lindsey, 2016; Hankey et al., 2017; Kim et al., 2019; Liu and Griswold, 2009; Lu et al., 2018; Miranda-Moreno et al., 2011; Miranda-Moreno and Fernandes, 2011; Ozbil et al., 2011; Pulugurtha and Repaka, 2013, 2008; Raford and Ragland, 2004; Sanders et al., 2017; Schneider et al., 2009, 2012, 2021; Tabeshian and Kattan, 2014)
	ns / –	(Hajrasouliha and Yin, 2015; Hankey et al., 2012; Kang, 2017, 2015; Maxwell, 2016; Qin and Ivan, 2001; Park et al., 2019; Pulugurtha and Repaka, 2013, 2008; Rodríguez et al., 2009)
Employment density, employment access, or commercial/office/non-residential space density	+	(Arnold et al., 2010; Behnam and Patel, 1977; Griswold et al., 2019; Hajrasouliha and Yin, 2015; Hankey and Lindsey, 2016; Kang, 2017, 2015; Kim et al., 2019; Liu and Griswold, 2009; Miranda-Moreno et al., 2011; Miranda-Moreno and Fernandes, 2011; Ozbil et al., 2011; Park et al., 2019; Pulugurtha and Repaka, 2013; Pushkarev and Zupan, 1971; Raford and Ragland, 2004; Sanders et al., 2017; Schneider et al., 2009, 2012, 2021; Sung et al., 2013; Tabeshian and Kattan, 2014)
	ns / –	(Hankey et al., 2012; Park et al., 2019; Pulugurtha and Repaka, 2013, 2008; Rodríguez et al., 2009; Sung et al., 2013)
Diversity		
Land use mix, entropy, balance, or % retail	+	(Ameli et al., 2015; Ewing et al., 2016; Ewing and Clemente, 2013; Hajrasouliha and Yin, 2015; Hamidi and Moazzeni, 2019; Liu and Griswold, 2009; Park et al., 2019; Sung et al., 2013)
	ns / –	(Ameli et al., 2015; Arnold et al., 2010; Ewing et al., 2016; Ewing and Clemente, 2013; Kang, 2018, 2017, 2015; Kim et al., 2019, 2017; Maxwell, 2016; Park et al., 2019)
Transit		
Distance to nearest rail/bus stop/station	–	(Ameli et al., 2015; Ewing et al., 2016; Ewing and Clemente, 2013; Hamidi and Moazzeni, 2019; Kang, 2017, 2015; Kim et al., 2019, 2017; Maxwell, 2016; Miranda-Moreno et al., 2011; Miranda-Moreno and Fernandes, 2011; Pushkarev and Zupan, 1971; Raford and Ragland, 2006; Sung et al., 2013, 2015)
	ns / +	(Hankey et al., 2012; Park et al., 2019; Raford and Ragland, 2006; Rodríguez et al., 2009)
Transit stop density	+	(Hankey and Lindsey, 2016; Hankey et al., 2017; Liu and Griswold, 2009; Lu et al., 2018; Miranda-Moreno et al., 2011; Miranda-Moreno and Fernandes, 2011; Park et al., 2019; Pulugurtha and Repaka, 2013, 2008; Schneider et al., 2009, 2021; Sung et al., 2013; Tabeshian and Kattan, 2014)
	ns / –	(Kang, 2017, 2015; Le et al., 2020)
Street network design		
Intersection density	+	(Hajrasouliha and Yin, 2015; Hamidi and Moazzeni, 2019)
	ns / –	(Ameli et al., 2015; Ewing et al., 2016; Ewing and Clemente, 2013; Hankey and Lindsey, 2016; Hankey et al., 2017; Kang, 2018, 2017, 2015; Lu et al., 2018; Maxwell, 2016; Park et al., 2020; Sung et al., 2013)
% 4-way intersections	+	(Miranda-Moreno et al., 2011; Miranda-Moreno and Fernandes, 2011; Park et al., 2019)
	ns / –	(Ameli et al., 2015; Ewing et al., 2016; Ewing and Clemente, 2013; Maxwell, 2016; Park et al., 2019; Sung et al., 2013)
Block length	+	(Ewing et al., 2016; Ewing and Clemente, 2013; Maxwell, 2016; Miranda-Moreno et al., 2011; Miranda-Moreno and Fernandes, 2011; Park et al., 2019; Tabeshian and Kattan, 2014)
	ns / –	(Ameli et al., 2015; Hamidi and Moazzeni, 2019; Park et al., 2019)
Space syntax (integration, reach, betweenness, etc.)	+	(Hajrasouliha and Yin, 2015; Hillier et al., 1993; Kang, 2018, 2017, 2015; Ozbil et al., 2011; Penn et al., 1998; Raford and Ragland, 2006, 2004)
	ns / –	(Kang, 2017, 2015)
Socioeconomics		
Household size	+	(Ameli et al., 2015; Ewing et al., 2016; Ewing and Clemente, 2013; Park et al., 2019)
	ns / –	(Hamidi and Moazzeni, 2019; Maxwell, 2016)
Mean/median income	–	(Hankey et al., 2017; Lu et al., 2018; Park et al., 2019; Pulugurtha and Repaka, 2013)
	ns / +	(Hankey et al., 2012; Hankey and Lindsey, 2016; Pulugurtha and Repaka, 2013, 2008; Rodríguez et al., 2009; Schneider et al., 2021; Tabeshian and Kattan, 2014)
Environmental		
Park density or proximity	+	(Kang, 2017, 2015)
	ns / –	(Kang, 2017, 2015; Miranda-Moreno and Fernandes, 2011; Schneider et al., 2021; Sung et al., 2013)
Slope or grade	–	(Kang, 2018, 2017, 2015; Kim et al., 2019, 2017; Liu and Griswold, 2009; Schneider et al., 2012; Sung et al., 2013, 2015)
	ns / +	(Griswold et al., 2019)

[a] Association with pedestrian volume: "+" positive, "–" negative, "ns" not statistically significant.

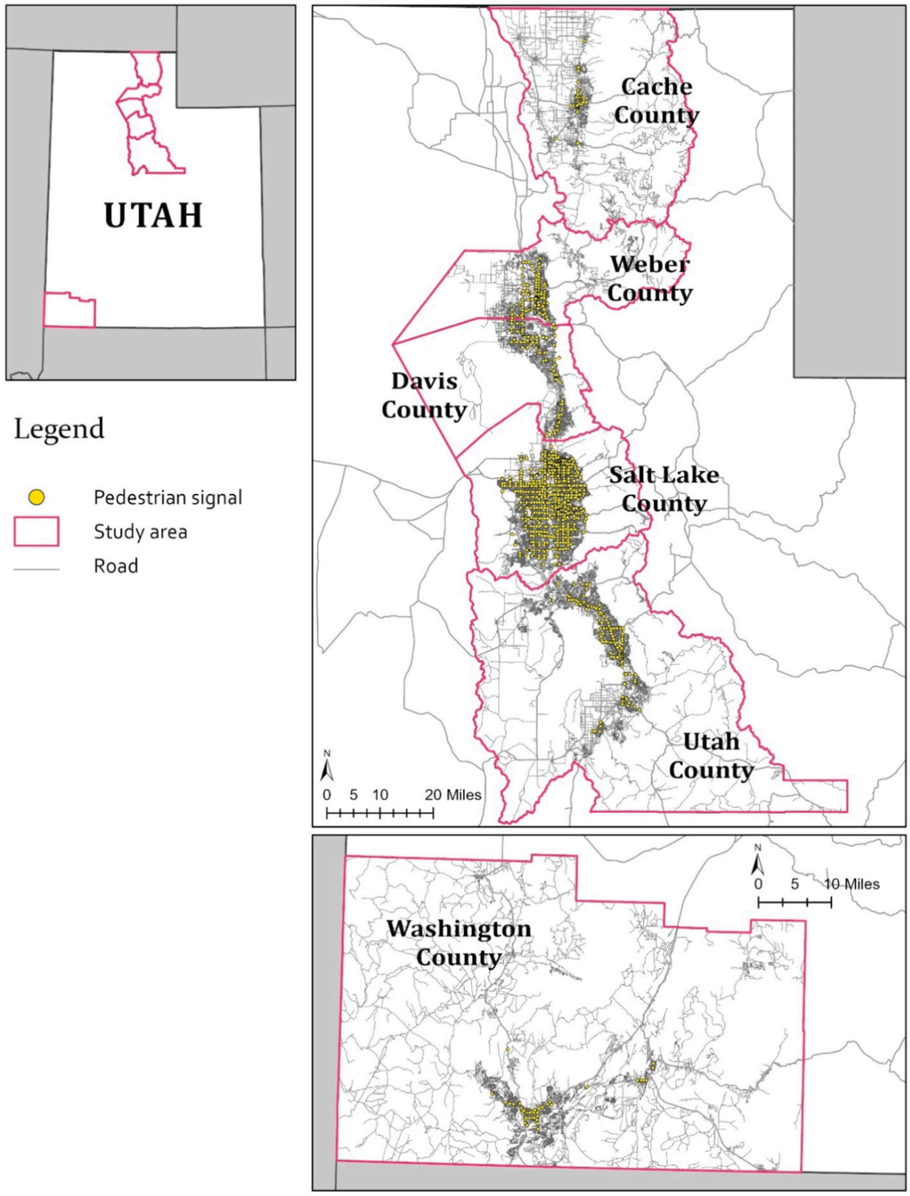

Fig. 1. Map of signalized intersections in the six most populous counties in Utah.

Table 3
Descriptive statistics for dependent variables.

Variable	*Min*	*Med*	*Max*	*Mean*	*SD*
Estimated annual average daily pedestrians (AADP)	1.08	116.13	6737.22	267.28	519.00
Weekdays (Monday–Friday)	1.12	133.15	7547.23	300.66	598.50
Weekends (Saturday–Sunday)	0.61	77.52	4712.21	183.82	352.54
Estimated annual average hourly pedestrians (AAHP)	0.31	33.87	1965.02	77.96	151.38
00:00–02:59	0.00	2.99	328.01	11.04	27.86
03:00–05:59	0.02	3.44	376.70	9.86	25.56
06:00–08:59	0.10	33.92	1889.48	71.33	135.64
09:00–11:59	0.36	40.89	2926.11	101.74	216.95
12:00–14:59	0.31	58.17	3757.55	137.90	288.31
15:00–17:59	0.62	67.85	3409.01	150.67	290.58
18:00–20:59	0.35	38.25	2566.70	98.00	201.30
21:00–23:59	0.05	15.81	946.61	43.10	86.37

3.2. Built environment data

Neighborhood built environment variables were measured for two different buffer widths—½-mile and ¼-mile—in a belief that the number of pedestrians may depend on the neighborhood environment at different scales. For example, the influence of road traffic volume on pedestrian activity may be only significant over a short distance while that of street network connectivity may be more extensive. A quarter-mile and a half-mile were selected as a standard walking distance beyond which walk frequency drops off rapidly; they are used in most travel behavior literature (Ewing and Clemente, 2013; Nagel et al., 2008). Thus, using the "Network Analyst" tool in the ArcGIS Pro software, we created street network-based buffers by ½-mile and ¼-mile for every signalized intersection.

For the predictors of pedestrian signal activity, we measured "D" variables—density, diversity, design, destination accessibility, and distance to transit—as well as socioeconomic factors. For density variables, we measured population density (number of 1000 people per square mile) and employment density (number of 1000 jobs per square mile). The population data came from the American Community Survey (ACS)

2013–2017 at the Census block group level, and the employment data (2017) were collected from the Longitudinal Employer-Household Dynamics (LEHD) at the Census block level. Then, the data were assigned to the buffers based on the relative areas of the Census boundaries (i.e., the spatial apportioning technique). For the land use variables, we compiled parcel-level land use maps from the Utah Automated Geographic Reference Center (AGRC) for the year 2019 and computed the percentage of residential parcels, percentage of commercial parcels, number of schools, number of places of worship, and total acreage of parks.

For a transit variable, we measured the number of transit stops in each buffer area. Transit stop location data in 2019 was available at OpenMobilityData (https://transitfeeds.com/) as a form of General Transit Feed Specification (GTFS). Also, two gross measures of street network design were computed, using intersection location data provided by the Metropolitan Research Center at the University of Utah. Intersection density (a measure of the block size) was computed as the number of intersections within a buffer divided by the gross area of the buffer in square miles. The proportion of four-way intersections (a measure of street connectivity) was computed as the number of four-way intersections divided by the total number of intersections within the buffer area.

Three demographic variables were also included—average household size, median household income, and average vehicle ownership—for block groups intersecting with the buffer. We hypothesized that more affluent residents with more vehicles available might walk less and drive more, while bigger households might walk more (Ewing et al., 2015; Owen et al., 2007). Data for demographic measures were gathered from the ACS (2017 5-year estimates) and assigned to the buffer using the spatial apportioning technique described above. Lastly, as a measure of traffic safety, we included road types for roads near the intersection. Road types were divided into three categories based on the cartographic code of road centerline data, provided by UDOT: highways (interstates, US and state highways, and associated ramps), major roads ("major local roads" such as arterials), and local roads (the rest, including collectors). (We wanted to include Annual Average Daily Traffic (AADT) volumes in the model, but they were not available for several signals and most intersections where one would want to apply these data. Also, preliminary models found AADT to be not significantly associated with pedestrian volumes.)

Table 4 shows descriptive statistics for the built environment variables. Within a given buffer width, all correlations between these variables were low-to-moderate (< 0.55) except for a negative correlation between residential and commercial land uses (−0.75). Also, the highest variance inflation factor (VIF) values in the regression models were lower than 5. Therefore, we conclude that multicollinearity among independent variables was not an issue.

Table 4
Descriptive statistics for independent variables.

Variable	¼-mile		½-mile	
	Mean	SD	Mean	SD
Population density (1000 per sq. mi.)	4.39	2.80	4.44	2.55
Employment density (1000 per sq. mi.)	5.60	8.10	4.85	6.31
Household size (average)	3.09	1.09	3.10	0.98
Household income ($1000; median)	59.75	23.21	60.27	22.40
Vehicle ownership (average)	1.68	0.51	1.69	0.47
% residential land use	31.02	22.72	37.17	21.37
% commercial land use	29.38	20.11	24.74	16.86
Intersection density (per sq. mi.)	97.97	49.01	100.32	38.86
% 4-way intersections	28.46	21.88	25.79	16.61
# schools	0.30	0.62	0.92	1.18
# places of worship	0.52	0.80	1.79	1.84
# transit stops	4.81	3.94	12.71	9.93
Park acreage	1.46	3.59	5.54	9.10

3.3. Direct-demand volume modeling

Consistent with many other studies using built environment characteristics to predict pedestrian volumes (see Table 1), we employed a log-linear regression model in which our dependent variable is transformed using the natural log function. We decided against applying a negative binomial (or Poisson-gamma mixture) regression model—traditionally used to model count data—because our pedestrian data are not actually count data; instead, they are averages of counts. We used the log transform because our data are strictly positive and are positively skewed. An implication of the log-transformed dependent variable is that we can interpret our estimated coefficients (when exponentiated) as proportional or percentage changes (rather than absolute changes) in pedestrian signal activity due to changes to our independent variables.

The pedestrian data in this study may have an issue of spatial autocorrelation, meaning that the estimated pedestrian activity at one signal is correlated with activity at nearby signals. Reasons for this might include walk trips that extend from one block to the next, similar demographics or urban form characteristics, or a large-scale destination in one block (e.g., a regional park, convention center, or theater). Moran's I statistic is a commonly-used measure to check for spatial autocorrelation. Any spatial pattern in the residuals violates the assumption of regression models that residuals are independent of each other and randomly distributed. Before controlling for the spatial autocorrelation, Moran's I for model residuals in this study ($p < .001$) indicated a strongly positive spatial relationship.

The spatial lag or error model can be used as a robust tool to deal with the spatial autocorrelation issue in ordinary least squares (OLS) regression. The Lagrange multiplier test is used to assess whether the autocorrelation is in the dependent variable or in the errors and helps in the choice of a spatial regression model. The robust Lagrange multiplier test indicated a spatial error model as the most suitable method, and thus, we employed spatial error models that treat spatial autocorrelation between the residuals of adjacent areas. We ran spatial error models using *errorsarlm* function (*spdep* package) in *R 3.6.1* software. The Moran's I values for the final models' residuals ($p > .1$), indicated no spatial autocorrelation.

As explained above, neighborhood environment variables were measured for two different buffer widths: ½-mile and ¼-mile. In our models, trial and error between the two buffer widths for each independent variable was used to arrive at the best-fit models. The best-fit models were chosen based on the statistical significance of the variable (i.e., p-value) and the goodness-of-fit of the model (i.e., lower AIC and BIC values).

3.4. Model validation

To test how well our models can predict actual pedestrian volumes, we evaluated the predictive performance of our models by running k-fold cross-validation (Fielding and Bell, 1997; Hair et al., 2006). Using the same data to estimate parameters and to test predictive accuracy may overestimate model validity. In k-fold cross-validation, the data are divided into k equal partitions. In this study, data were randomly divided into ten folds: 90% of the data (training data) used for model fitting and 10% of the data withheld for model validation in each iteration. The root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) are used as three measures of the prediction capability of regression models (Chai and Draxler, 2014; Willmott and Matsuura, 2005). This procedure is repeated for each of the k partitions, and the RMSE, MAE, and MAPE values are averaged to obtain the mean value.

4. Results

Table 5 shows three models for daily pedestrian activity (AADP) for all days, weekdays, and weekends, respectively. Lambda represents a

Table 5
Model results for AADP and AAHP.

n = 1494 signals	Day of week (AADP)						Time of day (AAHP)											
	All days		Mon–Fri		Sat–Sun		6 am–9 am		9 am–12 pm		12 pm–3 pm		3 pm–6 pm		6 pm–9 pm		9 pm–12 am	
Variable	B	sig[a]	B	sig[a]	B	sig[a]	B	sig[a]	B	sig[a]	B	sig[a]	B	sig[a]	B	sig[a]	B	sig[a]
(Intercept)	2.747	*	2.897	*	2.275	*	1.871	*	1.706	*	1.926	*	2.123	*	1.452	*	0.986	*
Population density (½-mile)[b]	0.326	*	0.344	*	0.373	*	0.258	*	0.297	*	0.338	*	0.347	*	0.402	*	0.553	*
Employment density (¼-mile)[b]	0.124	*	0.136	*	0.070	*	0.083	*	0.134	*	0.152	*	0.125	*	0.116	*	0.126	*
Household size (¼-mile)[b]	0.418	*	0.452	*	0.146		0.429	*	0.385	*	0.434	*	0.450	*	0.331	*	0.252	*
Household income (½-mile)	−0.010	*	−0.010	*	−0.008	*	−0.008	*	−0.009	*	−0.010	*	−0.010	*	−0.010	*	−0.014	*
Vehicle ownership (½-mile)	−0.198	*	−0.217	*	−0.103		−0.282	*	−0.191	*	−0.173	*	−0.196	*	−0.134	~	−0.141	~
% residential (¼-mile)	0.006	*	0.006	*	0.006	*	0.008	*	0.004	~	0.006	*	0.007	*	0.005	*	0.003	
% commercial (¼-mile)	0.019	*	0.019	*	0.022	*	0.014	*	0.020	*	0.021	*	0.020	*	0.021	*	0.019	*
Intersection density (½-mile)	0.004	*	0.004	*	0.004	*	0.003	*	0.004	*	0.004	*	0.004	*	0.004	*	0.003	*
% 4-way intersections (½-mile)	0.006	*	0.006	*	0.008	*	0.005	*	0.007	*	0.006	*	0.006	*	0.008	*	0.007	*
# schools (¼-mile)	0.155	*	0.170	*	0.065		0.247	*	0.117	*	0.169	*	0.160	*	0.079	~	0.030	
# places of worship (½-mile)	0.060	*	0.054	*	0.080	*	0.047	*	0.069	*	0.068	*	0.057	*	0.070	*	0.062	*
# transit stops (¼-mile)	0.068	*	0.069	*	0.066	*	0.061	*	0.075	*	0.074	*	0.072	*	0.070	*	0.062	*
Park acreage (½-mile)[b]	0.022	*	0.023	*	0.025	*	0.021	*	0.020	*	0.022	*	0.022	*	0.029	*	0.027	*
Road type (major road dummy)	0.242	*	0.245	*	0.245	*	0.266	*	0.235	*	0.225	*	0.263	*	0.227	*	0.215	*
Model diagnostics[c]	Lambda: 0.49		Lambda: 0.49		Lambda: 0.46		Lambda: 0.50		Lambda: 0.50		Lambda: 0.48		Lambda: 0.48		Lambda: 0.48		Lambda: 0.47	
	AIC: 3772		AIC: 3784		AIC: 3909.7		AIC: 3986.6		AIC: 3767.6		AIC: 3800		AIC: 3817		AIC: 3894.5		AIC: 4032.5	

[a] *: $p < .05$; ~: $p < .1$.
[b] Log-transformed.
[c] All Lambdas are $p < .001$.

coefficient on the spatially correlated errors (Anselin and Rey, 2010): it has a positive effect and is statistically significant in all models.

Most built environment variables—population density, employment density, % residential parcels, % commercial parcels, intersection density, % 4-way intersections, schools, places of worship, transit stops, and park acreage—were statistically significant at a $p < .05$ level and positively associated with the estimated average daily volumes of pedestrians. Among demographic variables, pedestrian volume increased with average household size and decreased with median household income and average vehicle ownership of households living near the intersection. Pedestrian volume increased significantly when the intersection contained major roads, compared with only highway or local road types.

Notable day-of-week differences were also found. As expected, the number of schools near the intersection was not significant in the weekend model; so were two other demographic variables: household size and vehicle ownership. Albeit statistically significant across the three daily models, a higher coefficient for the employment density variable was found on weekdays while the population density variable had a bigger effect size on weekends. Also, the coefficient for places of worship was higher in the weekend model.

Table 5 also shows six models for hourly pedestrian activity (AAHP) for specific times of day, in 3-h windows from 6 am to midnight. Lambda values had a positive effect and were statistically significant in all models. Again, most built environmental variables were positively associated with the pedestrian volumes across the day at a $p < .05$ significance level. Average household size (positively) and median household income (negatively) were also statistically significant in all time-of-day models of pedestrian volume.

The number of schools near an intersection was positively associated with pedestrian activity, but only during the daytime (6 am–6 pm). Residential land use became statistically non-significant during the nighttime (in the after-9 pm or before-6 am models; the latter models are not shown in the table). The slope coefficients of population density were higher during the nighttime (after-6 pm models) while those of employment density were higher during the daytime (models for 9 am–3 pm). The coefficient for being on a major road (as opposed to a highway or local road) was strongest during peak hours (6 am–9 am and 3 pm–6 pm).

After fitting the models with the full data, we assessed the predictive power of the nine models using 10-fold cross-validation. Intersections ($n = 1494$) were randomly split into ten equal-sized groups. The validation data set (10% of the data) was used to validate the model, which was fitted using the other 90% of the data through a spatial error model. As a result of the 10-fold cross-validation, we obtained average RMSE, MAE, and MAPE for each model. From the cross-validation results, the average RMSEs ranged from 0.932 (AADP model) to 1.027 (9 pm–12 am model); the average MAEs were between 0.699 (9 am–12 pm model) and 0.788 (9 pm–12 am model); and the average MAPEs ranged from 22.0% (Mon–Fri model) to 131.0% (9 pm–12 am model). These error values are comparable to those from the full model (RMSEs: 0.928–1.003; MAEs: 0.690–0.771; MAPEs: 21.8–108.4%), indicating that our predictive models are stable for new input data. A further exploration of errors show that pedestrian traffic volumes were underestimated in the areas with highest pedestrian volume such as downtowns and near university campuses, findings which call for additional explanatory variables or non-linear functions.

5. Discussion

To meet our study objective of examining relationships between the built environment and pedestrian activity, we developed direct demand built environment models of daily and hourly pedestrian crossing volumes at signalized intersections using a novel data source: volumes estimated using pedestrian push-button events from high-resolution traffic signal controller logs. Like in past research, we used log-linear regression and controlled spatial autocorrelation, and we examined traditional built environment measures like activity density, land use, transit access, street network design, and neighborhood sociodemographics. In contrast to previous work, we employed a continuously-collected measure of pedestrian activity estimated from signal data, measured over the course of one full year, and averaged per day and per hour. Notably, we also identified day-of-week and time-of-day variations in built environment relationships with walking volumes, which we believe to be a relatively unique contribution to the literature (see Lu et al. (2018) for one other example). Another contribution of our work is that we used a larger sample size of sites (1494 signalized intersections from different areas in Utah) than almost any other past effort, giving our analysis more power and potentially making our results more generalizable.

Indeed, all of our findings are consistent with theory and expectations (from past research) regarding links between walking and the built environment (see Table 2), which supports the validity of our pedestrian measures. Intersections with greater population and employment densities and higher percentages of nearby residential and commercial land uses saw more pedestrian activity (Ameli et al., 2015; Behnam and Patel, 1977; Ewing et al., 2016; Ewing and Clemente, 2013; Kim et al., 2019; Liu and Griswold, 2009; Miranda-Moreno et al., 2011; Miranda-Moreno and Fernandes, 2011; Ozbil et al., 2011; Park et al., 2019; Pulugurtha and Repaka, 2013, 2008; Schneider et al., 2012; Sung et al., 2013). Transit stop density was strongly and positively linked to walking (Miranda-Moreno et al., 2011; Miranda-Moreno and Fernandes, 2011; Park et al., 2019; Sung et al., 2013). Regarding sociodemographic characteristics, as has been found previously, pedestrian activity was greater in neighborhoods with larger household sizes (Ameli et al., 2015; Ewing et al., 2016; Ewing and Clemente, 2013; Park et al., 2019). Overall, these results continue to support research-informed built environment interventions and land use policies aimed at creating more walkable communities.

Our analysis was also able to uncover theoretically-consistent relationships between walking and other built environmental attributes for which past research has more commonly found null or theoretically-inconsistent findings. Signals in areas with greater street network connectivity had more pedestrian crossing events, which has been found in only a few prior studies for intersection density (Hajrasouliha and Yin, 2015; Hamidi and Moazzeni, 2019) and the percentage of four-way intersections (Miranda-Moreno et al., 2011; Miranda-Moreno and Fernandes, 2011; Park et al., 2019). Specific nearby destinations like parks also attracted more pedestrian crossings, which has only been found in studies by Kang (2017, 2015). Pedestrian volumes were greater in neighborhoods with lower median household incomes, which has been found in some studies (Hankey et al., 2017; Lu et al., 2018; Park et al., 2019; Pulugurtha and Repaka, 2013) but not in other studies (Hankey et al., 2012; Pulugurtha and Repaka, 2008; Rodríguez et al., 2009). One of our findings is perhaps contrary to expectation: the positive association of pedestrian activity with major roads. It could be that the design and traffic volumes on these streets encourage pedestrians to cross at the signal rather than at an unsignalized intersection (Schneider et al., 2012), or that pedestrian attractors (businesses, transit stops) are commonly located along these streets (Griswold et al., 2019).

The use of a continuously-recorded pedestrian data source also allowed us to examine time-of-day and day-of-week variations in these built environment relationships that are not feasible to consider when using only short-duration pedestrian counts. Many factors had similar relationships with pedestrian activity throughout the week and across the day, but a few did not. Population density seemed to be most relevant (with a larger coefficient) on weekends and during evening hours, when we expect more people to be at home. For example, a 10% increase in population density would be expected to yield a 4.0% increase ($1.10^{0.402}$) in evening hourly pedestrian volumes (6–9 pm), but only a 2.5% increase ($1.10^{0.258}$) during the morning (6–9 am). Lu et al. (2018) also found population density to have a larger coefficient during evening hours than during the day. Conversely, employment density played a bigger role on weekdays and during daytime hours: a 10% increase in employment density would be expected to generate 1.3% more ($1.10^{0.136}$) daily pedestrians during weekdays, but only 0.7% more ($1.10^{0.070}$) during on weekends. As expected, our models showed that intersections near schools had greater pedestrian activity, but only or especially when primary/secondary schools are in session: on weekdays and during morning and afternoon commuting hours. This finding supports traffic calming and safety efforts around primary/secondary schools, including school zone speed limits and crossing guards.

Despite these contributions, a limitation of this work is the use of pedestrian volumes estimated from traffic signal data as opposed to observed pedestrian counts or crossing volumes. Previous research on pedestrian behavior and the utilization of pedestrian push-buttons at signals has found that rates vary across locations such as by signal type (Kutela and Teng, 2020), in different situations like the presence/absence of approaching motor vehicles (Foster et al., 2014), and by age, gender, and other pedestrian characteristics (Kutela and Teng, 2020). These factors and their aggregated versions (i.e., motor vehicle traffic volumes and neighborhood socio-demographics) have not been considered in the models upon which our estimated pedestrian volume data are based (Singleton et al., 2020b; Singleton and Runa, 2021). However (as previously mentioned), research from Utah and other states (Blanc et al., 2015; Kothuri et al., 2017; Li and Wu, 2021; Singleton et al., 2020b; Singleton and Runa, 2021) has found pedestrian push-button event data to be highly correlated with observed pedestrian crossing volumes. So, any improvement in the accuracy of our models' dependent variables through the addition of factors like these would likely be modest.

Another limitation is that the locations where pedestrian signal data are available may not be entirely representative. These data are not available at signals without pedestrian detection: in our study, these included some high-pedestrian downtown intersections that operate without push-buttons, as well as a few intersections in heavily-industrial areas and isolated freeway interchanges. Also, signalized intersections tend to be more highly concentrated along larger, arterial roadways and in urban areas, so our findings may not be completely generalizable to non-signalized intersections, and our data may capture more utilitarian walk trips. That said, more than 90% of Utah's population lives in an urban area, and we did find more walking near parks. It could be advantageous to combine signal-based estimates of pedestrian volumes with data from permanent pedestrian counters on trails and in other more recreational contexts in order to improve the generalizability of direct demand models. Overall, these methods may be most appropriate for moderately urban to suburban locations. Nevertheless, this trait is fortunate, since (in the US) these tend to be the locations most lacking in pedestrian data and where tradeoffs have to be made between priorities (e.g., in signal timing) for pedestrians vs. motor vehicle drivers.

Finally, there are opportunities to improve upon our analysis through additional research. Future studies could examine seasonal variations in daily pedestrian activity at signalized interactions, which would consider effects due to weather variables such as temperature, precipitation, and wind (Runa and Singleton, 2021). Also, because pedestrian traffic volumes may not be linearly related to all built environment variables, future studies may use non-linear regression such as generalized additive models (Park et al., 2020) or machine learning algorithms such as gradient boosting decision trees or random forests (Cheng et al., 2019; Ding et al., 2018). We expect that by using long-term automated counts derived from traffic signal event data, our pedestrian measures can potentially do a better job of reducing the random variability arising from short-term (usually ⟨12h) counts, thus yielding more robust relationships with measures of the built environment. However, this topic—quantifying error associated with estimates of pedestrian volumes using different durations of count data (Johnstone et al., 2018; Nordback et al., 2019)—is another subject for further study. Research should also continue to explore the feasibility and accuracy of other pedestrian detection methods—video image processing (Rahman et al., 2019), lidar (Zhao et al., 2019), and others—for pedestrian volume monitoring applications.

Despite these limitations and opportunities for future work, we think our theoretically-consistent findings about built environment relationships with walking—and our ability to detect day-of-week and time-of-day variations in those relationships—demonstrate the utility of traffic signal data sources for direct demand pedestrian volume modeling. There are hundreds of thousands of traffic signals across the US (NTOC, 2012), many with pedestrian push-buttons (more than 85% in Utah). Also, many states and regions (including Utah, Georgia, and the Phoenix, Las Vegas, and Orlando areas) have or are actively developing ATSPM systems to archive pedestrian detections and other signal events. These trends make our methods increasingly applicable for the development of locally-calibrated direct demand pedestrian volume models. Additionally, the ultimate objective of direct demand models is to predict pedestrian volumes in areas and for locations without current pedestrian data. In fact, the specific models presented in this paper can be applied, using built environment data, to estimate average daily/hourly pedestrian volumes at thousands of unsignalized intersections through Utah (Singleton et al., 2020a). Such estimates would be valuable for various transportation planning, design, and operational tasks, including as a measure of exposure for pedestrian safety studies. Overall, this work provides planners with more tools to model, analyze, and plan for pedestrians with greater temporal resolution.

Declaration of Competing Interest

The authors have no competing interests to disclose.

Acknowledgements

This work was funded by the Utah Department of Transportation through project UT19.504. The authors alone are responsible for the preparation and accuracy of the information, data, analysis, discussions, recommendations, and conclusions presented herein. The contents do not necessarily reflect the views, opinions, endorsements, or policies of the Utah Department of Transportation or the US Department of Transportation. The Utah Department of Transportation makes no representation or warranty of any kind, and assumes no liability therefore. The comments from several anonymous reviewers significantly improved this work.

References

Ameli, S.H., Hamidi, S., Garfinkel-Castro, A., Ewing, R., 2015. Do better Urban Design qualities Lead to more walking in Salt Lake City, Utah? J. Urban Des. 20, 393–410. https://doi.org/10.1080/13574809.2015.1041894.

Anselin, L., Rey, S., 2010. Properties of tests for spatial dependence in linear regression models. Geogr. Anal. 23, 112–131. https://doi.org/10.1111/j.1538-4632.1991.tb00228.x.

Arnold, L., Donlon, J., Jones, M.G., Ledbetter, L., Ragland, D.R., Ryan, S., 2010. Seamless Travel: Measuring Bicycle and Pedestrian Activity in San Diego County and its Relationship to Land Use, Transportation, Safety, and Facility Type. University of California Institute of Transporation Studies, Berkeley, CA.

Behnam, J., Patel, B.G., 1977. A method for estimating pedestrian volume in a central business district. Transp. Res. Rec. 629, 22–26.

Blanc, B., Johnson, P., Figliozzi, M., Monsere, C., Nordback, K., 2015. Leveraging signal infrastructure for nonmotorized counts in a statewide program. Transp. Res. Rec. J. Transp. Res. Board 2527, 69–79. https://doi.org/10.3141/2527-08.

Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)?–arguments against avoiding RMSE in the literature. Geosci. Model Dev. 7, 1247–1250.

Cheng, L., Chen, X., De Vos, J., Lai, X., Witlox, F., 2019. Applying a random forest method approach to model travel mode choice behavior. Travel Behav. Soc. 14, 1–10. https://doi.org/10.1016/j.tbs.2018.09.002.

Day, C.M., Premachandra, H., Bullock, D.M., 2011. Rate of pedestrian signal phase actuation as a proxy measurement of pedestrian demand. In: Presented at the 90th Annual Meeting of the Transportation Research Board, Washington, DC. http://docs.lib.purdue.edu/civeng/24.

Day, C.M., Taylor, M., Mackey, J., Clayton, R., Patel, S.K., Xie, G., Li, H., Sturdevant, J.R., Bullock, D., 2016. Implementation of automated traffic signal performance measures. ITE J. Inst. Transp. Eng. 86, 26–34.

Desyllas, J., Duxbury, E., Ward, J., Smith, A., 2003. Pedestrian Demand Modelling of Large Cities: An Applied Example from London. (CASA Working Papers 62). London, UK.

Ding, C., Cao, X., Næss, P., 2018. Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. Transp. Res. Part A Policy Pract. 110, 107–117. https://doi.org/10.1016/j.tra.2018.02.009.

Ewing, R., Cervero, R., 2010. Travel and the built environment. J. Am. Plan. Assoc. 76, 265–294. https://doi.org/10.1080/01944361003766766.

Ewing, R., Clemente, O., 2013. Measuring Urban Design: Metrics for Livable Places. Island Press, Washington, DC.

Ewing, R., Handy, S., 2009. Measuring the Unmeasurable: Urban Design qualities related to walkability. J. Urban Des. 14, 65–84. https://doi.org/10.1080/13574800802451155.

Ewing, R., Tian, G., Goates, J., Zhang, M., Greenwald, M.J., Joyce, A., Kircher, J., Greene, W., 2015. Varying influences of the built environment on household travel in 15 diverse regions of the United States. Urban Stud. 52, 2330–2348. https://doi.org/10.1177/0042098014560991.

Ewing, R., Hajrasouliha, A., Neckerman, K.M., Purciel-Hill, M., Greene, W., 2016. Streetscape features related to pedestrian activity. J. Plan. Educ. Res. 36, 5–15. https://doi.org/10.1177/0739456X15591585.

Federal Highway Administration (FHWA), 2016. Traffic Monitoring Guide. U.S. Department of Transportation, Washington, DC. https://www.fhwa.dot.gov/policyinformation/tmguide/.

Fielding, A.H., Bell, J.F., 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. Environ. Conserv. 24, 38–49.

Foster, N., Monsere, C.M., Carlos, K., 2014. Evaluating Driver and Pedestrian Behaviors at Enhanced, Multilane, Midblock Pedestrian Crossings. Transp. Res. Rec. 2464, 59–66. https://doi.org/10.3141/2464-08.

Griswold, J.B., Medury, A., Schneider, R.J., Amos, D., Li, A., Grembek, O., 2019. A pedestrian exposure model for the California state highway system. Transp. Res. Rec. J. Transp. Res. Board 2673, 941–950. https://doi.org/10.1177/0361198119837235.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., Tatham, R., 2006. Multivariate Data Analysis, vol. 5. Pearson Prentice Hall, NJ, pp. 207–219. No. 3.

Hajrasouliha, A., Yin, L., 2015. The impact of street network connectivity on pedestrian volume. Urban Stud. 52, 2483–2497. https://doi.org/10.1177/0042098014544763.

Hamidi, S., Moazzeni, S., 2019. Examining the relationship between Urban Design qualities and walking behavior: empirical evidence from Dallas, TX. Sustainability 11, 2720. https://doi.org/10.3390/su11102720.

Hankey, S., Lindsey, G., 2016. Facility-demand models of peak period pedestrian and bicycle traffic: comparison of fully specified and reduced-form models. Transp. Res. Rec. J. Transp. Res. Board 2586, 48–58. https://doi.org/10.3141/2586-06.

Hankey, S., Lindsey, G., Wang, X., Borah, J., Hoff, K., Utecht, B., Xu, Z., 2012. Estimating use of non-motorized infrastructure: models of bicycle and pedestrian traffic in Minneapolis. MN. Landsc. Urban Plan. 107, 307–316. https://doi.org/10.1016/j.landurbplan.2012.06.005.

Hankey, S., Lu, T., Mondschein, A., Buehler, R., 2017. Spatial models of active travel in small communities: merging the goals of traffic monitoring and direct-demand modeling. J. Transp. Health 7, 149–159. https://doi.org/10.1016/j.jth.2017.08.009.

Hillier, B., 2007. Space Is the Machine: A Configurational Theory of Architecture. Space Syntax, London, UK.

Hillier, B., Penn, A., Hanson, J., Grajewski, T., Xu, J., 1993. Natural movement: or, configuration and attraction in urban pedestrian movement. Environ. Plan. B Plan. Des. 20, 29–66. https://doi.org/10.1068/b200029.

Johnstone, D., Nordback, K., Kothuri, S., 2018. Annual average nonmotorized traffic estimates from manual counts: quantifying error. Transp. Res. Rec. J. Transp. Res. Board 2672, 134–144. https://doi.org/10.1177/0361198118792338.

Kang, C.-D., 2015. The effects of spatial accessibility and centrality to land use on walking in Seoul, Korea. Cities 46, 94–103. https://doi.org/10.1016/j.cities.2015.05.006.

Kang, C.-D., 2017. Measuring the effects of street network configurations on walking in Seoul, Korea. Cities 71, 30–40. https://doi.org/10.1016/j.cities.2017.07.005.

Kang, C.-D., 2018. The S + 5Ds: spatial access to pedestrian environments and walking in Seoul, Korea. Cities 77, 130–141. https://doi.org/10.1016/j.cities.2018.01.019.

Kim, T., Sohn, D.-W., Choo, S., 2017. An analysis of the relationship between pedestrian traffic volumes and built environment around metro stations in Seoul. KSCE J. Civ. Eng. 21, 1443–1452. https://doi.org/10.1007/s12205-016-0915-5.

Kim, S., Park, S., Jang, K., 2019. Spatially-varying effects of built environment determinants on walking. Transp. Res. Part A Policy Pract. 123, 188–199. https://doi.org/10.1016/j.tra.2019.02.003.

Kothuri, S., Nordback, K., Schrope, A., Phillips, T., Figliozzi, M., 2017. Bicycle and pedestrian counts at signalized intersections using existing infrastructure. Transp. Res. Rec. J. Transp. Res. Board 2644, 11–18. https://doi.org/10.3141/2644-02.

Kutela, B., Teng, H., 2020. Evaluating the influential factors for pushbutton utilization at signalized midblock crosswalks. Saf. Sci. 122, 104533. https://doi.org/10.1016/j.ssci.2019.104533.

Kuzmyak, J.R., Walters, J., Bradley, M., Kockelman, K.M., 2014. Estimating Bicycling and Walking for Planning and Project Development: A Guidebook. Transportation Research Board, Washington, D.C. https://doi.org/10.17226/22330.

Le, M., Geedipally, S.R., Fitzpatrick, K., Avelar, R.E., 2020. Estimating pedestrian volumes for signalized and stop-controlled intersections. Transp. Res. Rec. J. Transp. Res. Board 2674, 799–808. https://doi.org/10.1177/0361198120932161.

Li, X., Wu, Y.-J., 2021. Real-time estimation of pedestrian volume at button-activated midblock crosswalks using traffic controller event-based data. Transp. Res. C 122, 102876. https://doi.org/10.1016/j.trc.2020.102876.

Liu, X., Griswold, J., 2009. Pedestrian volume modeling: a case study of San Francisco. Yearb. Assoc. Pacific Coast Geogr. 71, 164–181. https://doi.org/10.1353/pcg.0.0030.

Lu, T., Mondschein, A., Buehler, R., Hankey, S., 2018. Adding temporal information to direct-demand models: Hourly estimation of bicycle and pedestrian traffic in Blacksburg, VA. Transp. Res. D Transp. Environ. 63, 244–260. https://doi.org/10.1016/j.trd.2018.05.011.

Maxwell, J.A., 2016. Designing for'life between buildings': Modeling the Relationship between Streetscape Qualities and Pedestrian Activity in Glasgow. University of Strathclyde, Scotland.

Miranda-Moreno, L.F., Fernandes, D., 2011. Modeling of pedestrian activity at signalized intersections: land use, urban form, weather, and spatiotemporal patterns. Transp. Res. Rec. J. Transp. Res. Board 2264, 74–82. https://doi.org/10.3141/2264-09.

Miranda-Moreno, L.F., Morency, P., El-Geneidy, A.M., 2011. The link between built environment, pedestrian activity and pedestrian–vehicle collision occurrence at signalized intersections. Accid. Anal. Prev. 43, 1624–1634. https://doi.org/10.1016/j.aap.2011.02.005.

Munira, S., Sener, I.N., 2017. Use of direct-demand modeling in estimating non-motorized activity: a meta-analysis. In: Safety through Disruption (Safe-D) National University Transportation Center. A&M Transportation Institute, Texas. https://safed.vtti.vt.edu/wp-content/uploads/2020/07/UTC-Safe-D_Direct-Demand-Model-for-PedBike_TTI-Report_12Oct17_Final.pdf.

Nagel, C.L., Carlson, N.E., Bosworth, M., Michael, Y.L., 2008. The relation between neighborhood built environment and walking activity among older adults. Am. J. Epidemiol. 168, 461–468. https://doi.org/10.1093/aje/kwn158.

Nordback, K., Kothuri, S., Johnstone, D., Lindsey, G., Ryan, S., Raw, J., 2019. Minimizing annual average daily nonmotorized traffic estimation errors: how many counters are needed per factor group? Transp. Res. Rec. J. Transp. Res. Board 2673, 295–310. https://doi.org/10.1177/0361198119848699.

Noyce, D.A., Bentzen, B.L., 2005. Determination of pedestrian push-button activation duration at typical signalized intersections. Transp. Res. Rec. J. Transp. Res. Board 1939, 63–68. https://doi.org/10.1177/0361198105193900108.

NTOC, 2012. 2012 National Traffic Signal Report Card: Technical Report. Washington, DC.

Owen, N., Cerin, E., Leslie, E., DuToit, L., Coffee, N., Frank, L.D., Bauman, A.E., Hugo, G., Saelens, B.E., Sallis, J.F., 2007. Neighborhood walkability and the walking behavior of Australian adults. Am. J. Prev. Med. 33, 387–395. https://doi.org/10.1016/j.amepre.2007.07.025.

Ozbil, A., Peponis, J., Stone, B., 2011. Understanding the link between street connectivity, land use and pedestrian flows. Urban Des. Int. 16, 125–141. https://doi.org/10.1057/udi.2011.2.

Park, K., Ewing, R., Sabouri, S., Larsen, J., 2019. Street life and the built environment in an auto-oriented US region. Cities 88, 243–251. https://doi.org/10.1016/j.cities.2018.11.005.

Park, K., Ewing, R., Sabouri, S., Choi, D., Hamidi, S., Tian, G., 2020. Guidelines for a polycentric region to reduce vehicle use and increase walking and transit use. J. Am. Plan. Assoc. 86, 236–249. https://doi.org/10.1080/01944363.2019.1692690.

Penn, A., Hillier, B., Banister, D., Xu, J., 1998. Configurational Modelling of urban movement networks. Environ. Plan. B Plan. Des. 25, 59–84. https://doi.org/10.1068/b250059.

Pulugurtha, S.S., Repaka, S.R., 2008. Assessment of models to measure pedestrian activity at signalized intersections. Transp. Res. Rec. J. Transp. Res. Board 2073, 39–48. https://doi.org/10.3141/2073-05.

Pulugurtha, S.S., Repaka, S.R., 2013. An assessment of models to estimate pedestrian demand based on the level of activity. J. Adv. Transp. 47, 190–205. https://doi.org/10.1002/atr.159.

Pushkarev, B., Zupan, J.M., 1971. Pedestrian travel demand. Highw. Res. Rec. 355, 37–53.

Qin, X., Ivan, J.N., 2001. Estimating pedestrian exposure prediction model in rural areas. Transp. Res. Rec. J. Transp. Res. Board 1773, 89–96. https://doi.org/10.3141/1773-11.

Raford, N., Ragland, D., 2004. Space syntax: innovative pedestrian volume modeling tool for pedestrian safety. Transp. Res. Rec. 1878, 66–74. https://doi.org/10.3141/1878-09.

Raford, N., Ragland, D., 2006. Pedestrian Volume Modeling for Traffic Safety and Exposure Analysis: The Case of Boston, Massachusetts. Safe Transportation Research & Education Center, UC Berkeley.

Rahman, M., Islam, M., Calhoun, J., Chowdhury, M., 2019. Real-time pedestrian detection approach with an efficient data communication bandwidth strategy. Transp. Res. Rec. J. Transp. Res. Board 2673, 129–139. https://doi.org/10.1177/0361198119843255.

Rodríguez, D.A., Brisson, E.M., Estupiñán, N., 2009. The relationship between segment-level built environment attributes and pedestrian activity around Bogota's BRT stations. Transp. Res. Part D Transp. Environ. 14, 470–478. https://doi.org/10.1016/j.trd.2009.06.001.

Runa, F., Singleton, P.A., 2021. Assessing the impacts of weather on pedestrian signal activity at 49 signalized intersections in northern Utah. Transp. Res. Rec. J. Transp. Res. Board. https://doi.org/10.1177/0361198121994111, 0361198121994111.

Ryus, P., Ferguson, E., Laustsen, K.M., Schneider, R.J., Proulx, F.R., Hull, T., Miranda-Moreno, L., 2014. Guidebook on pedestrian and bicycle volume data collection, Transportation Research Board. 10.17226/22223.

Sanders, R.L., Frackelton, A., Gardner, S., Schneider, R., Hintze, M., 2017. Ballpark method for estimating pedestrian and bicyclist exposure in Seattle, Washington. Transp. Res. Rec. J. Transp. Res. Board 2605, 32–44. https://doi.org/10.3141/2605-03.

Schneider, R.J., Arnold, L.S., Ragland, D.R., 2009. Pilot model for estimating pedestrian intersection crossing volumes. Transp. Res. Rec. J. Transp. Res. Board 2140, 13–26. https://doi.org/10.3141/2140-02.

Schneider, R.J., Henry, T., Mitman, M.F., Stonehill, L., Koehler, J., 2012. Development and application of volume model for pedestrian intersections in San Francisco, California. Transp. Res. Rec. J. Transp. Res. Board 2299, 65–78. https://doi.org/10.3141/2299-08.

Schneider, R.J., Schmitz, A., Qin, X., 2021. Development and validation of a Seven-County regional pedestrian volume model. Transp. Res. Rec. J. Transp. Res. Board. https://doi.org/10.1177/0361198121992360, 0361198121992360.

Singleton, P.A., Runa, F., 2021. Pedestrian traffic signal data accurately estimates pedestrian crossing volumes. Transp. Res. Rec. J. Transp. Res. Board 1–12. https://doi.org/10.1177/0361198121994126.

Singleton, P.A., Totten, J.C., Orrego-Oñate, J.P., Schneider, R.J., Clifton, K.J., 2018. Making strides: state of the practice of pedestrian forecasting in regional travel models. Transp. Res. Rec. J. Transp. Res. Board 2672, 58–68. https://doi.org/10.1177/0361198118773555.

Singleton, P.A., Park, K., Lee, D.H., 2020a. Predicted Pedestrian Volumes at Intersections (62k) in Utah. Utah State University. https://www.arcgis.com/home/item.html?id=32af57787d864a7996af76d0d9055cb2.

Singleton, P.A., Runa, F., Humagain, P., 2020b. Utilizing Archived Traffic Signal Performance Measures for Pedestrian Planning & Analysis. Salt Lake City, UT.

StreetLight InSight, 2018. StreetLight Active Mode Methodology, Data Sources and Validation [White Paper]. StreetLight Data, San Francisco, CA.

Sturdevant, J.R., Overman, T., Raamot, E., Deer, R., Miller, D., Bullock, D.M., Day, C.M., Brennan, T.M., Li, H., Hainen, A., 2012. Indiana traffic signal hi resolution data logger enumerations. West Lafayette, Indiana. https://doi.org/10.4231/K4RN35SH.

Sung, H.-G., Go, D.-H., Choi, C.G., 2013. Evidence of Jacobs's street life in the great Seoul city: identifying the association of physical environment with walking activity on streets. Cities 35, 164–173. https://doi.org/10.1016/j.cities.2013.07.010.

Sung, H., Go, D., Choi, C., Cheon, S., Park, S., 2015. Effects of street-level physical environment and zoning on walking activity in Seoul, Korea. Land Use Policy 49, 152–160. https://doi.org/10.1016/j.landusepol.2015.07.022.

Tabeshian, M., Kattan, L., 2014. Modeling nonmotorized travel demand at intersections in Calgary, Canada: use of traffic counts and Geographic Information System data. Transp. Res. Rec. J. Transp. Res. Board 2430, 38–46. https://doi.org/10.3141/2430-05.

Taylor, M., Mackey, J., 2018. Automated traffic signal performance measures. In: Presented at the 2018 UDOT Annual Conference. the 2018 UDOT Annual Conference, Sandy, UT.

Willmott, C., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Clim. Res. 30, 79–82. https://doi.org/10.3354/cr030079.

Wu, X., Lui, H.X., 2014. Using high-resolution event-based data for traffic modeling and control: an overview. Transp. Res. C 42, 28–43. https://doi.org/10.1016/j.trc.2014.02.001.

Zhao, J., Xu, H., Liu, H., Wu, J., Zheng, Y., Wu, D., 2019. Detection and tracking of pedestrians and vehicles using roadside LiDAR sensors. Transp. Res. C 100, 68–87. https://doi.org/10.1016/j.trc.2019.01.007.