Advances in pedestrian travel monitoring: Temporal patterns and spatial characteristics using pedestrian push-button data from Utah traffic signals

Prasanna Humagain
Utah State University
prasanna.hmg@usu.edu

Patrick A. Singleton (corresponding author)
Utah State University
patrick.singleton@usu.edu

Abstract: In this study, we advanced pedestrian travel monitoring using a novel data source: pedestrian push-button presses obtained from archived traffic signal controller logs at more than 1,500 signalized intersections in Utah over one year. The purposes of this study were to: (1) quantify pedestrian activity patterns; (2) create factor groups and expansion/adjustment factors from these temporal patterns; and (3) explore relationships between patterns and spatial characteristics. Using empirical clustering, we classified signals into five groups, based on normalized hourly/weekly counts (each hour's proportion of weekly totals, or the inverse of the expansion factors), and three clusters with similar monthly adjustment factors. We also used multinomial logit models to identify spatial characteristics (land use, built environment, socio-economic characteristics, and climatic regions) associated with different temporal patterns. For example, we found that signals near schools were much more likely to have bimodal daily peak hours and lower pedestrian activity during out-of-school months. Despite these good results, our hourly/weekday patterns differed less than in past research, highlighting the limits of existing infrastructure for capturing all kinds of activity patterns. Nevertheless, we demonstrated that signals with push-button data are a useful supplement to existing permanent counters within a broader pedestrian traffic monitoring program.

Keywords: Pedestrian travel monitoring, pedestrian push-buttons, empirical clustering, factor groups, land use

1   Introduction

Despite recent advances and interest from researchers and practitioners, pedestrian monitoring and data collection remains incomplete and insufficient, especially compared to motorized data collection. In practice, there are two kinds of methods for counting pedestrian activity at intersections: manual and automatic approaches (FHWA, 2016; Greene-Roesel et al., 2008). Manual counts involve collecting pedestrian volumes (by an observer) in real-time in situ or later using video recording. Although manual counts are accurate and advantageous for making modal distinctions (walking vs. cycling) and determining directional flows (left, straight, or right), the accuracy of counts depends upon characteristics

of observers (e.g., attentiveness) (Diogenes et al., 2007). More critically, manual counts are infeasible over longer time periods, because of the need for direct human supervision. Another method based on automated instruments—such as microwave, ultrasonic, infrared etc.—are feasible for longer-duration counts, and thus are ideal for identifying variations in pedestrian activity over time (Bu et al., 2007; Green-Roesel et al., 2008). However, automated counts can be susceptible to adverse weather or crowding, are expensive to install, and sometimes require periodic validation from manual counts (FHWA, 2016; Ryus et al., 2014).

1.1   Pedestrian push-button data to measure pedestrian activity

Alternatively, one novel source of pedestrian data is from pedestrian push-buttons at signalized intersections. Many (but not all) traffic signals require people walking who want to cross an approach to press a pedestrian push-button to request (actuate) the walk phase. Given readily available hardware and software, each pedestrian push-button press event can be timestamped, logged (Smaglik et al., 2007; Sturdevant et al., 2012), archived, and made available (for example) through the Automated Traffic Signal Performance Measures (ATSPM) system (ATKINS, 2016; Day, Bullock et al., 2014; Day, Taylor et al., 2016). Such high-resolution traffic signal controller log data are relatively ubiquitous in both time and space (available 24/7 at many intersections), making them a potentially rich source of information about pedestrian activity levels. Some of the limitations of existing methods—such as requirement of manual labor or upfront costs for installation of automated counters—and the lack of pedestrian data could be addressed by the use of this novel pedestrian data source.

Until recently, few studies investigated the use of pedestrian data from traffic signal controller logs to estimate walking activity. Day et al. (2011) analyzed data on pedestrian actuations per hour at one signalized intersection in Indiana over an 18-month period, finding impacts of time-of-day, day-of-week, weather and other seasonal effects, special events, and a change in pedestrian phase configuration on pedestrian actuations. Similarly, Blanc et al. (2015) and Kothuri et al. (2017) conducted studies of pedestrian activity at one intersection in Oregon that had actuated pedestrian crossings (using push-button detection) for all four crosswalks. The two Oregon studies used video data to manually count pedestrians, which they then compared to pedestrian actuations for each crosswalk, usually finding correlations of around 0.80 or greater. Recently, a large-scale validation study of pedestrian push-button use and walking activity at signalized intersections was conducted in Utah by Singleton et al. (2020). The authors compared hourly pedestrian signal activity metrics derived from push-button presses against observed pedestrian counts—obtained from manual counts of over 20,000 hours of videos recorded in 2019 for 320 crosswalks at 90 signalized intersections—using simple non-linear regression models. The models' estimated pedestrian volumes were strongly correlated with observed pedestrian crossing volumes (0.84) and had a low mean absolute error (3.0 pedestrians per hour) (Singleton et al., 2020; Singleton & Runa, 2021). Overall, these studies demonstrate that traffic signal data can be used to estimate pedestrian crossing volumes and monitor levels of pedestrian activity at intersections. We utilize pedestrian data from traffic signals in this paper.

1.2   Applications of continuously measured pedestrian data

Temporally-rich pedestrian data—measured continuously over time—has many applications. Quantifying and understanding the characteristics of pedestrian activity patterns in different spatial locations over time can assist planners and/or researchers in any (or all) of the following ways:

1. *Planning*: Pedestrian data can help planners to prioritize pedestrian infrastructure investments in specific areas and predict the impacts of new transportation or urban development projects on

walking.

2. *Safety*: Pedestrian safety analysis could use temporal patterns of pedestrian activity to better quantify risks related to exposure to traffic at crossings.
3. *Traffic operations*: Hourly distributions of pedestrian activity by location can assist with optimizing traffic signal timing for pedestrian delay or safety, as well as scheduling/permitting maintenance or construction work for areas and times with low pedestrian activity.
4. *Traffic monitoring*: Automated pedestrian counters cannot be deployed in all areas, so long-term count data are used to develop expansion factors that translate short-duration (e.g., manual) counts into estimated average annual daily pedestrian volumes, information which is useful for all of the activities listed above.

This fourth application (traffic monitoring) is the area to which our paper contributes.

1.3   Pedestrian expansion factors

To develop expansion factors, locations with similar pedestrian activity patterns (quantified either daily or weekly) are often grouped together into "factor groups" (FHWA, 2016; Medury et al., 2019; Ryus et al., 2017), each with a unique set of expansion factors. Short-duration pedestrian volume measurements (e.g., manual peak-period or daily counts) are then multiplied by the expansion factors—for the specific factor group to which that short-duration count location best belongs—in order to estimate long-term average pedestrian volumes more precisely (FHWA, 2016; Ryus et al., 2017). In current practice, there are two common approaches to constructing factor groups of multiple locations with similar pedestrian activity patterns. The first method is the *land-use classification* approach (Medury et al., 2019), which involves classifying locations based upon their surrounding land-use characteristics, under the assumption that locations with similar land uses will generate similar pedestrian activity patterns. Studies implementing this approach have identified distinct patterns for commercial areas, employment areas, university areas, trail areas, etc. (Medury et al., 2019; Schneider et al., 2009). The second method is the data-driven *empirical clustering* approach, which essentially groups locations based upon their pedestrian activity patterns, referred to as clusters. In short, the clustering algorithm works by minimizing differences in patterns within each cluster, while simultaneously maximizing differences in patterns between clusters. Miranda-Moreno and Lahti (2013) classified bicycle traffic patterns into four distinct groups as utilitarian, mixed-utilitarian, mixed-recreational, and recreational. Griswold et al. (2018) compared land use and empirical clustering approach and concluded that both approaches provided better results than a "single factor" method (where all locations are combined into single factor group). No matter the approach, the process of constructing factor groups is limited by the number and variety of locations with long-duration pedestrian count data.

1.4   Research Objectives

In this study, we aim to overcome limitations surrounding the lack of long-term automated pedestrian count data for traffic monitoring through the use of pedestrian push-button information from hundreds of sites in one US state. Specifically, we utilize high-resolution data collected from traffic signal controller logs at more than 1,500 signalized intersections throughout Utah—available from the Utah Department of Transportation (UDOT)'s ATSPM system—to investigate the temporal patterns of pedestrian activity and develop expansion factors and factor groups that relate to spatial characteristics. As such, the objectives of this paper are threefold, to:

1. Quantify and understand pedestrian activity patterns at signalized intersections, using continuous, archived data from pedestrian push-buttons at more than 1,500 signalized intersections in Utah.

2. Calculate time-of-day/day-of-week expansion factors and create factor groups based on empirical clustering of pedestrian activity patterns at signalized intersections, while accounting for seasonal variation.
3. Explore relationships between pedestrian factor groups and land use, built environment, and socio-economic neighborhood characteristics.

2   Data and methods

In this section, we present the data and an overview of the analysis methods. First, we describe calculating the pedestrian activity metrics for two temporal dimensions—hourly/weekday patterns, and monthly (seasonal) patterns—from traffic signal controller log data, as well as assembling data on spatial characteristics from various sources. Second, we explain the analysis methods employed, including empirical clustering, regression modeling, and expansion/adjustment factor accuracy. The data and scripts used in this paper are publicly available (Singleton, Runa, & Humagain, 2021).

2.1   Data sources and preparation

2.1.1   Pedestrian traffic signal data

Traffic signal controller log data from most of the over 2,000 signalized intersections in Utah were collected from UDOTs ATSPM system (UDOT, 2020) for one full year (July 2017 through June 2018). In total, data from 1,697 signals with pedestrian push-buttons were usable. The remainder of the signals either did not have pedestrian push-buttons (either in isolated rural/industrial locations or in the heart of downtown Salt Lake City) or were missing data for a significant portion of the year.

In order to prepare time series pedestrian datasets for clustering, a suitable metric that defines intersection-level "pedestrian activity" from traffic signal data was required. For this purpose, we relied upon the research by Singleton et al., (2020) that validated pedestrian push-button data against observed pedestrian counts using over 20,000 crossing-hours of observations in Utah. That research, using regression models and various fit statistics, determined that a new pedestrian activity metric of imputed pedestrian calls registered "45B" was the best predictor of actual pedestrian crossing volumes in many cases. More details about this validation and modeling process can be found elsewhere (Singleton et al., 2020; Singleton & Runa, 2021). Specifically, the 45B pedestrian activity metric is defined as:

- For each pedestrian phase, in a time-ordered sequence of traffic signal controller events with just events {0, 21, 90}, the number of 90 events immediately preceded by a 0 or 21 event, where:
    - Event 0, Phase On: This event occurs with the activation of the phase on, such as the start of green or the start of the walk interval.
    - Event 21, Pedestrian Begin Walk: This event occurs with the activation of the walk indication for a particular phase.
    - Event 90, Pedestrian Detector On: This event occurs when the signal from the pedestrian push-button is activated, after any delay or extension is processed, for a particular pedestrian detector channel.

In simple terms, the pedestrian activity metric 45B (imputed pedestrian calls registered) counts the number of times the walk signal appeared as a result of a pedestrian push-button press.

We analyzed two types of temporal patterns in pedestrian activity: hourly/weekday patterns, and monthly (or seasonal) patterns. For *hourly/weekday patterns*, we did the following for each intersection i: First, we calculated the pedestrian activity metric (45B) for all pedestrian phases over the entire year, removing any hours with missing data (i.e., due to communication outages or maintenance work). Sec-

ond, we averaged these year-long hourly observations into 168 values $v_{i,t,d}$, one for each of the unique hour-of-day t and day-of-week d combinations (e.g., 4-5pm Mondays). Third, we calculated normalized counts $\bar{v}_{i,t,d}$ according to the following equation (Griswold et al., 2018; Ryus et al., 2014):

$$\bar{v}_{i,t,d} = \frac{v_{i,t,d}}{\sum_{t=1}^{24} \sum_{d=1}^{7} v_{i,t,d}} \tag{1}$$

These normalized counts $\bar{v}_{i,t,d}$ are really the average hourly counts as a proportion of total average weekly counts of pedestrian activity at each signal, or essentially the (inverse of) hour-to-week expansion factors, specific to each intersection. By averaging across the entire year, this process mitigates some of the effects of temporal variation caused by special events, abnormal weather, or other unusual occurrences. These intersection-specific normalized counts (inverse expansion factors) of pedestrian activity (45B) were used as the data input into the empirical clustering analysis for hourly/weekday patterns.

For *monthly (seasonal) patterns*, we did the following for each intersection i: First, we took the whole-year hourly pedestrian activity (45B) dataset from the first step of the previous paragraph, and summed the hourly values to generate 365 daily totals. Second, for each month m, we calculated the monthly average daily volume ($d_{i,m}$); we also calculated the overall annual average daily volume ($y_{i,y}$). Third, we calculated the 12 monthly adjustment factors ($m_{i,m}$) according to the following equation:

$$m_{i,m} = \frac{d_{i,m}}{y_{i,y}} \tag{2}$$

These intersection-specific monthly adjustment factors were used as the data input into the empirical clustering analysis for monthly (seasonal) patterns.

2.1.2 Spatial data

To explore relationships between temporal patterns in pedestrian activity and spatial characteristics, we assembled land use, built environment, and socio-economic attributes for the area surrounding each signalized intersection. Specifically, measures were calculated using quarter-mile network buffers. Data came from various sources, including population and employment data from the 2013-2017 American Community Survey and the 2017 Longitudinal Employer-Household Dynamics dataset for Census block groups, as well as 2019 land use and transportation data from the Utah Automated Geographic Reference Center. (See Singleton, Park, & Lee (2021) for details on these data.) Due to a lack of data, this information was available for only 1,161 signals. Descriptive statistics for these attributes are presented in Table 1.

Table 1. Descriptive statistics for land use, built environment, and socio-economic attributes (*N* = 1,161)

Attribute	Mean	SD
Land-use attributes		
Residential land use (%)	31.264	22.006
Commercial land use (%)	30.756	19.360
Industrial land use (%)	2.007	8.190
Schools (#)	0.344	0.665
Places of worship (#)	0.593	0.850
Parks (acres)	1.537	3.622
Built environment attributes		
Population density (1,000/mi²)	5.263	2.923
Employment density (1,000/mi²)	8.365	12.888
Intersection density (#/mi²)	105.228	46.549
4-way intersections (%)	30.988	20.036
Transit stops (#)	43.065	23.491
Socio-economic attributes		
Vehicle ownership (#, mean)	1.646	0.418
Household size (#, mean)	2.972	0.850
Household income ($1,000)	57.655	20.152

2.2 Analysis methods

For each type of temporal pattern (hourly/weekday, monthly), we conducted a series of analyses: (1) empirical cluster analysis to identify clusters of signals with similar temporal patterns; (2) multinomial logit regression modeling to understand spatial factors associated with temporal clusters; and (3) calculating the accuracy of applying the expansion/adjustment factors.

2.2.1 Identifying temporal patterns using empirical cluster analysis

Critical steps in the cluster analysis process—selecting a (dis)similarity measure, choosing a clustering algorithm, and determining an optimal number of clusters—are discussed in this section.

Because the objective of this study focuses on grouping intersections based on similar hourly/weekly patterns, we used a structural-based (dis)similarity measure—temporal correlation (CORT)—since it allows us to compare the relative trajectories of normalized counts across intersections. Basically, CORT measures the proximity of temporal variation between two time series, which aligns better with our objective than other conventional distance measures such as Euclidean distance, which works with difference in magnitude between data points (Montero & Vilar, 2014). The equation for CORT is:

$$d_{CORT}\left(F_i, F_j\right) = \frac{\sum_{t=1}^{T-1}\left(F_{i(t+1)} - F_{it}\right)\left(F_{j(t+1)} - F_{jt}\right)}{\sqrt{\sum_{t=1}^{T-1}\left(F_{i(t+1)} - F_{it}\right)^2}\sqrt{\sum_{t=1}^{T-1}\left(F_{j(t+1)} - F_{jt}\right)^2}} \tag{3}$$

where F_i and F_j represent two time series i and j, measured over T time points t.

In terms of clustering algorithms, we applied a k-means algorithm, which basically has the objective of minimizing differences within each cluster and maximizing differences between other clusters. The k-means algorithm is found to be computationally efficient and can group intersections based on subtle nuances in temporal patterns of pedestrian activity.

The final step in cluster analysis is to determine the optimal number of clusters (between one and the number of observations) that adequately represent patterns within a dataset. Common tools to assist in the selection of the number of clusters include the Calinski-Harabasz criterion, the sum of squared differences, and the average silhouette width:

- Calinski-Harabasz (CH) criterion: This is the ratio of between-cluster variation to within-cluster variation, so a higher value reflects distinct clusters.
- Sum of squared differences (SSD): This calculates the sum of squared differences between each observation's values and the mean values of the cluster to which the observation belongs. Smaller differences indicate more homogenous clusters.
- Average silhouette width (ASW): This measures the similarity (ranging from –1 to +1) of patterns of observations within each cluster (cohesion) compared to observations of other clusters (separation). A more positive value implies that observations are well matched within clusters and poorly matched to neighboring clusters.

By performing k-means clustering for various numbers of clusters (2 to 7 for hourly/weekday, 2 to 5 for monthly), we computed the various fit statistics and also visualized the patterns of resulting clusters. Generally, Table 2 shows that CH, SSD, and ASW values decreased with increased numbers of clusters, although not exclusively so. More clusters generally means an increase in both the distinctiveness and compactness of clusters; hence, the optimum number of clusters can be the number at which more/fewer clusters provides neither a significant improvement nor degradation in the fit statistics. Hence, the optimum number of hourly/weekday clusters was determined to be five, as it provides reasonable fit statistics and a relatively lower decrease in SSD (compared to six clusters). Similarly, the optimum number of monthly clusters was determined to be three, as it provides a reasonable tradeoff between low SSD and higher ASW.

Table 2. Fit statistics for various numbers of clusters

	Hourly/weekday clusters			Monthly (seasonal) clusters		
# clusters	CH	SSD	ASW	CH	SSD	ASW
2	1,424.71	4,715.61	0.72	700.38	668.42	0.54
3	1,082.93	3,226.85	0.54	**585.93**	**539.92**	**0.40**
4	1,052.84	3,612.17	0.54	539.65	475.59	0.30
5	**985.61**	**2,285.66**	**0.57**	502.74	307.44	0.27
6	973.47	1,970.04	0.31			
7	995.75	1,520.18	0.28			

2.2.2   Predicting clusters with spatial factors using multinomial logit regression models

In order to explain which signals belonged in each cluster on temporal patterns of pedestrian activity, we performed multinomial logit regression using the spatial factors shown in Table 1 (land use, built environment, and socio-economic attributes) as explanatory variables. Variables with statistically significant coefficients indicate spatial characteristics associated with signals having different temporal patterns in pedestrian activity.

2.2.3   Assessing the accuracy of expansion/adjustment factors

When applying the cluster results and expansion/adjustment factors to convert short-duration counts to longer-term average volumes, there will be some discrepancy even with the same data due to using expansion/adjustment factors based on the cluster mean values. For hourly/weekday factors, this expansion accuracy is expressed as the absolute percentage error of the expanded weekly counts at a location relative to the average expanded weekly counts of the cluster to which that location belongs. As presented in Griswold et al. (2018) and Medury et al. (2019), the expansion accuracy $\epsilon_{i,t,d}^c$ for hourly/weekday expansion factors is given by the following equation:

$$\epsilon_{i,t,d}^c = \left| \frac{\bar{v}_{i,t,d}}{\gamma_{i,t,d}^c} \right| \times 100 \qquad (4)$$

where: $\bar{v}_{i,t,d}$ is the normalized count and $\gamma_{i,t,d}^c$ is the applicable expansion factor for cluster c, location i, and time period t. A similar equation applies to the accuracy of monthly adjustment factors.

3   Results

3.1   Hourly/weekday patterns

3.1.1   Results of hourly/weekly clusters

To recap, we used the CORT (dis)similarity measure and the k-means algorithm to classify the normalized counts of pedestrian activity at 1,697 signalized intersections into five clusters. The cluster analysis results are summarized in Table 3 and the text below, and the mean and distributions of the hourly/weekly patterns are depicted in Figure 1.

Table 3. Summary of hourly/weekly cluster results

Pattern	Cluster (#, %)	Visual characteristics
Uniform	1 (871, 51.3)	Evening peak, increase from morning to evening, weekdays > weekends, peak hour volume ~1-1.5% of weekly volume
	2 (278, 16.4)	Evening peak, increase from morning to evening, weekdays > weekends, peak hour volume ~1-1.5% of weekly volume
Bimodal	3 (302, 17.8)	Morning and evening peaks, evening > morning, weekdays > weekends, peak hourly volume ~1.5-2% of weekly volume
	4 (188, 11.1)	Morning and evening peaks, evening > morning, weekdays > weekends, peak hourly volume ~1.5-2% of weekly volume
	5 (58, 3.4)	Morning and evening peaks, evening > morning, weekdays > weekends, peak hourly volume ~2-2.5 of weekly volume

Overall, the hourly/weekday clusters can be classified into two general patterns of pedestrian activity: (a) unimodal, with one (usually evening) peak hour that is approximately 1-1.5% of the weekly total, and (b) bimodal, with two distinct peak hours (usually evening is greater than morning) and where the (usually evening) peak hour is approximately 1.5-2% of the weekly total. Besides these general observations, the clusters themselves show some (albeit more minor) differences. Unimodal clusters 1 and 2 are slightly differentiated in their daytime vs. evening patterns: for cluster 1, the pattern is somewhat more uniform (or smooth) than for cluster 2, and the mean is slightly more peaked. The bimodal clusters 3,

4, and 5 are distinguished by the magnitude of their peaks—cluster 5's peaks are more than 2% of the weekly total, whereas peaks for clusters 3 and 4 are 1.5-2%—and somewhat by the difference between the morning and evening peaks (difference: cluster 5 > cluster 4 > cluster 3).

There were also some similarities between all the hourly/weekday clusters. Unsurprisingly, pedestrian activity was highest during daytime and evening hours, with most intersections recording little-to-no activity overnight. Peak pedestrian hours were more common in the afternoon and early evening than in the morning. Weekend pedestrian activity (especially on Sundays) was lower than on weekdays, but often without a clear single peak hour (usually mid-day). Tuesdays often had the largest peak hour of pedestrian activity, while Mondays and Fridays tended to have slightly lower peaks than other weekdays (although, Mondays had the highest peaks for clusters 3 and 4).
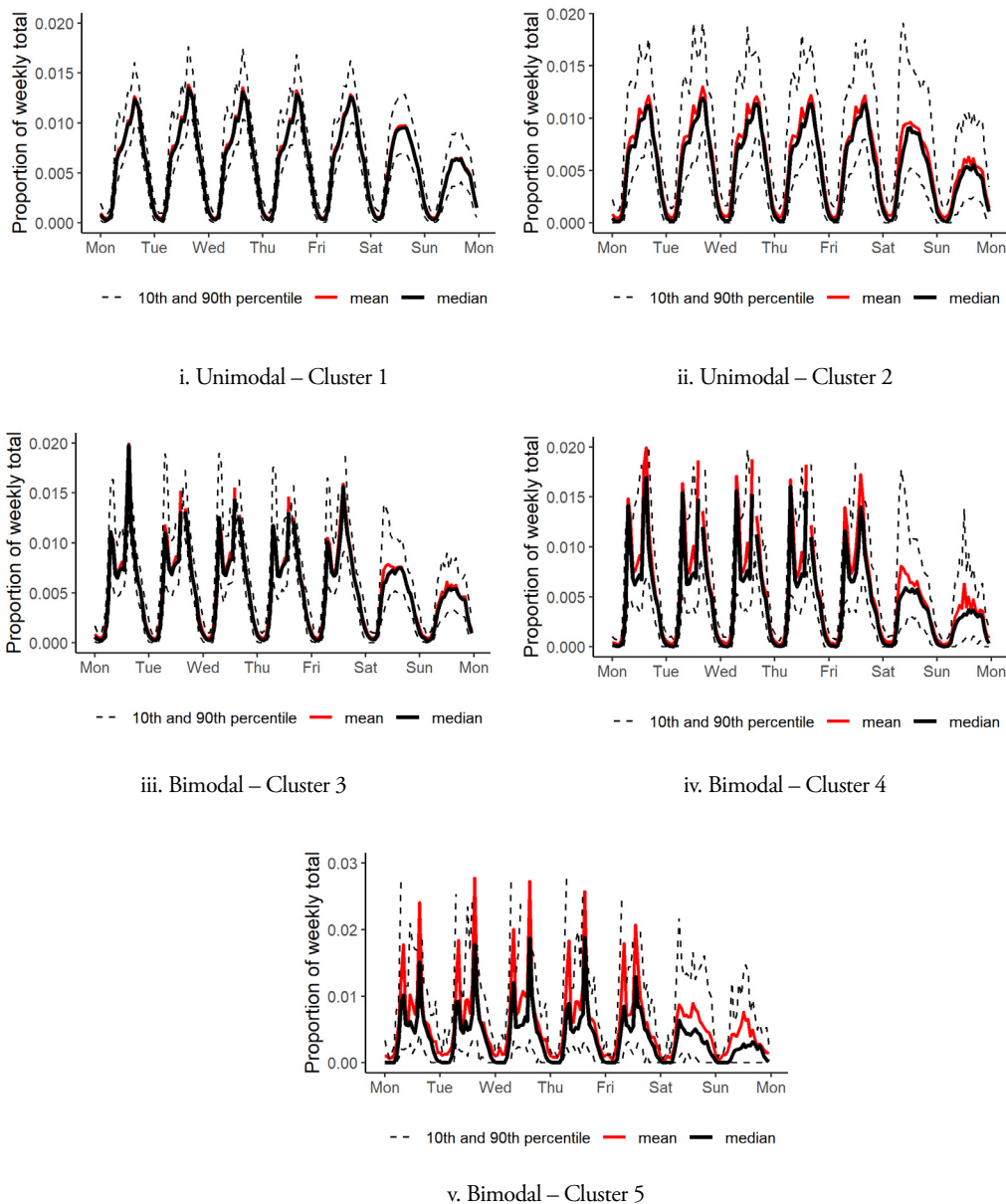

i. Unimodal – Cluster 1

ii. Unimodal – Cluster 2

iii. Bimodal – Cluster 3

iv. Bimodal – Cluster 4

v. Bimodal – Cluster 5

Figure 1. Mean and distribution of pedestrian activity patterns by hourly/weekday cluster

3.1.2   Spatial factors affecting hourly/weekly patterns

Some past studies have investigated the influential role of land use, the built environment, and socio-economic characteristics in shaping temporal patterns of pedestrian activity across locations (e.g., Hankey et al., 2012; Medury et al., 2019; Schneider et al., 2009). The presence of offices, schools/colleges, and different land-use characteristics surrounding count locations are found to influence the hourly/weekday temporal patterns. For instance, locations near schools displayed multiple peaks on weekdays and relatively lower pedestrian activity during weekends, whereas recreational trails had higher activity during weekday evenings and relatively higher pedestrian activity during weekends (Medury et al., 2019). We add to this literature using our larger dataset of over 1,000 signalized intersections in Utah.

To understand the relationships between pedestrian activity patterns and spatial characteristics, we estimated a multinomial logit model on 1,161 signalized intersections with such data, where membership in an hourly/weekly cluster (1 to 5) was the dependent variable and spatial characteristics were the independent variables. Results are presented in Table 4 and described below.

Table 4. Multinomial logit model results of hourly/weekday cluster membership

	Cluster-specific coefficients (ref. = 1)			
Variable	*2*	*3*	*4*	*5*
Intercept	0.751	1.409	0.006	-0.576
Population density (1,000/mi²)	**-0.234**	**-0.240**	**-0.373**	
Residential land use (%)	**-0.035**			
Commercial land use (%)	**-0.046**	**-0.053**	**-0.055**	
Industrial land use (%)		**-0.034**		**0.058**
Intersection density (#/mi²)		**-0.005**	**-0.010**	**-0.019**
4-way intersections (%)		**-0.017**	**-0.018**	**-0.039**
Schools (#)		**0.515**	**1.039**	**1.166**
Vehicle ownership (#, mean)	**-0.555**			
Household size (#, mean)	**0.275**		*0.325*	
Household income ($1,000)	**0.021**	**0.012**	**0.016**	
McFadden pseudo-R²	0.170			
Sample size (N)	1,161			

Statistical significance: **bold** for *p* < 0.05, *italics* for *p* < 0.10, not shown for p > 0.10.

The model results help to explain why we see some of the differences in the pedestrian activity patterns across clusters. Notably, the bimodal patterns (multiple peaks) of clusters 3, 4 and 5 is partially explained by the result that these locations were much more likely to located within a quarter-mile walking distance of one or more schools, indicated by the significant positive coefficients for number of schools. Signals were also more likely to have a bimodal pattern in areas with less street network connectivity, as shown by the significant negative coefficients for intersection density and percentage of four-way intersections. Based on the results for population density, percentage of commercial land use, and household income, signals were more likely to have the smooth unimodal pattern of cluster 1 when they were in areas with greater population density, more commercial land uses, and lower household incomes. Looking at differences between unimodal signals, belonging to cluster 1 was more likely in neighborhoods with more residential land uses, greater vehicle ownership, and smaller household sizes.

3.1.3   Expansion factor accuracy

Figure 2 displays the average expansion accuracy for each hour in the week by hourly/weekly cluster. Overall, expansion accuracy is greater (lower error) for clusters and during times with higher pedestrian activity levels. Average error is less than 75-100% for all clusters when counts are taken during daytime hours, but greater than 75% (and as much as 150%) when overnight counts are expanded. Expansion errors for daytime counts at signals in clusters 1 and 3 (the largest clusters) are around 25% or less, suggesting that only a few hours of counts at these locations may be enough to accurately estimate longer-term pedestrian volumes. Conversely, counts taken at signals in the smallest cluster (5) may need to be of a longer duration in order to produce similarly accurate estimates of pedestrian activity.
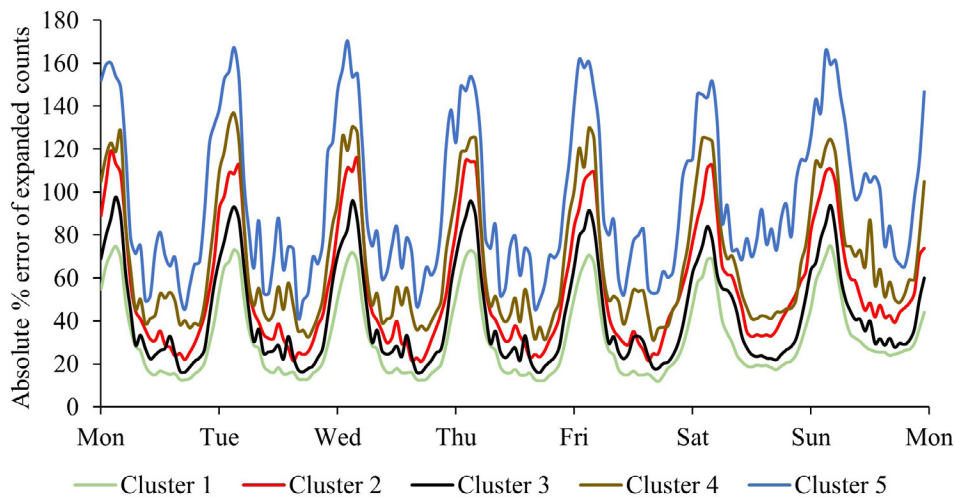
Figure 2. Expansion accuracy for hourly/weekday clusters

3.2   Monthly patterns

3.2.1   Seasonal (monthly) clusters

The normalized counts (inverse expansion factors) for the hourly/weekday clusters shown in Figure 1 depict average hourly and weekday pedestrian activity patterns expressed as a proportion of weekly totals. Homogeneity within those clusters may obscure other sources of temporal variations in pedestrian activity patterns between locations, such as those differences due to seasonal variation. In fact, factoring processes in traffic monitoring to convert short-duration counts to annual average daily volumes require seasonal adjustment factors as well.

Therefore, we performed a similar empirical clustering process to generate monthly clusters of similar seasonal pedestrian activity patterns. In traffic monitoring, such seasonal variation in activity at intersections is addressed during the calculation of annual volume by using monthly adjustment factors. After calculating the monthly adjustment factors as described in section 2.1.1, some of the intersections were removed due to unusually high factors above 3.0, which could have resulted from missing data or technical errors, resulting in a total of 1,446 intersections. To recap, the optimum number of monthly (seasonal) clusters was determined to be three (based on fit statistics and visualization).

The mean values of the adjustment factors for the three monthly clusters are shown in Figure 3. On average, the largest cluster A (1,076, 74.4%) has the least variation in pedestrian activity patterns from month-to-month, peaking in the summer months (especially June, but with a slight decrease in July)

and bottoming out in the winter months (especially December). Higher pedestrian activity during the summer is warranted because of pleasant weather (warm and dry) throughout most of Utah. Similarly, cold temperature and snow conditions result in lower levels of pedestrian activity during Utah winter months. The other two clusters B and C show greater (and different) variations on this trend. Signals in cluster B (111, 7.7%) have sort of a more extreme version of cluster A's pattern, with even higher relative volumes in the summer (from June through September) and lower volumes in winter (especially February). The pattern suggests that pedestrian activity for signals in cluster B is more sensitive to weather. Conversely, cluster C (259, 17.9%) shows a distinctly different pattern, peaking in September and having less-than-average pedestrian activity during the summer months (June through August). Signals in cluster C could be near schools or universities, which are usually not in session during these three summer months. These hunches about the reasons motivating seasonal variations in pedestrian activity could be confirmed through comparisons with spatial characteristics, as presented in the following section.
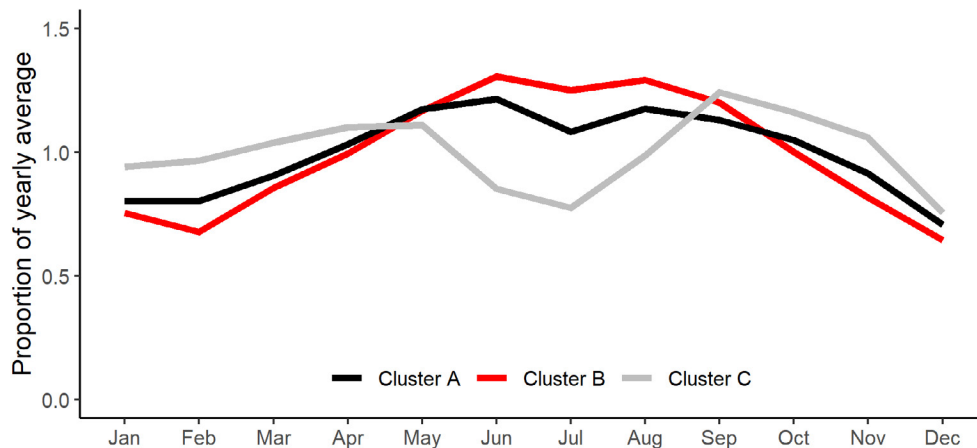

Figure 3. Means of pedestrian activity patterns by monthly cluster

3.2.2   Spatial factors affecting seasonal (monthly) patterns

As described earlier, seasonal variation in pedestrian activity patterns are mostly influenced by climatic conditions (i.e., snow, temperature, rainfall) (Runa, 2020). To help explain these seasonal variations in Utah, we estimated another multinomial logit model on 1,161 signalized intersections, this time where the dependent variable was membership in a monthly cluster (A, B, or C). In addition to the same land use, built environment, and socio-economic characteristics as used previously, we also added climatic divisions classifications from the National Climate Data Center. The assumption is that signals in the same climatic division experience similar weather patterns throughout the year. As shown in Figure 4, Utah contains seven climatic divisions, although most signals in our dataset are located in the North Central region (also known as the Wasatch Front), with some in the Northern Mountains and Dixie regions. (We did not have spatial characteristics data for signals in the other climatic divisions.) Note that although many of the signals appear to lie along the border of the North Central and Northern Mountains divisions, almost all are truly located in the urbanized valleys of Utah's Wasatch Front (such as in Salt Lake City). These signals experience weather patterns that are much more similar to each other than they are to higher-elevation locations in the Northern Mountains (such as in Park City).
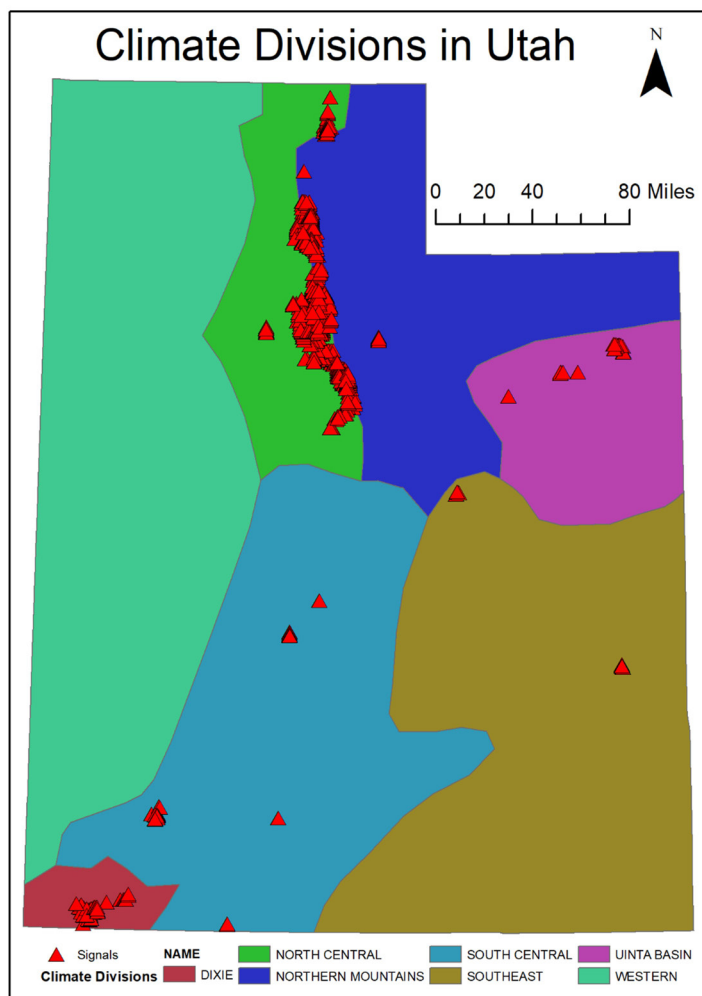
Figure 4. Map of signalized intersections and climate divisions

The results of the multinomial logit model for monthly clusters are shown in Table 5; the largest cluster A is the reference alternative. Signals were less likely to be in cluster A and more likely to be in clusters B or C if they were in areas with greater employment density, lower traffic volumes, and larger household sizes. Cluster B was associated with less commercial land uses, greater intersection density, and fewer places of worship. Signals in the Dixie climatic region were much more likely to belong to cluster B or C than to A. Typically, this region experiences hotter summers and milder winters than other areas of Utah, which could explain the greater monthly variation for cluster B and the summer trough of cluster C. Also, signals near schools were much more likely to be in cluster C, which supports our hypothesis about a lack of school attendance being an explanation for lower pedestrian activity levels in summer months.

Table 5. Multinomial logit model results of monthly cluster membership

Variables	*Cluster-specific coefficients (ref. = A)*	
	B	*C*
Intercept	*-2.233*	**-3.521**
Climate division: Dixie		
(ref. = North Central)	**2.852**	**4.666**
Employment density (1,000/mi²)	**0.029**	**0.046**
Commercial land uses (%)	**-0.031**	
4-way intersections (%)	**0.019**	
Schools (#)		**0.734**
Places of worship (#)	*-0.399*	
Vehicle ownership (#, mean)	**-1.001**	**-0.880**
Household size (#, mean)	**0.722**	**1.053**
McFadden pseudo-R²	0.186	
Sample Size (N)	1,161	

Statistical significance: bold for $p < 0.05$, *italics* for $p < 0.10$, not shown for $p > 0.10$.

3.2.3   Adjustment factor accuracy

The average error (expansion/adjustment accuracy) for the three monthly clusters is shown in Figure 5. Accuracy is remarkably good (around 10% error) for the largest cluster A all year round, but even cluster C has reasonably good expansion/adjustment accuracy (10-20% error) throughout most of the year. The smallest cluster B shows the potential for more error (30-40%). As with the hourly/weekday clusters, accuracy for the monthly clusters tends to be worse during months with lower levels of pedestrian activity (December for all clusters, July for cluster C).
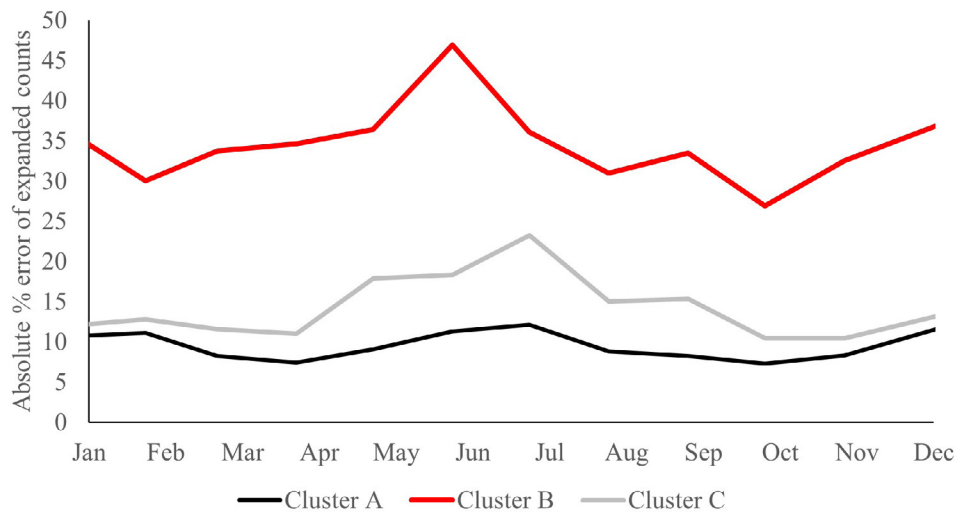
Figure 5. Expansion accuracy for monthly clusters

3.3   Cross-classification of hourly/weekday and monthly clusters

As hourly/weekday and monthly pedestrian activity pattern trends differ in nature (due to variations in climatic conditions and some spatial characteristics), it is useful to examine the frequency of signals in each of the hourly/weekday clusters that belonged to a particular seasonal trend produced by the monthly clusters. Therefore, we cross-tabulated the number of intersections that belonged to each combination of hourly/weekday and monthly clusters. Table 6 indicates that most unimodal hourly/weekday clusters (1 and 2) pertain to the more uniform seasonal variations of cluster A. However, most of intersections that belonged to the bimodal hourly/weekday clusters 4 and 5 (those with high relative peak pedestrian activity levels) were grouped under monthly cluster C, where pedestrian activity dips during summer months. This makes sense, given that a bimodal daily pattern and a summer lull in pedestrian activity are both indicative of school-driven pedestrian activity patterns.

Table 6. Cross-classification of hourly/weekday and monthly clusters (N = 1,446)

Hourly/weekday clusters	Monthly Clusters		
	A	B	C
1	661	38	85
2	152	36	39
3	199	11	50
4	53	19	70
5	11	7	15

4   Discussion and conclusions

In this study, we provided an empirical clustering approach to grouping locations with similar long-term pedestrian activity patterns using pedestrian push-button data from over 1,500 signalized intersections in Utah. After calculating the proxy measure of pedestrian activity (imputed pedestrian calls registered), we performed cluster analysis to: classify signals based on the normalized hourly/weekly counts (each hour's proportion of weekly totals, or the inverse of the expansion factors), and account for seasonal variation using monthly adjustment factors. We also used multinomial logit models to identify spatial and climatic characteristics (land use, built environment, and socio-economic characteristics, as well as climatic regions) that help predict and explain which locations see different temporal patterns in pedestrian activity levels. Finally, we assessed the accuracy of applying the expansion/adjustment factors.

The ultimate objective of this work was to investigate the temporal patterns of pedestrian activity and develop expansion/adjustment factors and factor groups that relate to spatial characteristics. Utilizing the hourly/weekday and monthly clusters that we developed, each with an average temporal pattern, one can expand a short-duration count (for a specific hour, weekday, and month) through multiplication and/or division to get an estimate of the long-term average pedestrian volume at a particular location. In the process of achieving this aim, our paper made several contributions to travel monitoring for pedestrian travel.

- Most notably, we utilized a much greater quantity of pedestrian data than has been possible to examine before, specifically one year of data from 1,697 signalized intersections throughout Utah. This larger sample size allowed us to examine more nuanced differences in pedestrian activity patterns and have the power to identify significant associations with spatial characteristics.
- Our use of separate hourly/weekday and monthly clusters allowed us to distinguish the influences of time-of-day and day-of-week from seasonal variations.

- We also considered the expansion/adjustment accuracy for our factor groups (clusters), which has implications for the selection of times-of-day and durations for short-term pedestrian counts.
- In the remaining section of this paper, we discuss key findings, their applications for understanding pedestrian behavior and monitoring pedestrian traffic, and study limitations.

Overall, the clustering of average hourly proportions of weekday counts revealed common hourly/weekday pedestrian activity patterns at most signalized intersections, as shown for the five clusters altogether in Figure 6. All clusters saw their highest pedestrian volumes during weekday daytime hours, with peak pedestrian volumes in the afternoon or early evening, lower volumes on weekends, and slightly lower volumes on Mondays and Fridays. In fact, the differences between clusters were more nuanced. Three clusters (3, 4, and 5) showed bimodal morning and evening peaks, with the other two clusters (1 and 2) having just a single evening peak. The weekday peak hours varied from 1 to 2.5 percent of weekly totals, depending on the cluster. The fact that the clustering algorithm picked up even these small differences in temporal patterns highlights the utility of an empirical data-driven approach to constructing pedestrian factor groups. On the other hand, the small differences between locations shown here suggests that a coarser factor grouping might not result in significantly inferior count expansion results.

Compared to past research defining pedestrian or non-motorized count factor groups (Medury et al., 2019; Miranda-Moreno & Lahti, 2013; Schneider et al., 2009), the hourly/weekday factor groups (clusters) we identified are not as distinct. This is likely due to several factors, most notably the limitation of our source data. Research relying on permanent non-motorized counters can cover a wider range of types of temporal use patterns—commuting vs. recreational vs. mixed—precisely because the locations where counters are deployed were selected to capture a variety of behaviors and uses. In this study, we relied upon existing infrastructure (traffic signals) in locations that were not chosen with pedestrian count programs in mind. In other words, location selection was exogenous to our study purpose. Since most traffic signals are located in areas with higher traffic volumes (such as along arterials) or where walking for utilitarian/transportation purposes is expected, they may not be able to detect the full variety of pedestrian temporal patterns that exist, such as along trails or in recreational areas.
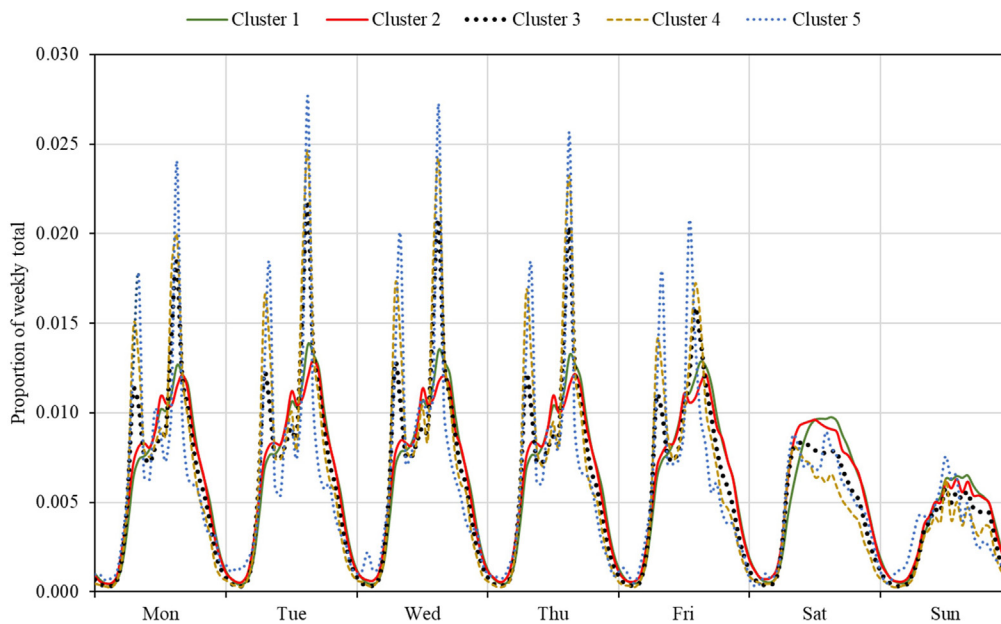

Figure 6. Means of pedestrian activity patterns by hourly/weekday cluster: 1 ($N = 871$), 2 ($N = 278$), 3 ($N = 302$), 4 ($N = 188$), and 5 ($N = 58$)

Despite the less noticeable differences between hourly/weekday pedestrian activity patterns in different clusters, the model illuminated land use, built environment, and socio-economic differences that help to explain these pattern variations. Notably, having schools nearby significantly increased the chances of a signal having a bimodal pedestrian activity pattern, coincident with the morning and afternoon/early evening time periods bracketing the school day. Several other built environment attributes—population density, residential and commercial land uses, connected street networks—were linked to signals belonging to cluster 1, with unimodal, smoother, and less varied weekday pedestrian activity patterns. This information about spatial characteristics—along with insight into expansion accuracy—aids planners and pedestrian traffic monitoring program managers by suggesting the types of locations where different expansion/adjustment factors are needed or even where longer or shorter short-duration counts are needed to provide accurate pedestrian volume estimates.

Our clustering of monthly adjustment factors showed more significant differences in the seasonal patterns of pedestrian activity across intersections (see Figure 3) than were found for hourly/weekday patterns. Specifically, we captured both the general seasonal trend (higher pedestrian activity in summer, lower activity during winter months) and also trends specific to certain locations, such as the drop in pedestrian activity during out-of-school months from June through August. By linking these monthly groupings to spatial characteristics, we confirmed that this latter specific pattern occurred more often near schools and universities. We also demonstrated that pedestrian activity was more sensitive to weather in certain regions (southwestern Utah) with higher summer temperatures and mild winters.

Another finding of this study is about expansion factor accuracy. Confirming past research, our results suggest that the expansion accuracy is cyclical in nature, with higher errors during low-volume overnight hours and greater accuracy at daytime. This implies that manual counts should be conducted at intersections during the daytime, and longer counts may be beneficial, especially during off-peak hours and at locations with more variability (i.e., cluster 5). The expansion accuracy from empirical clustering should be more accurate than the "single factor" approach of having just one factor group, as shown by previous studies (Griswold et al., 2018; Medury et al., 2019).

4.1   Limitations

The study is not without additional limitations. First, the pedestrian activity metrics derived from pedestrian push-button data does not provide the actual pedestrian volume, and may contain errors, because of imperfect correlation and non-linearities. However, correlation between push-button data and volumes is high (Singleton et al., 2020; Singleton & Runa, 2021), and we suspect the benefits of being able to analyze data from hundreds if not thousands of locations outweigh the inaccuracies of the source data. Second, the clustering approach produced five clusters, but in reality, there were only two distinct groupings (unimodal and bimodal). However, the tradeoff between cluster fit and the number of clusters is difficult to control for, and empirical clustering does offer the benefit of identifying smaller differences that may be obscured using a different approach. Third, the sample size decreased somewhat when incorporating spatial characteristics, which might cause bias or lack of generalizability of the spatial analysis results. Fourth, the relatively coarse nature of the climatic divisions used in the study could not exactly pin-point what causes the variation in seasonal patterns (i.e., snow, rainfall, wind), and there may be additional influential weather variations within each climatic division. Fifth, the use of empirical clustering and Utah-specific climate zones may limit the generalizability of these temporal patterns to areas outside of Utah. Nevertheless, a similar process may be useful for developing pedestrian expansion and adjustment factors in other states. Despite these limitations, in summary, we have demonstrated that traffic signals with pedestrian push-button data are a very useful supplement to—but not a complete replacement for—existing permanent counters within a broader pedestrian traffic monitoring program.

4.2   Future research

There are many opportunities to refine this research or extend it in new directions. Future research could look at using alternative data sources (such as the Google Places API) to calculate the attributes near intersections, which could provide more detailed insights into specific land uses or place types than some of the aggregate metrics used in our study. Correlating traffic signal-based pedestrian activity levels with weather is a potential fruitful area of inquiry. In this regard, more fine-grained data—about temperature, humidity, snow, air quality, etc.—collected from nearby weather stations could be assembled and correlated with pedestrian activity patterns at intersections (Runa & Singleton, 2021). If there are common patterns in how pedestrian activity changes when it snows, rains, etc., it may be possible to develop expansion/adjustment factors that work for short-duration counts conducted during mildly-inclement weather. Additionally, another promising area for future research could be the investigation of the effects of major events such as concerts or sporting events on changes in pedestrian activity compared to normal days. More research could also be done using these pedestrian traffic signal data to inform the duration and timing of short-term pedestrian counts. For example, one could extend this study to determine the average expansion/adjustment accuracy of different count durations (anywhere from one hour to one week), in an attempt to find the optimum tradeoff between cost and accuracy. Finally, research also could look at using pedestrian push-button data to calculate and compare other types of traffic monitoring count data expansion factors, such as hour-to-year or day-to-year expansion factors, as these have been suggested as potentially more accurate alternatives for estimating annual average volumes (Medury et al., 2019).

References

ATKINS. (2016). *Automated traffic signal performance measures reporting details. Atlanta, GA: Georgia Department of Transportation.* Retrieved from https://udottraffic.utah.gov/ATSPM/Images/ATSPM_Reporting_Details.pdf

Blanc, B., Johnson, P., Figliozzi, M., Monsere, C., & Nordback, K. (2015). Leveraging signal infrastructure for nonmotorized counts in a statewide program: Pilot study. *Transportation Research Record: Journal of the Transportation Research Board, 2527,* 69–79. https://doi.org/10.3141/2527-08

Bu, F., Greene-Roesel, R., Diogenes, M. C., & Ragland, D. R. (2007). *Estimating pedestrian accident exposure: Automated pedestrian counting devices report.* Berkeley, CA: UC Berkeley Traffic Safety Center. Retrieved from https://escholarship.org/uc/item/0p27154n

Day, C. M., Premachandra, H., & Bullock, D. M. (2011). *Rate of pedestrian signal phase actuation as a proxy measurement of pedestrian demand.* Paper presented at the 90th Annual Meeting of the Transportation Research Board, Washington, DC. Retrieved from https://docs.lib.purdue.edu/civeng/24/

Day, C. M., Bullock, D. M., Li, H., Remias, S. M., Hainen, A. M., Freije, R. S., ... & Brennan, T. M. (2014). *Performance measures for traffic signal systems: An outcome-oriented approach.* West Lafayette, IN: Purdue University. Retrieved from https://doi.org/10.5703/1288284315333

Day, C. M., Taylor, M., Mackey, J., Clayton, R., Patel, S. K., Xie, G., ... & Bullock, D. (2016). Implementation of automated traffic signal performance measures. *ITE Journal, 86*(8), 26–34. https://trid.trb.org/view/1418795

Diogenes, M. C., Greene-Roesel, R., Arnold, L. S., & Ragland, D. R. (2007). Pedestrian counting methods at intersections: A comparative study. *Transportation Research Record: Journal of the Transportation Research Board, 2002*(1), 26–30. https://doi.org/10.3141/2002-04

Federal Highway Administration (FHWA). (2016). *Traffic monitoring guide.* Washington, DC: U.S. Department of Transportation. Retrieved from https://www.fhwa.dot.gov/policyinformation/tmguide/

Greene-Roesel, R., Diogenes, M. C., Ragland, D. R., & Lindau, L. A. (2008). *Effectiveness of a commercially available automated pedestrian counting device in urban environments: Comparison with manual counts.* Presented at the 87th Annual Meeting of the Transportation Research Board, Washington, DC. Retrieved from https://escholarship.org/uc/item/2n83w1q8

Griswold, J. B., Medury, A., Schneider, R. J. & Grembek, O. (2018). Comparison of pedestrian count expansion methods: Land-use groups versus empirical clusters. *Transportation Research Record: Journal of the Transportation Research Board, 2672*(43), 87–97. https://doi.org/10.1177/0361198118793006

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. *Landscape and Urban Planning, 107*(3), 307–316. https://doi.org/10.1016/j.landurbplan.2012.06.005

Kothuri, S., Nordback, K., Schrope, A., Phillips, T., & Figliozzi, M. (2017). Bicycle and pedestrian counts at signalized intersections using existing infrastructure: Opportunities and challenges. *Transportation Research Record: Journal of the Transportation Research Board, 2644,* 11–18. https://doi.org/10.3141/2644-02

Medury, A., Griswold, J. B., Huang, L. & Grembek, O. (2019). Pedestrian count expansion methods: Bridging the gap between land-use groups and empirical clusters. *Transportation Research Record: Journal of the Transportation Research Board, 2673*(5), 720–730. https://doi.org/10.1177/0361198119838266

Miranda-Moreno, L. F., & Lahti, A. C. (2013). Temporal trends and the effect of weather on pedestrian volumes: A case study of Montreal, Canada. *Transportation Research Part D: Transport and Environment, 22,* 54–59. https://doi.org/10.1016/j.trd.2013.02.008

Montero, P., & Vilar, J. A. (2014). TSclust: An R package for time series clustering. *Journal of Statistical Software, 62*(1), 1–43. https://doi.org/10.18637/jss.v062.i01

Runa, F. (2020). *The effect of weather on pedestrian activity at signalized intersections in Utah* (master's thesis). Logan, UT: Utah State University. https://doi.org/10.26076/cdbb-1171

Runa, F., & Singleton, P. A. (2021). Assessing the impacts of weather on pedestrian signal activity at 49 signalized intersections in Northern Utah. *Transportation Research Record: Journal of the Transportation Research Board, 2675*(6), 406–419. https://doi.org/10.1177/0361198121994111

Ryus, P., Ferguson, E., Laustsen, K. M., Proulx, F. R., Schneider, R. J., Hull, T., & Miranda-Moreno, L. (2014). *Methods and technologies for pedestrian and bicycle volume data collection* (NCHRP web-only document 205). Washington, DC: Transportation Research Board. Retrieved from https://doi.org/10.17226/23429

Ryus, P., Butsick, A., Proulx, F. R., Schneider, R. J., & Hull, T. (2017). *Methods and technologies for pedestrian and bicycle volume data collection: Phase 2* (NCHRP Web-Only Document 205). Washington, DC: Transportation Research Board. Retrieved from https://doi.org/10.17226/24732

Schneider, R. J., Arnold, L. S., & Ragland, D. R. (2009). Methodology for counting pedestrians at intersections: Use of automated counters to extrapolate weekly volumes from short manual counts. *Transportation Research Record: Journal of the Transportation Research Board, 2140*(1), 1–12. https://doi.org/10.3141/2140-01

Singleton, P. A., Park, K., & Lee, D. H. (2021). Varying influences of the built environment on daily and hourly pedestrian crossing volumes at signalized intersections estimated from traffic signal controller event data. *Journal of Transport Geography, 93,* 103067. https://doi.org/10.1016/j.jtrangeo.2021.103067

Singleton, P. A., & Runa, F. (2021). Pedestrian traffic signal data accurately estimates pedestrian crossing volumes. *Transportation Research Record: Journal of the Transportation Research Board, 2675*(6), 429–440. https://doi.org/10.1177/0361198121994126

Singleton, P. A., Runa, F., & Humagain, P. (2020). *Utilizing archived traffic signal performance measures for pedestrian planning and analysis.* Taylorsville, UT: Utah Department of Transportation. https://rosap.ntl.bts.gov/view/dot/54924

Singleton, P., Runa, F., & Humagain, P. (2021). *Singletonpa/ped-signal-data* [data set]. Zenodo. Retrieved from https://doi.org/10.5281/zenodo.4759088

Smaglik, E. J., Sharma, A., Bullock, D. M., Sturdevant, J. R., & Duncan, G. (2007). Event-based data collection for generating actuated controller performance measures. *Transportation Research Record: Journal of the Transportation Research Board, 2035*(1), 97–106. https://doi.org/10.3141/2035-11

Sturdevant, J. R., Overman, T., Raamot, E., Deer, R., Miller, D., Bullock, D. M., ... & Remias, S. M. (2012). *Indiana traffic signal hi resolution data logger enumerations.* West Lafayette, IN: Purdue University. http://doi.org/10.4231/K4RN35SH

Utah Department of Transportation (UDOT). 2020. *Automated traffic signal performance measures, UDOT.* Retrieved from https://udottraffic.utah.gov/ATSPM/