ACTIVE TRANSPORTATION COUNTS FROM EXISTING ON-STREET SIGNAL AND DETECTION INFRASTRUCTURE

Final Report

SPR 857

Oregon Department of Transportation

ACTIVE TRANSPORTATION COUNTS FROM EXISTING ON-STREET SIGNAL AND DETECTION INFRASTRUCTURE

Final Report

SPR 857

By

Sirisha Kothuri, Ph.D., Senior Research Associate
Elizabeth Yates, Graduate Research Assistant
Joseph Broach, Ph.D., Research Associate
Portland State University
P.O. Box 751
Portland, OR 97207

Patrick Singleton, Ph.D., Assistant Professor
Mahyar Vahedi Saheli, Graduate Research Assistant
Utah State University
4110 Old Main Hill
Logan, UT 84321

for

Oregon Department of Transportation
Research Section
555 13th Street NE, Suite 1
Salem OR 97301

and

Federal Highway Administration
1200 New Jersey Avenue SE
Washington, DC 20590

December 2023

Technical Report Documentation Page

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA-OR-RD-24-03		

4. Title and Subtitle	5. Report Date
Active Transportation Counts from Existing On-Street Signal and Detection Infrastructure	February 2024
	6. Performing Organization Code

7. Author(s)	8. Performing Organization Report No.
Sirisha Kothuri, Patrick Singleton, Mahyar Vahedi Saheli, Elizabeth Yates, Joseph Broach	

9. Performing Organization Name and Address		10. Work Unit No. (TRAIS)
Portland State University P.O. Box 751 Portland, OR 97207	and Utah State University 4110 Old Main Hill Logan, OR 84321	11. Contract or Grant No.

12. Sponsoring Agency Name and Address		13. Type of Report and Period Covered
Oregon Dept. of Transportation Research Section 555 13th Street NE, Suite 1 Salem, OR 97301	and Federal Highway Admin. 1200 New Jersey Avenue SE Washington, DC 20590	Final Report 14. Sponsoring Agency Code

15. Supplementary Notes

16. Abstract

This study's objective was to use data from existing traffic signal infrastructure to estimate pedestrian volumes. Pedestrian push-button actuations were collected from signal controller logs at 49 intersections in western Oregon and an additional 16 intersections in eastern Oregon. These actuations were then compared to observed pedestrian counts, totaling over 34,000 people, obtained from video recordings. After exploring various options, a simple quadratic relationship was modeled using a single measure of pedestrian signal activity: the number of push-button presses (filtered to remove multiple presses within 15 seconds). The model's predictions showed a correlation of 0.86 with observed pedestrian volumes and had an average error of ±2.4 pedestrians per hour. These results suggest that existing traffic signal infrastructure data can be used to estimate pedestrian volumes in Oregon with reasonable accuracy. Using such pedestrian volume estimates can lead to improvements in pedestrian traffic monitoring, safety assessments of exposure, and equity and health analyses.

17. Key Words	18. Distribution Statement
Actuated, Pedestrians, Pedestrian Counts, Traffic Signals, Traffic Signal Controllers	Copies available from NTIS, and online at http://www.oregon.gov/ODOT/TD/TP_RES/

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	99	

SI* (MODERN METRIC) CONVERSION FACTORS

APPROXIMATE CONVERSIONS TO SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
in	inches	25.4	millimeters	mm
ft	feet	0.305	meters	m
yd	yards	0.914	meters	m
mi	miles	1.61	kilometers	km
		AREA		
in^2	square inches	645.2	millimeters squared	mm^2
ft^2	square feet	0.093	meters squared	m^2
yd^2	square yards	0.836	meters squared	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	kilometers squared	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	ml
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	meters cubed	m^3
yd^3	cubic yards	0.765	meters cubed	m^3
NOTE: Volumes greater than 1000 L shall be shown in m^3.				
		MASS		
oz	ounces	28.35	grams	g
lb	pounds	0.454	kilograms	kg
T	short tons (2000 lb)	0.907	megagrams	Mg
		TEMPERATURE (exact)		
°F	Fahrenheit	(F-32)/1.8	Celsius	°C

APPROXIMATE CONVERSIONS FROM SI UNITS

Symbol	When You Know	Multiply By	To Find	Symbol
		LENGTH		
mm	millimeters	0.039	inches	in
m	meters	3.28	feet	ft
m	meters	1.09	yards	yd
km	kilometers	0.621	miles	mi
		AREA		
mm^2	millimeters squared	0.0016	square inches	in^2
m^2	meters squared	10.764	square feet	ft^2
m^2	meters squared	1.196	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	kilometers squared	0.386	square miles	mi^2
		VOLUME		
ml	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	meters cubed	35.315	cubic feet	ft^3
m^3	meters cubed	1.308	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.205	pounds	lb
Mg	megagrams	1.102	short tons (2000 lb)	T
		TEMPERATURE (exact)		
°C	Celsius	1.8C+32	Fahrenheit	°F

*SI is the symbol for the International System of Measurement

ACKNOWLEDGEMENTS

We would like to thank Josh Roll, ODOT's research coordinator, for expertly managing the study. We extend our gratitude to the members of the Technical Advisory Committee, whose valuable guidance and feedback ensured the success of this project. Special thanks to Shawn Strasser, Alison Tanaka, Matt Dorado, and John Fasana for providing the traffic signal controller log data. We also greatly appreciate the efforts of the students who collected the pedestrian data used in this project:

- Portland State University: Liam Anthony, Jason Grajales, Salvador Mercado, Matthew Prak.
- Utah State University: Caitlyn Andrus, Alexis Aragon, Allie Boyer, Melissa Brown, Maya Cottam, Mercy Smith, Jared Webb, Peyton Webb, Emma Wheeler.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

ODOT's mission is to provide a safe and reliable multimodal transportation system that connects people and helps Oregon's communities and economy thrive. In order to fulfill this mission, it is important to have a robust understanding of the usage of each mode within the system as well as the linkages between modes. Currently there is no systemic accounting of pedestrian traffic across the transportation network, making it difficult to understand systemic crash safety and prioritize projects. Analysts must rely on imperfect proxies such as land use density, race, income, and transit usage in the absence of measured pedestrian traffic data. Additionally, ODOT is developing a new key performance measure for reporting pedestrian travel to the Oregon Transportation Commission (OTC), which will require pedestrian traffic counts. While significant efforts have been made in the last decade to establish active transportation counting programs, the spread of permanent counters has been limited due to cost and resource considerations. This trend is likely to continue. A noteworthy possibility is to utilize existing traffic signal infrastructure to count pedestrians by using push-button actuation systems as permanent traffic counters. Permanent counters are critical for assessing temporal patterns and for developing adjustment and expansion factors which can be applied to short-duration counts to estimate longer-duration active transportation flows on the network. Existing permanent traffic counters tend to be situated on trails and therefore are not able to provide information on the amount of walking activity that takes place on the statewide street network and on sidewalks and in crosswalks. These counters also require significant investments in the installation and maintenance of specialized equipment, which is often proprietary.

Currently, existing on-street infrastructure, such as pedestrian pushbuttons, continuously collects data, which are stored in an ODOT data warehouse. However, these data have rarely been utilized to understand total pedestrian volumes at those sites. As of August 2023, there were around 900 signals on the ODOT network logging pedestrian actuation data. If accurate adjustment factors could be developed, using data from existing infrastructure (such as pedestrian push-button actuations) for pedestrian activity would kickstart a nonmotorized traffic data collection program in Oregon. By combining this actuation data with information collected from short-duration counts, it would be possible to estimate walking volumes across the entire system. These data would provide necessary information for a more comprehensive systemic safety analysis and assist in prioritizing funding allocation decisions. A similar successful effort was recently completed in Utah (Singleton, Runa, and Humagain, 2020), showing promising results. Therefore, it is necessary to understand the accuracy and robustness of these pedestrian push-button data in Oregon and how existing on-street infrastructure can be leveraged to compliment a nonmotorized traffic data collection program.

The intent of this research is to examine the feasibility of collecting pedestrian volumes from existing on-street infrastructure, such as pedestrian pushbuttons, at a large scale. Achieving this goal involved developing adjustment factors that convert pedestrian data collected from existing on-street infrastructure to pedestrian volumes, and then determining the transferability of the methods developed. The remainder of this report is organized into the following chapters:

Chapter 2.0 reviews literature and techniques about pedestrian counting techniques—including manual counts, automated counts, traffic signal pedestrian push-button data, and other emerging data sources—as well as methods for estimated pedestrian volumes.

Chapter 3.0 describes the data collection process, including assembling data, selecting study locations, retrieving pedestrian traffic signal data from controller logs, extracting observed pedestrian counts from video recordings, and integrating the various data together.

Chapter 4.0 describes the data analysis process, including data analysis principles; the consideration of three dependent variables, eight independent variables, eight functional forms, and 14 segmentation variables; and the use of K-fold cross-validation and five model validation statistics to assess the performance of the models.

Chapter 5.0 presents the results of a multi-step data analysis process, including descriptive statistics; the inspection and selection of dependent variables, independent variables, functional forms, and segmentation variables; and the final recommended model(s).

Chapter 6.0 concludes with a summary of the research project's key findings, as well as a discussion of the study's limitations and opportunities for future work.

2.0 LITERATURE REVIEW

Counts are critical for measuring nonmotorized travel, monitoring trends, planning for new infrastructure, and in safety analyses. They are also useful in assessing health, economic, and equity impacts of nonmotorized travel and for calibrating and validating travel demand models. Motorized travel monitoring techniques have been well established; however, nonmotorized counting is fairly new, with the methods and technologies still evolving.

When examining non-motorized travel modes, bicycle travel and pedestrian travel are unique from one another. Whereas cyclists (or bicycles) tend to follow more vehicle-like behavior patterns, pedestrians travel in all directions and do not necessarily adhere to specific paths. Additionally, they tend to walk in groups, which can create accuracy issues when using certain technologies to obtain counts (FHWA, 2016). These factors make counting pedestrians challenging, and currently there is no systemic monitoring of pedestrian traffic across the transportation network. This chapter reviews the literature on pedestrian counting methods and technologies, pedestrian push-button actuations from traffic signal data, other emerging sources of pedestrian data, and ways to model pedestrian volumes.

This literature review includes peer-reviewed journal articles, conference papers, technical reports, and guidebooks produced by state and federal transportation agencies. These documents were obtained by searching journal archives such as those maintained by the Transportation Research Board (i.e., TRID) and Google Scholar, general search engines like Google, transportation agency websites such as ODOT, and the reference lists of the identified documents.

2.1 COUNTING TECHNIQUES

Pedestrian volumes can be obtained using both manual and automated counting methods. While automated technologies for counting pedestrians are continuously evolving, fewer technologies exist to count pedestrians than bicyclists (Nordback et al., 2016). Commonly used automated technologies for counting pedestrians include video cameras (in-field and from video footage), passive and active infrared sensors, and radio beams. Less-commonly-used technologies include thermal cameras and pressure or acoustic pads. Other technologies that capture pedestrian activity include pedestrian actuations from pushbuttons, or a subset of counts from Wi-Fi and Bluetooth readers. All of these technologies for counting pedestrians are described further in the subsections below.

2.1.1 Manual Counts

2.1.1.1 In-Field

Manual counts can include both in-field counts and counts from video. In-field manual counts are recorded using paper sheets, traffic count boards, counters, or smartphone apps. These counts can capture additional features such as gender and risky behaviors and are generally appropriate for short duration counts due to high cost. Manual counts are also necessary for validating automated counters (Ryus et al., 2014a). Data collector

training, motivation, and management are critical for obtaining accurate counts (Ryus et al., 2014a). One study using manual intersection counts found undercount rates ranging between 8% and 25% for 15-minute intervals compared to manual counts obtained from video (Diogenes et al., 2007).

2.1.1.2 From Video

Manual counts can also be collected from video footage collected in the field. These counts can also yield additional characteristics such as gender and risky behaviors, however these characteristics may be more difficult to note from the video than directly in the field. The biggest advantage of video recording when compared to in-field counts is the ability to slow down or speed up the video to get accurate observations. Pedestrians in groups or crowds can be counted more accurately than in real-time, while time periods with no pedestrians can be watched quickly, thus yielding accurate data often with less effort than in-field counts. This technique also allows for a single data collector to reduce data from multiple sites. However, it also requires the installation of permanent or temporary cameras (or the use of existing cameras), which can be expensive, and the storage and transfer of large video files. Like in-field counts, the method of obtaining counts from video is also best suited for short-duration counts due to the high cost involved with the data reduction process.

2.1.2 Automated Counting Technology

2.1.2.1 Infrared

Infrared devices to count nonmotorized users can be categorized into passive and active. Passive infrared sensors detect users by comparing the temperature of the background to the infrared radiation emitted by persons passing in front of the sensor (Ryus et al., 2014a). Passive devices use a signal transmitter on only one side of the detection area (FHWA, 2016). Active infrared sensors count users using an infrared beam between an emitter and a receiver located on opposite sides of a path, and a count is registered when the beam is broken due to a user crossing (Ryus et al., 2014a). Both passive and active infrared counters cannot distinguish between pedestrians and cyclists. If differentiation is necessary, they can be installed in conjunction with bicycle-only counting technology such as inductive loops. The pedestrian count can then be obtained by subtracting the bicycle count from the combined count provided by the infrared sensors.

Installation of infrared sensors is critical for accuracy. Passive sensors work best when installed pointing towards a fixed object (Ryus et al., 2014a). The receiver and transmitter for active infrared sensors need to be installed facing each other with a clear line of sight (Ryus et al., 2014a). One study found that the passive and active infrared sensors tested worked best when they were placed at a 45-degree angle to the roadway (Jones et al., 2010). These sensors are portable and easy to install and can be used in combination with other technologies. However, neither of these sensors is suitable for on-street monitoring (i.e., counting pedestrians in crosswalks).

Results regarding accuracy are mixed, with some studies finding that passive sensors undercount pedestrians (Schneider et al., 2012, Ryus et al., 2014a), while others have found evidence of overcounting (Kothuri et al., 2017). Undercounting was also observed among active infrared sensors (Ryus et al., 2014a, Jones et al., 2010). Testing conducted on two passive infrared sensors and one active infrared sensor as part of NCHRP Research Report 797 revealed annual average percent differences of 11% and 33% for the passive devices and 12% for the active device (Ryus et al., 2014a). Occlusion can also be a problem, particularly on facilities with high volumes of pedestrians (Greene-Roesel et al., 2008).

2.1.2.2 Video Image Processing

Both pedestrian and cyclist counts can be obtained from video cameras using automated algorithms to capture both intersection and screenline counts. The technology and the algorithms have been continuously evolving over the past few years. The cameras are portable and easy to install, and the collected video can be used for other purposes. This technology can provide crosswalk counts pedestrian counts, however it is expensive and therefore used only for short duration counts.

The accuracy of video image processing can depend on camera height and angle, as well as being affected by environmental conditions such as glare, fog, rain. One study found that nighttime detection misses are up to 50% and false positives are 3.4% of the ground truth counts, which are much higher than for daytime detection (Yoneyama et al., 2005). Some studies have developed custom algorithms to count and separate pedestrians from other users using cluster-based appearance modeling and online tracking (Li et al., 2012, Kocamaz et al., 2016), and machine learning algorithms with Kalman filtering approaches (García et al., 2012). The detection rate varied between 87% and 98% for these studies. Liu et al. (2017) used cameras mounted on transit service doors and counted people boarding and alighting using a convolutional neural network (CNN) with other classifiers and obtained a 93% accuracy rate. Another study used RGB-D cameras mounted on the ceilings of public buses to count boardings and alightings and obtained accuracy rates of 91% to 94% for people exiting and 75% to 85% for people boarding (Sun et al., 2019). Other research has used one CNN or two cascade CNN's to estimate crowd count and density from an image derived from RGB cameras (Zhang et al., 2016, Sindagi and Patel, 2017). The images derived from RGB cameras suffer from the known issue of poor performance during night and during adverse weather conditions (Lesani et al., 2020). To overcome these drawbacks, studies have used thermal images and automated algorithms to count pedestrians (Dai et al., 2007; Gomez et al., 2018). Detection rates using thermal images vary between 99% (Dai et al., 2007) and 54% (Gomez et al., 2018).

2.1.2.3 Thermal

Thermal cameras have been proposed as an alternative to video cameras because of their ability to eliminate the issues associated with adverse weather conditions and poor lighting. Thermal devices generate infrared images by detecting body heat (Ryus et al., 2014a). Studies have tested thermal cameras for vehicle detection and achieved high

detection rates for vehicles (Iwasaki et al., 2008; MacCarley et al., 2000). Thermal video has also been used to detect pedestrians at night, which poses particular challenges due to low light conditions thereby increasing the crash risk. Xu et al. (2005) used a support vector machine (SVM) to detect and classify pedestrians using a thermal camera attached to a moving vehicle and found occlusion to be of significant concern in heavy traffic conditions. Another study recommended combining both traditional video and thermal camera technology to take advantages of both (Krotosky and Trivedi, 2007). One study found a 15% improvement in detection rate over regular cameras (Balsys et al., 2009). Recently, Fu et al. (2017) tested traditional and thermal video cameras and found that while the traditional video cameras narrowly outperformed thermal video cameras during daytime, the thermal video cameras demonstrated higher accuracy during low visibility and in the presence of shadows. Kothuri et al. (2017) tested the ability of thermal cameras to count pedestrians, bicycles, and motor vehicle traffic at an intersection. They found significant undercounting of bicyclists and pedestrians on a sidewalk, whereas the on-street vehicle volumes were overcounted. They also noted a potential link between temperature and accuracy (Kothuri et al., 2017).

2.1.2.4 LiDAR

Recently, LiDAR (Light Detection and Ranging) has been emerging as a technology for detection and counting. It is currently used in a wide range of applications including automated vehicles, the generation of road geometry data, and traffic monitoring (De Silva et al., 2017; Jung et al., 2018; Zhang et al., 2019). LiDAR sensors emit laser beams at an object and then estimate the distance to the object using time in flight and the energy of the reflected beams (Lesani et al., 2020).

LiDAR systems can be classified as one, two- (2D), or three-dimensional (3D) based on channels and coverage area. 2D and 3D systems have been evaluated in a wide variety of applications. 2D systems include an array of beams, where each beam covers a straight line and a set of beams covers a two-dimensional plane in space (Lesani et al., 2020). 3D systems include two different types of solid and rotating LiDAR sensors, which have the ability to scan the surrounding space in three dimensions (Pacala and Edala, 2017; Tarko et al., 2018). One study evaluated the use of 3D LiDAR mounted on a wheelchair to detect pedestrians walking around the wheelchair in sidewalks and found a 96% detection rate, although the sample was very small (Savtchenko, 2011). Premebida et al. (2009) tested two 3D LiDAR sensors mounted on a vehicle to detect and classify pedestrians. They reported 97% accuracy when classifying the detected objects as pedestrians; however, the overall detection rate was not reported (Premebida et al., 2009). Akamatsu et al. (2014) tested a 3D based LiDAR sensor for tracking pedestrians entering a building and reported an error of 2%; however, it was not clear if this error was absolute or not. Lesani et al. (2020) used a 2D LiDAR sensor and clustering algorithms to detect, count, and identify the direction of travel for each pedestrian. Their results showed a 97% accuracy rate when counting pedestrians, with a false positive rate of 1.1%, overcounting error of 0.7% and undercounting errors of 1.3% and 2.7% for the two test sites (Lesani et al., 2020). The cost of these LiDAR systems can be high, which could be a limiting factor for widespread adoption.

2.1.2.5 Radio Beams

Radio beams use a radio beam transmitted from a transmitter to a receiver placed on opposite sides of the facility for detection. Users are detected when they pass in front of the device, causing the beam to break. Devices that use multiple radio frequencies can differentiate between pedestrians and cyclists (Ryus et al., 2014a). These can be used for screenline counts on sidewalks and on paths for both short-duration and permanent counts. They are subject to occlusion errors (Ryus et al., 2014a). NCHRP Research Report 797 tested two radio beam products, only one of which was able to count pedestrians and cyclists separately. The annual average daily pedestrians (AADP) for two products were 53% for pedestrians and 28% for the combined count (Ryus et al., 2014b).

2.1.2.6 Pressure and Acoustic Sensors

Pressure sensors detect changes in weight, while acoustic sensors detect the passage of energy waves through the ground (FHWA, 2016). For both these sensors, the sensing device is placed underneath or near the detection area. These are mostly used on unpaved trails or paths, where the sensor can be placed underneath the path without much damage (FHWA, 2016). Pressure and acoustic sensors do require that pedestrians or cyclists pass directly over them, so they are most suited to situations where pedestrians or cyclists will travel single file (Ryus et al., 2014a). However, they have also been used in Europe on sidewalks and at curbside pedestrian signal waiting areas as an alternative to pushbuttons (FHWA, 2016). Some models of these devices are capable of differentiating between pedestrians and cyclists and also can be used to gather directional information (FHWA, 2016). The accuracy of these devices has not been tested.

2.1.3 Traffic Signal Pedestrian Push-Button Data

2.1.3.1 Controller Event Logs

Traffic signals (and signal controller technologies) are ubiquitous components of urban transportation systems that monitor and manage traffic operations at intersections. As such, traffic signals offer potential opportunities to utilize existing on-street infrastructure for travel monitoring and traffic counts. Regarding pedestrian travel, many traffic signals include pedestrian pushbuttons that people walking who want to cross an approach can press to request the walk indication. Pushbuttons are active sensors that (barring equipment malfunctions) confirm that a person was present at that location at that time. Of course, pushbuttons are imperfect measures of pedestrian counts: one person may press the button multiple times, a group of people may only press the button once, or people may cross without pressing the button. In addition, some traffic signals operate in pedestrian recall and rest-in-walk (for some crossings and times-of-day), which means that the walk signal appears without having to press the button. However, if detection events could be recorded, stored, and shared, pedestrian data from traffic signal controller logs could be a useful automated source of information (that already exists at minimal additional cost) on pedestrian activity.

Until recently, this rich set of traffic signal controller event data was not being systematically logged. To obtain logs, researchers had to create custom devices and software to physically plug into controllers and save events signal-by-signal, for example to calculate pedestrian delay (Kothuri, 2014). About 15 years ago, Smaglik et al. (2007) developed a general method and module for automatically logging time-stamped event data from traffic signal controllers. These high-resolution data loggers record many types of traffic signal events, including active phase changes, phase control and overlap events, and vehicle and pedestrian detection events. Each record includes a timestamp (measured to the nearest tenth-of-a-second), an event code, and an event parameter representing a phase or overlap number, detector channel, or other information (Sturdevant et al., 2012; Li et al., 2020).

Several pedestrian-relevant traffic signal events are commonly logged:

- Event code 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 code 21, Pedestrian Begin Walk: This event occurs with the activation of the walk indication for a particular phase.

- Event code 22, Pedestrian Begin Clearance: This event occurs with the activation of the flashing don't walk indication for a particular phase.

- Event code 23, Pedestrian Begin Solid Don't Walk: This event occurs when the don't walk indication becomes solid, with the termination of the flashing don't walk indication.

- Event code 45, Pedestrian Call Registered: This event occurs when a call to service for a particular phase is registered from pedestrian demand. Note that this event may not occur if pedestrian recall is set for the phase.

- Event codes 89 and 90, PedDetector Off and PedDetector On: These events occur when the signal from the pedestrian push-button is deactivated or activated, after any delay or extension is processed, for a particular pedestrian detector channel. Multiple pedestrian detection events may occur for a single pedestrian call registered.

Table 2.1 is an example high-resolution traffic signal event log containing pedestrian-relevant data. The location is signal ID 4113, the intersection of SE 82nd Ave and SE Foster in Portland, OR. All events are for phase 4, associated with the northern crosswalk across SE 82nd Ave. About 25 minutes after midnight on 21 June 2022, a pedestrian walking arrived at the intersection and pressed the push-button twice in quick succession: from 14.8 to 15.1 and from 15.5 to 15.7 seconds (event 90 on, event 89 off). Upon the first push-button press, the signal controller also registered a pedestrian call (event 45), because it was the first pedestrian detection event for this phase during this cycle; no pedestrian call was registered during the second push-button press. About one second later, at 15.9 seconds, phase 4 was served (event 0) and the WALK indication turned on

(event 21). Seven seconds later, the FLASHING DON'T WALK indication (event 22) started. Sixteen seconds afterwards, at 38.9 seconds, the steady DON'T WALK indication (event 23) appeared, signaling the end of the crossing event.

Table 2.1: Example high-resolution traffic signal controller event log

Location	Timestamp	Event Code	Event Parameter
4113	06/21/2022 00:25:14.800	90	4
4113	06/21/2022 00:25:14.900	45	4
4113	06/21/2022 00:25:15.100	89	4
4113	06/21/2022 00:25:15.500	90	4
4113	06/21/2022 00:25:15.700	89	4
4113	06/21/2022 00:25:15.900	0	4
4113	06/21/2022 00:25:15.900	21	4
4113	06/21/2022 00:25:22.900	22	4
4113	06/21/2022 00:25:38.900	23	4

2.1.3.2 Data Access

To harness the power of high-resolution traffic signal controller log data for signal systems operation and management, the Automated Traffic Signal Performance Measures (ATSPM) system has been developed to convert raw data into useful performance measures (Atkins, 2016; Day et al., 2014; Day et al., 2016). ATSPMs provide real-time and historical performance measures about traffic signal operations, for example, corridor progression, splits, and detections. There is currently one pedestrian metric: pedestrian delay, which measures the time difference between a pedestrian call registered (event 45) and the next subsequent begin walk indication (event 21) for the associated phase. A recent NCHRP guidebook (Nevers et al., 2020) describes the process of selecting, implementing, and utilizing ATSPMs.

As depicted in Figure 2.1, ATSPMs rely upon modern traffic signal controllers that record high-resolution events, robust communications systems to transfer the data, data storage systems to archive and store the data, and programs to query and retrieve the data for display to the end user. ATSPMs have been developed and implemented by state and local transportation agencies across the country. For example, as of late 2021, utilizing intergovernmental agreements and a robust state-wide fiber-optic network, the Utah Department of Transportation (UDOT) has integrated 2,187 state- and locally owned traffic signals into their ATSPM system. Such integration also allows for the raw data on detections (such as pedestrian push-button events) to be accessed, queried (such as through SQL), and processed (using any number of data analysis programs, such as R or Python) for conversion into counts or volume estimates.

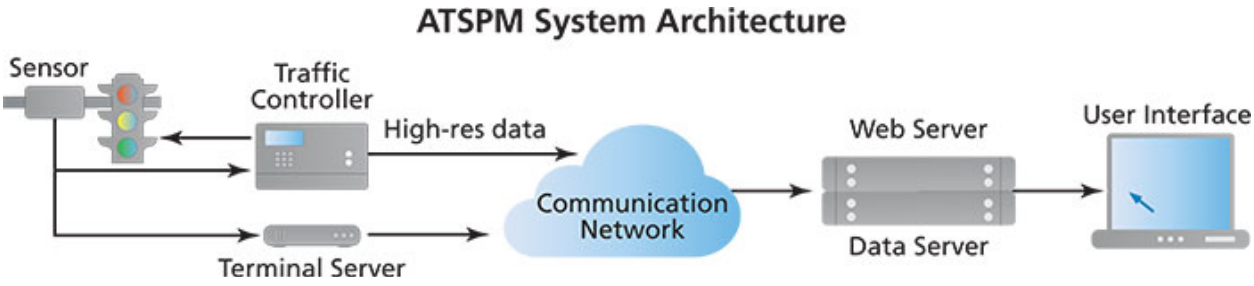


Figure 2.1: Automated Traffic Signal Performance Measures (ATSPM) system architecture

2.1.3.3 Push-Button Data and Pedestrian Counts

High-resolution traffic signal data—especially records of pedestrian detections and pedestrian phase calls—could provide valuable information about pedestrian activity, measured continuously over time and across many hundreds or thousands of locations. (Of course, this opportunity is limited by the number of traffic signals with high-resolution event logging capabilities that have phases with walk indications and crossings with pushbuttons.) Nevertheless, as previously mentioned, push-button data are imperfect measures of pedestrian activity, given that some people may press the button multiple times and others may not press the button before crossing. As a result, the relationship between pedestrian traffic signal data and pedestrian counts or volumes likely depends on pedestrian push-button utilization, crossing behavior, and traffic signal compliance. Previous research on pedestrian behavior and the utilization of pedestrian pushbuttons at signals has found that rates vary across locations such as by signal type (Kutela and Tang, 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 Tang, 2020). Other, less directly relevant research has investigated push-button utilization in order to quantify the effectiveness of providing visual/audible feedback about a button press (Van Houten et al., 2006) or measured button-press duration to inform the timing for an extended press that would extend the walk interval at accessible pedestrian signals (Noyce and Bentzen, 2005).

A few studies in recent years have investigated the use of pedestrian data from traffic signal controller event logs to estimate walking activity at signalized intersections. Early work focused on single or a couple of intersections and short durations. Recent research in Utah expanded on this early work and investigated many dozen intersections and various time periods. Still, more research is needed in order to refine and generalize these methods to apply in various geographies and different settings. The following paragraphs summarize these studies.

Working at one signalized intersection in Indiana, Day et al. (2011) analyzed traffic signal data on pedestrian phase actuations per hour for over 18 months. They were able to identify patterns of pedestrian signal activity and variations based on time-of-day, day-of-week, weather and other seasonal effects, special events, and a change in the configuration of the pedestrian phases (implementation of an exclusive pedestrian phase). One key contribution is that the authors demonstrated that it was feasible to record pedestrian actuations over a long period of time with minimal additional cost, although

they did not compare actuations with observed pedestrian counts or develop pedestrian volume estimates.

In Oregon, Blanc et al. (2015) conducted a 24-hour pilot study of pedestrian activity at one signal with actuated pedestrian crossings (using push-button detection) on all four crossings. The authors used video data to manually count 596 pedestrians, and then compared those counts to the 482 pedestrian phases recorded in the controller event logs. They developed adjustment factors for each phase (and for the intersection overall), which ranged from 0.84 to 1.50 pedestrians per phase. When comparing hourly pedestrian counts to pedestrian actuations for each crosswalk, the authors found correlations of 0.83 or greater, demonstrating how traffic signal pedestrian data could potentially approximate (with adequate accuracy) pedestrian crossing volumes at a signalized intersection. Finally, they demonstrated how one might apply the adjustment factors to pedestrian phase data and calculate estimates of daily and annual average daily pedestrian volumes.

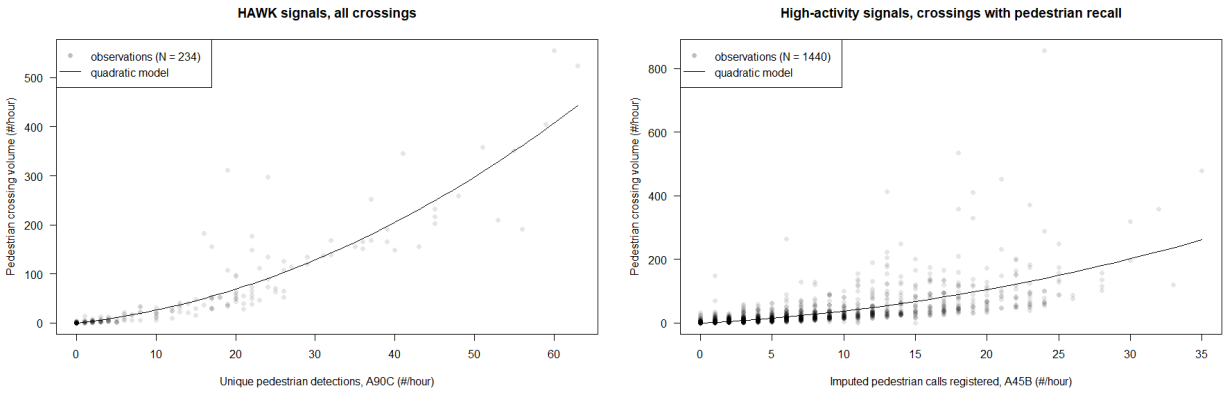
Two years later, Kothuri et al. (2017) returned to the same Oregon intersection to try replicating the previous study's findings. During daytime hours across nearly three days, the authors used video data to manually count 818 pedestrians; signal controller log data counted 723 pedestrian phases over the same time period. Compared to the previous study, adjustment factors were of roughly the same magnitude (0.95 to 1.21 pedestrians per phase), and correlations of hourly observations were nearly as good in most cases (around 0.80, although one crossing was about 0.67).

Recently, a large-scale research effort in Utah investigated the feasibility of using pedestrian traffic signal data for pedestrian volume estimation (Singleton et al., 2020; Singleton and Runa, 2021). The authors randomly selected 90 signalized intersections across Utah to study, ensuring locations in different regions and with different levels of pedestrian activity (high, medium, low). They then used existing overhead video cameras operated by the Utah DOT to record at least 24 (and often more) hours of video at each signal, and later manually count pedestrians. Over 175,000 pedestrians were counted from over 10,000 hours of video, covering different months, weekdays, and hours throughout 2019. Using traffic signal event data from the Utah DOT's ATSPM system, they then calculated the number of unique pedestrian push-button presses (those with a time difference of 15+ seconds from the previous press) and pedestrian calls registered for each phase in each hour with video data. Finally, the authors estimated simple non-linear (quadratic and piecewise linear) regression models predicting hourly pedestrian crossing volumes as a function of the constructed measures of pedestrian signal data, in different situations. For ease of application, the models did not include traffic volumes or neighborhood characteristics (demographics, socioeconomics, land use, built environment, etc.), 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; pedestrian hybrid beacon (HAWK) vs. traditional signal). See Figure 2.2 for the different situations. Using data for more than 22,500 crossing-hours of observations, the correlation between hourly observed pedestrian counts and model-estimated pedestrian crossing volumes was 0.84 overall. Most models had correlations close to 0.90. Results

indicated that crossings with pedestrian recall had poorer performance, because pedestrians did not need to press the push-button to receive the walk indication. Overall, the mean error was ±3 pedestrians per hour (Singleton et al., 2020; Singleton and Runa, 2021). The authors later applied their methods to estimate AADP crossing volumes for almost all signalized intersections in the state of Utah.

Other recent work in Arizona has tested the ability of traffic signal and push-button data to predict pedestrian crossing volumes at traffic signals. Li and Wu (2021) first developed an algorithm to estimate pedestrian volumes at two mid-block crossings (both with pedestrian hybrid beacons) using push-button event-based data, and then compared their model estimates to hourly pedestrian volumes observed by manually counting seven days of video recordings. Their methods yielded pedestrian volume estimates with a mean error of ±2.3 and ±1.8 pedestrians per hour at the two study locations. In a follow-up study, the same authors (Li, Xu, and Wu, 2021) looked at 70 signalized intersections in and around Tucson and collected ground-truth pedestrian data manually or from video-based sensors. This time, they used Bayesian additive regression tree (BART) method and considered both traffic signal event-based data (cycles, push-button presses, time between presses, etc.) and information about transit trips and local points of interest (destinations). Their methods achieved a correlation of 0.92 between hourly observed and estimated pedestrian volumes, which was better than a random forest model or stepwise linear regression.

Altogether, these results suggest that pedestrian traffic signal data can be used to estimate pedestrian crossing volumes with reasonable accuracy, at least considering the cost savings of utilizing existing on-street signal and detection infrastructure to add hundreds or even thousands of new permanent counters to a pedestrian traffic monitoring system.



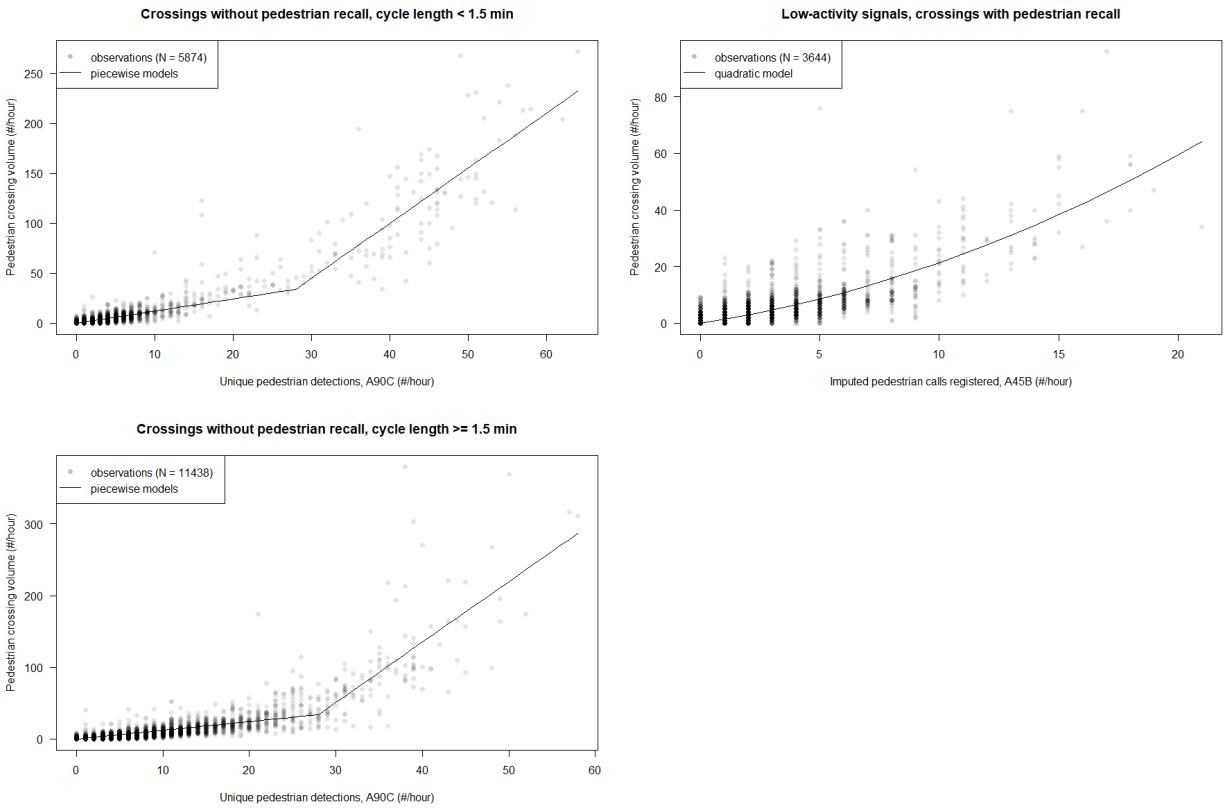


Figure 2.2: Pedestrian volume estimation models for different situations (Singleton and Runa, 2021)

2.1.4 Emerging Sources

While manual and automated counts (including those from pushbuttons at traffic signals) provide good information, they only do so for specific locations on the network. These methods are unable to identify the same individual at various points in the network to define trip routes, origins, destinations, and travel times (Lesani and Miranda-Moreno, 2018). To overcome these limitations, studies have focused on data from Bluetooth (Bullock et al., 2010; Malinoskiy et al., 2012) and Wi-Fi access points (Danalet et al., 2014; Musa and Eriksson, 2012).

The advantages of Bluetooth sensors compared to other technologies include lower costs, ability to collect large quantities of real-time data, and ease of installation (Lesani and Miranda-Moreno, 2018). They are suitable for both short-duration and permanent counts. The limitations of Bluetooth sensors include the small sample of smartphone users who use Bluetooth leading to low detection rates. Additionally, keeping Bluetooth enabled on the smartphone can also drain the battery. Therefore, most of the Bluetooth applications are on vehicular networks, where the Bluetooth devices are found in vehicles. There are very limited studies evaluating their use for counting pedestrians and cyclists. Detection rates for Bluetooth devices are between 5% and 12% (Malinovskiy et al., 2012). Specific to pedestrians, one study used Bluetooth sensors to measure pedestrian travel times at airport security checkpoints (Utsch and Liebig, 2012), while another

study used Bluetooth sensors at two locations to determine pedestrian travel times and dwell times at sensor locations (Malinovskiy et al., 2012).

Some studies have also examined Wi-Fi devices connected to Wi-Fi access points as an alternative to Bluetooth devices. However, Wi-Fi access tracking requires that devices be connected to a specific wireless network which encompasses the entire detection area (Lesani and Miranda Moreno, 2018). Therefore, the study area cannot be increased without changing the network coverage, which is a significant limitation. Due to these shortcomings, there is interest in developing independent Wi-Fi sniffer systems for pedestrian activity, which will allow for the capture of MAC addresses that are not limited to a specific network; a passive WiFi scanner could capture all available Wi-Fi probe signals broadcasted by Wi-Fi devices that are attempting to identify available access points (Lesani and Miranda Moreno, 2018).

One study evaluated pedestrian flow techniques using Wi-Fi and Bluetooth scanners and found that Wi-Fi overestimated and Bluetooth underestimated flows compared to the number of boarding pass scans (Schauer et al., 2014). Another study used Bluetooth and Wi-Fi sensors, with the Bluetooth data being used to assign a unique ID to each user and the Wi-Fi to obtain the user's location (Vu et al., 2010). Abedi et al. (2015) evaluated the impacts of small and large antenna gains on tracking the movements of pedestrians and cyclists based on MAC address datasets. Other applications include the estimation of wait time in bus terminals (Kurkcu and Ozbay, 2017), counting pedestrians in public transit stations using Wi-Fi traces (Bai et al., 2017), and a method to improve detection rate and estimate precision of pedestrian flow (Du et al., 2017). A recent study developed an integrated Wi-Fi–Bluetooth system to monitor pedestrian and cyclist activity, differentiated between the two sets of users, and proposed an extrapolation method that combined count and MAC data (Lesani and Miranda Moreno, 2018). The results revealed high detection rates for the combined system, high correlation between estimated and ground truth speeds and extrapolated Wi-Fi counts and ground truth counts, and low classification errors (Lesani and Miranda Moreno, 2018).

Another emerging source is the crowdsourced third-party data. These sources use anonymized location data from apps, smartphones, and navigation devices in connected vehicles to develop count estimates. With some exceptions (like fitness-tracking apps), these data sources—which include multi-app location-based services (LBS) in addition to Wi-Fi, Bluetooth, and cell-tower pings—tend to be mode-unspecified (Lee and Sener, 2020). As a result, they require algorithms to automatically identify travel characteristics such as mode, which may be imprecise and proprietary. For example, StreetLight Data offers bicycle and pedestrian data from mode-unspecified LBS data; their described process involves a machine learning (random forest) model trained on a variety of mode-tagged GPS data sources (including household travel surveys) and geospatial context data, which provides a probability of a ping being of a particular mode, and pings are aggregated into trips based on changes in predicted mode and/or to/from stationary (StreetLight, 2020). Similarly, another provider (Replica) uses LBS data alongside other information (points-of-interest, credit transactions, etc.) to simulate trip-making, including a machine learning mode choice model that includes walking (Replica, 2020). LBS data provides the ability to access counts across a wide network, however there are few published studies yet that have validated these counts against actual volumes. For example, StreetLight's pedestrian validation process involved aggregate comparisons with household travel surveys (e.g., trip lengths, durations, and speeds) and disaggregate comparisons with permanent counters at only

around 12 locations in Philadelphia and Ottawa (StreetLight, 2020). In general, the use of LBS data for bicycle traffic monitoring is more advanced and better validated than uses for pedestrian traffic monitoring.

2.2 MODELING METHODS

In the absence of sensors that collect pedestrian counts and monitor levels of walking activity at all times in all locations, some models are necessary to estimate or approximate pedestrian volumes in locations and for times when data are not being collected. The most basic models are standard tools in the traffic monitoring world: expansion and adjustment factors that can be used to convert short duration counts to longer-term average volumes. Other models include those already described in Sections 2.1.3.3 and 2.1.4 about estimating pedestrian volumes from push-button data or extrapolating volumes from location-based-services data that are samples of the population. Additionally, there are two other common ways to obtain pedestrian volume estimates (Kuzmyak et al., 2014): from direct-demand models that utilize relationships between walking and the built environment, and from regional travel demand forecasting models used by metropolitan planning organizations. The following sections summarize these different pedestrian modeling methods. All methods would benefit greatly from additional automated pedestrian counts that could be used for estimation and validation purposes.

2.2.1 Expansion and Adjustment Factors

Due to the expense of continuous traffic count technologies, most traffic monitoring programs include a combination of automatic counters permanently installed in a few locations, plus short-term counts conducted in many other locations over a much shorter time period, usually anywhere from a few hours to a week (FHWA, 2016). Data from the continuous automatic counters are used, first to identify factor groups of locations with similar hourly/daily/weekly patterns, and second to calculate expansion and/or adjustment factors. These factors provide a simple model for how much a shorter-duration count should be adjusted or expanded to estimate a longer-duration average volume, such as annual average daily traffic. For example, a one-hour count from 4:00-5:00PM may need to be multiplied (expanded) by 10 to get an estimate of the daily volume; this hour captures one-tenth of the daily volume, on average. Or perhaps a single-day count in May should be multiplied (adjusted) by 0.95 to get an estimate of the annual average daily volume; this implies counts in May are 5.3% (1/0.95) higher than the annual average. In general, a shorter-duration count is multiplied by expansion and/or adjustment factor(s)—for the specific factor group applicable to that location—to estimate a long-term average volume at that location. Various technologies are available for doing these short- and long-duration counts, as previously described in Sections 2.1.1 and 2.1.2.

Pedestrian traffic monitoring approaches follow a similar process, usually involving factor groups and expansion/adjustment factors (Ryus et al., 2014a). While the calculation of factors is straightforward, the determination of factor groups is more subjective (Griswold et al., 2018). There are two common approaches. One adopts a land-use classification approach, in which locations are classified based on their surrounding land use; the assumption is that locations with similar land development patterns will generate similar pedestrian activity patterns. Patterns have been identified for areas with high employment density, near universities, along recreational trails, etc. (Medury et al., 2019; Schneider et al., 2009a). The second method is to use empirical

clustering, a data-driven approach that groups locations based on the similarity of their pedestrian activity patterns. For instance, non-motorized traffic has been classified into patterns spanning the range from utilitarian (especially commuting, with AM/PM weekday peaks) to recreational (more traffic on weekends) (Miranda-Moreno and Lahti, 2013). No matter the approach, one challenge is selecting an appropriate factor group—with a similar activity pattern and surrounding context—for a given short-term count location. The reliability and generalizability of factor groups and their expansion/adjustment factors depends on the number and variety of locations with long-duration pedestrian count data. For example, a recent study used pedestrian push-button activity to develop empirical clusters and relate those clusters to the built environment, sociodemographic, and climate variables (Humagain and Singleton, 2021). However, the identified pedestrian activity patterns were less varied than in previous studies, because the data relied upon existing traffic signals and likely underrepresented recreational patterns and locations.

2.2.2 Direct-Demand Models

Direct-demand models predict pedestrian volumes at specific locations from measures of the surrounding streetscape, street network, land uses, and built environment (Munira and Sener, 2017; Schneider et al., 2021; Singleton et al., 2021). Research in this area has evolved from two separate approaches with different motivations, although both paths produce a similar product (a direct-demand pedestrian model). One approach is concerned with understanding relationships between urban design and walking activity (to inform planning), so it often focuses on complex measures of urban form and the built environment. The other approach is motivated by a need to quantify pedestrian exposure for use in safety analyses (Schneider et al., 2009b). No matter the approach, direct-demand models require real-world data on pedestrian volumes; more locations and longer-time periods of pedestrian data collection can produce more accurate and generalizable forecasts.

Pedestrian volume direct-demand models utilize various measures of the built environment. This research finds that, in general, walking is greater in areas with greater residential and employment density, in locations closer to and with more transit stops, and often in places with a greater diversity of land uses (Ewing and Cervero, 2010; Saelens and Handy, 2008). Other variables tested in direct-demand models include measures of street network designs (intersection density, the percentage of four-way intersections, etc.), demographics and socioeconomics (household size and income), and streetscape and urban design qualities (like imageability or transparency) (Singleton et al., 2021).

A key limitation of many direct-demand models is a lack of ground-truth counts of pedestrian volumes in many locations and for long durations. Most models utilize manually collected short duration counts of people walking along streets segments or crossing at intersections. Counts are often done for just one or two hours and rarely exceed 12 hours; one exception is the week-long automated counts of pedestrians in one study in Virginia (Hankey et al., 2017; Lu et al., 2018). Similarly, most research uses data from a few dozen to a few hundred locations; some exceptions are around 1,000 intersections used in work from Montreal (Miranda-Moreno and Fernandez, 2011) or California (Griswold et al., 2019), and the 10,000 street segments counted in South Korea (e.g., Kim et al., 2019). Studying fewer locations limits the generalizability of findings. The temporal restriction on pedestrian count data limits most direct-demand models' abilities to

consider temporal variations in relationships between the built environment and pedestrian volumes.

Recent work in Utah has attempted to mitigate some of these limitations by developing a direct-demand model that is based on pedestrian volume estimates from traffic signal push-button data. Singleton, Park, and Lee (2021) collected a full year of pushbutton data from nearly 1,500 traffic signal controllers throughout Utah, and then applied the models created by Singleton and Runa (2021) to estimate average pedestrian crossing volumes. They then developed a direct-demand model based on various measures of the built environment (activity density, street network design, transit access), land use (residential, commercial, parks, schools, etc.), and socio-demographics (household size, income, vehicle ownership). Notably, because they used one year of continuously collected data, the authors were able to develop pedestrian volume models segmented by weekday vs. weekend and by hour-of-the-day. In this way, the models identified temporally varying influences of the built environment: for example, schools attract pedestrian activity, but only on weekdays and during daytime hours. While the direct-demand model did not use observed counts (instead, it estimated volumes based on observed push-button activity), the authors speculated that the loss of accuracy due to this characteristic was more than made up for by the benefits of using continuously collected data over an entire year (Singleton et al., 2021). Also, they then applied their models to predict pedestrian volumes at over 60,000 other intersections in Utah, offering rough estimates of pedestrian activity and exposure for future safety analyses.

2.2.3 Regional Travel Demand Forecasting Models

As part of their transportation planning mandates, metropolitan planning organizations (MPOs) operate regional travel demand forecasting models (TDMs) that predict travel by various modes as an output of different land use and transportation project scenarios. Originally developed to plan for growing automobile travel, such tools did not begin to consider walking or non-motorized modes until the late 1980s and early 1990s, due to limitations with collecting walk trips in household travel surveys and the challenges of representing walk trips in a network of major roads and large zones. Throughout the 1990s and continuing into the 2000s, increased attention to relationships between (and policies involving) travel demand and the built environment led to a gradual increase in the representation of pedestrian travel (and walking-relevant measures of the built environment) within regional TDMs. In 2005, about half of large MPOs (54%) modeled non-motorized traffic in some way (TRB, 2007; VHB, 2007). By 2017, three-quarters (75%) of large MPOs modeled non-motorized travel, and more than two-thirds (69%) of those distinguished between walking and bicycling (Singleton et al., 2018).

A variety of frameworks have been developed to model pedestrian travel within regional TDMs. Frameworks can be distinguished by the stage in the traditional four-step trip/tour modeling process at which walk trips are predicted (Singleton et al., 2018). Walk trips can be generated completely separately from other modes, in a process somewhat similar to the direct-demand models described in the previous section (except that this process models walk trips ends rather than pedestrian volumes). Alternatively, walk trips can be separated (using a walk mode split model) immediately after trip/tour generation (Clifton et al., 2016a), which allows for the inclusion of pedestrian-relevant measures of the built environment (Lefebvre-Ropars et al., 2017) or the add-on modeling of pedestrian destination choice (Clifton et al., 2016b). A similar

splitting process may happen after trip/tour destination choice. By far the most common method to model pedestrian travel in TDMs is within a trip/tour mode choice model, whether as a single non-motorized mode or with walking as a separate mode (from bicycling and other motorized modes). Pedestrian representation within activity-based models typically utilizes the mode choice framework. As of 2017, no large MPO was assigning pedestrian traffic to networks using a route choice model (Singleton et al., 2018).

There are several barriers to pedestrian modeling at a regional level, although some MPOs have taken specific approaches to overcome these challenges (Singleton et al., 2018). The large zonal and aggregate network structures used in most TDMs are not necessarily compatible with pedestrian travel, as many walk trips may be intrazonal and utilize local streets and paths that are not modeled. To address this issue, some MPOs are using micro-zones that approach the size of a block as well as all-streets-and-paths networks for pedestrian modeling or to calculate pedestrian travel times and impedances. Relatedly, TDMs can be made more sensitive to walkability-related policy by including more fine-grained measures of the built environment into models of key decisions such as trip generation, destination choice, and/or mode choice. But by far the biggest challenge reported by MPO modelers in 2017 was a lack of pedestrian behavior data to estimate pedestrian models (Singleton et al., 2018). This challenge is directly linked to the small sample sizes and general difficulty of measuring walking through region-wide household travel surveys: short walk trips may be forgotten, and many respondents may not make any walk trips. As MPOs adopt more GPS-based household travel surveys, missing walk trips may become less of an issue, but oversampling households in walkable areas may still be necessary to achieve sufficient sample sizes for pedestrian model estimation. The use of external data sources on pedestrian volumes (such as from location-based services data) could also help to improve the validation of these processes within TDMs.

2.3 SUMMARY

Counts are critical for measuring nonmotorized travel, monitoring trends, planning for new infrastructure, and in safety analyses. They are also useful in assessing health, economic, and equity impacts of nonmotorized travel and for calibrating and validating travel demand models. While vehicular counts are easily available, currently there is no systemic accounting of pedestrian traffic across the transportation network. Various technologies such as traditional and thermal cameras, radar, LiDAR, radio beams, pressure, and acoustic sensors are available to count pedestrians; however, due to cost and other considerations, they are not deployed everywhere. Direct-demand and travel demand models have also been estimated to provide pedestrian volume estimates; however, these tend to be region specific and do not transfer well to other regions.

One fairly ubiquitous data source is pedestrian push-button actuations from traffic signals, which provide a proxy for pedestrian demand. Push-button actuation has been traditionally used to serve the pedestrian phase. Recent work in Utah has examined the feasibility of using pedestrian traffic signal data for pedestrian volume estimation and found good correlations between actuations and demand. Simple non-linear (quadratic and piecewise linear) regression models predicting hourly pedestrian crossing volumes were constructed and applied, allowing for the calculation of annual average daily pedestrian volumes. Using these averages, a direct-demand pedestrian volume model was estimated incorporating various measures of the built environment,

land use, and socio-demographics. These models were subsequently used to predict pedestrian volumes at over 60,000 other intersections in Utah, offering estimates of pedestrian exposure for safety analysis.

This study used pedestrian push-button actuation data from signalized intersections in Oregon to develop estimates of pedestrian volumes. The next chapter describes the data collection process including the selection of sites.

3.0 DATA COLLECTION

This chapter describes the data collection process including assembling an inventory of locations, site selection, data collection protocol, data extraction, validation, and data integration.

3.1 INVENTORY OF LOCATIONS

To compile an inventory of locations where high-resolution data was being collected, the research team sent out an email requesting a list of such locations to various agencies across the state. The email solicitation was sent out to Oregon Department of Transportation (ODOT), Portland Bureau of Transportation (PBOT), Washington County, and the cities of Corvallis, Salem, Eugene and Medford. From these agencies, 803 locations were obtained from ODOT, PBOT, Washington County, and Salem, while the other cities reported not having high-resolution data. From these 803 locations, we further filtered sites to be in the Willamette Valley or close to the Portland metropolitan area (to facilitate the video data collection protocol in Section 3.3), and that had at least some pedestrian push-button activity: greater than or equal to 10 events per day, when available. This yielded a population size of 433 potential study locations, shown in Figure 3.1.

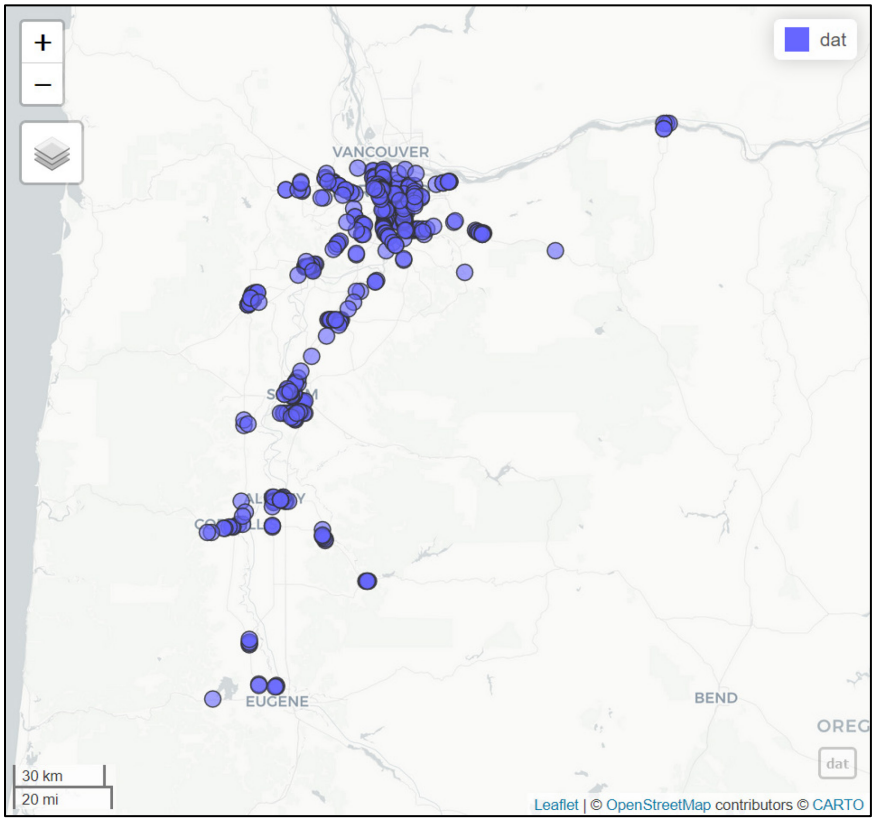


Figure 3.1: Map of potential study locations

3.2 SITE SELECTION

The objective of the site selection process was to select a variety of locations to study, to ensure coverage of different contexts, and to explore if there are differences in factors or accuracy across contexts. To achieve this objective, we assembled GIS datasets from a variety of sources (e.g., ODOT GIS, FHWA HPMS, Oregon Spatial Data Library). We linked attributes (listed below) from these datasets to each signal location from the inventory that we assembled, using geospatial matching. These attributes were chosen because of their impact on pedestrian volume, and because we wanted to sample locations to ensure variety across these attributes.

For each location, we noted the ODOT region, county, MPO, urbanized area, and city where the signal was located. Information on the presence of transit stops and schools near each signal location (if present) was also gathered. Roadway characteristics such as functional class, number of through lanes, speed limit, and AADT were obtained for the intersecting roadways where each signal was located. The land use and built environment context was captured through the place type variable, which was obtained from the Oregon Department of Land Conservation and Development (DLCD). The categories used for sampling were Low / non-MPO (Rural, Rural near Major Center, MPO Low Density, Isolated City, City near Major Center); Medium MPO (MPO Residential, MPO Employment); and High MPO (MPO Mixed Use, MPO TOD). Additionally, for ODOT signals, we processed raw traffic signal data to calculate the number of pedestrian push-button presses, using a 15-second filter, as has been done in previous studies (Singleton and Runa, 2021). These data were averaged to calculate AADP (annual average daily pedestrian events) for each signal. The signals were grouped into categories based on AADP: Low: < 100; Medium: 100–250; High: > 250. Overall, the attributes assembled were:

- ODOT Region
- County
- MPO
- Urbanized Area
- City
- Transit stops
- Schools
- Functional class
- Number of through lanes
- Speed limit
- AADT

- Place type (https://www.oregon.gov/lcd/CL/Pages/Place-Types.aspx)
- Pedestrian actuations (for ODOT signals)

For signals managed by the City of Portland (PBOT) and Washington County, we selected study locations manually. Part of this was based on staff recommendations, and part based on obtaining a variety of conditions from the attributes listed above. We selected 11 PBOT signals and three Washington County signals.

For signals managed by ODOT, we first filtered for those that were recording pedestrian signal data. Then, we grouped signals into nine categories, based on a 3x3 tabulation of place types and pedestrian activity. We then selected up to four signals from each of the nine categories. For categories with few options (less than 20 signals), signals were selected manually to ensure that study sites were in a variety of locations and not too close to each other. For categories with many options (3 of the 9), signals were selected using random sampling. Finally, a few (5) signals were switched due to selected signals being too close together. We selected a total of 35 ODOT signals.

Some additional locations were originally selected (6) and others even had videos recorded (5), but these 11 locations had to be removed prior to video data collection/extraction because of construction or a lack of available traffic signal data for the time period. In total, we ended up with 49 study locations. These study locations are mapped by agency in Figure 3.2 and listed in Table 3.1. (As noted in the table, one of these 49 locations was removed for lack of signal data.)

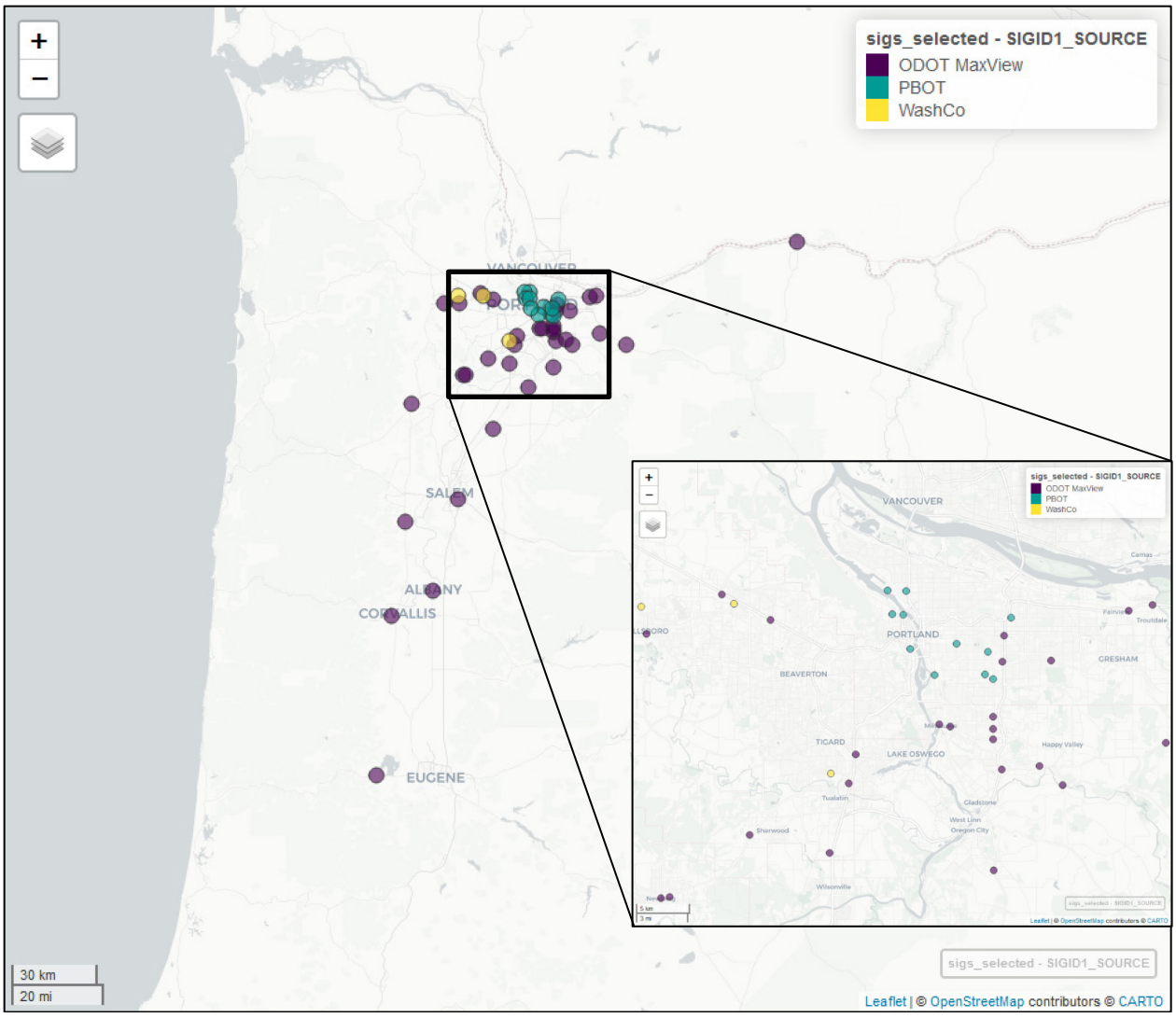


Figure 3.2: Map of selected study locations

Table 3.1: List of selected study locations

Agency	Local ID	Location	Start Date	End Date	Hours of Video	
					Collected	Used
ODOT	231 / 22534	OR8_Adair @ 10th_W	6/21/22	6/23/22	48	48
	244 / 03009	OR22 at Hawthorne	9/7/22	9/9/22	48	48
	408 / 03047	OR99W at Main St	6/6/22	6/8/22	48	48
	412 / 23071	OR224 @ Springwater Rd	6/29/22	7/1/22	48	48
	437 / 22143	I-84 EB Ramp @ 238th	6/16/22	6/18/22	48	48
	440 / 04051	Ellsworth at 6th	6/6/22	6/8/22	48	48
	445 / 22036	OR99E @ Harrison/17th Street	6/2/22	6/4/22	48	48
	449 / 22044	OR99E @ Ivy St	6/28/22	6/30/22	48	48
	455 / 22067	OR212 @ 135th	7/6/22	7/8/22	48	47
	462 / 23044	OR212@ Richey Rd	6/2/22	6/4/22	48	48
	470 / 22553	99W @ SunsetBlvd/ElwertRd	6/5/22	6/7/22	48	47
	505 / 04022	OR34 at 35th	6/6/22	6/8/22	48	48
	579 / 05046	OR126 at Territorial Hwy	6/21/22	6/23/22	48	48
	585 / 23024	US26 EB @ Meining	7/6/22	7/8/22	48	45
	619 / 22145	I-84 EB Ramp at 257th/S. Frontage Rd	6/16/22	6/18/22	48	48
	792 / 22404	OR8WB_Baseline @ 1st Ave (OR219)	6/21/22	6/23/22	48	48
	798 / 23051	I-84WB at 2nd Street	6/20/22	6/22/22	48	48
	827 / 03020	OR99E at Young	6/28/22	6/30/22	48	47
	862 / 22437	US26EB @ Bethany	6/27/22	6/29/22	46.75	45
	873 / 22093	OR213_82nd @ Otty	7/6/22	7/8/22	48	47
	875 / 03109	5th at Adams	6/28/22	6/30/22	48	48
	883 / 22064	OR212 @ 82nd Drive	7/12/22	7/14/22	48	48
	893 / 22468	OR141 @ 95th/Commerce Circle	9/8/22	9/10/22	48	48
	895 / 23011	US26 at 148th	6/16/22	6/18/22	48	48
	901 / 22104	I-205 SB at Stark Street	6/1/22	6/3/22	48	48
	902 / 22100	US26 at 92nd	6/21/22	6/23/22	47	47
	914 / 22052	OR 224 at Oak Street	9/7/22	9/9/22	48	48
	926 / 22471	US26EB@Cornelius Pass	6/24/22	6/26/22	48	48

	942 / 22326	OR217/Kruse@I-5NB	6/20/22	6/22/22	48	0*
	943 / 22463	OR141 @72nd/Bridgeport	6/20/22	6/22/22	48	48
	953 / 22085	OR213 at Sunnyside/Harmony	6/2/22	6/4/22	48	47
	957 / 22090	OR213 at Causey	7/6/22	7/8/22	48	47
	971 / 03051	1st at College	6/2/22	6/4/22	48	48
	978 / 03056	OR99W at Villa	6/6/22	6/8/22	43.88	43
	1040 / 22001	OR213 @ Molalla	6/1/22	6/3/22	48	48
PBOT	1026	N Interstate @ Going	9/8/22	9/10/22	48	48
	1095	N Anchor @ Channel	7/7/22	7/9/22	48	48
	2156	NE Halsey and NE 102nd	6/21/22	6/23/22	47	47
	3048	NW Vaughn @ 23rd	7/7/22	7/9/22	48	44
	3060	NW Naito at 15th Ave	9/8/22	9/10/22	48	48
	4025	SE Foster @ 72nd	6/21/22	6/23/22	48	48
	4044	SE Cesar Chavez @ Hawthorne	6/1/22	6/3/22	48	47
	4055	SE McLoughlin @ 17th	6/28/22	6/30/22	48	48
	4113	SE 82nd @ Foster	6/21/22	6/23/22	47	47
	4127	SE Division @ 76th	9/8/22	9/10/22	48	48
	5288	SW Naito @ Lincoln	6/7/22	6/9/22	48	48
Washington County	38	Durham – 92nd - Tigard	6/6/22	6/8/22	48	48
	32	Evergreen – Rock Creek Trail - Hillsboro	7/13/22	7/15/22	48	48
	40	Glencoe – North HS Driveway - Hillsboro	6/8/22	6/10/22	48	48
				Total	2,343.63	2,279

* Ultimately, this signal's video-extracted data were unable to be used due to signal data being unavailable.

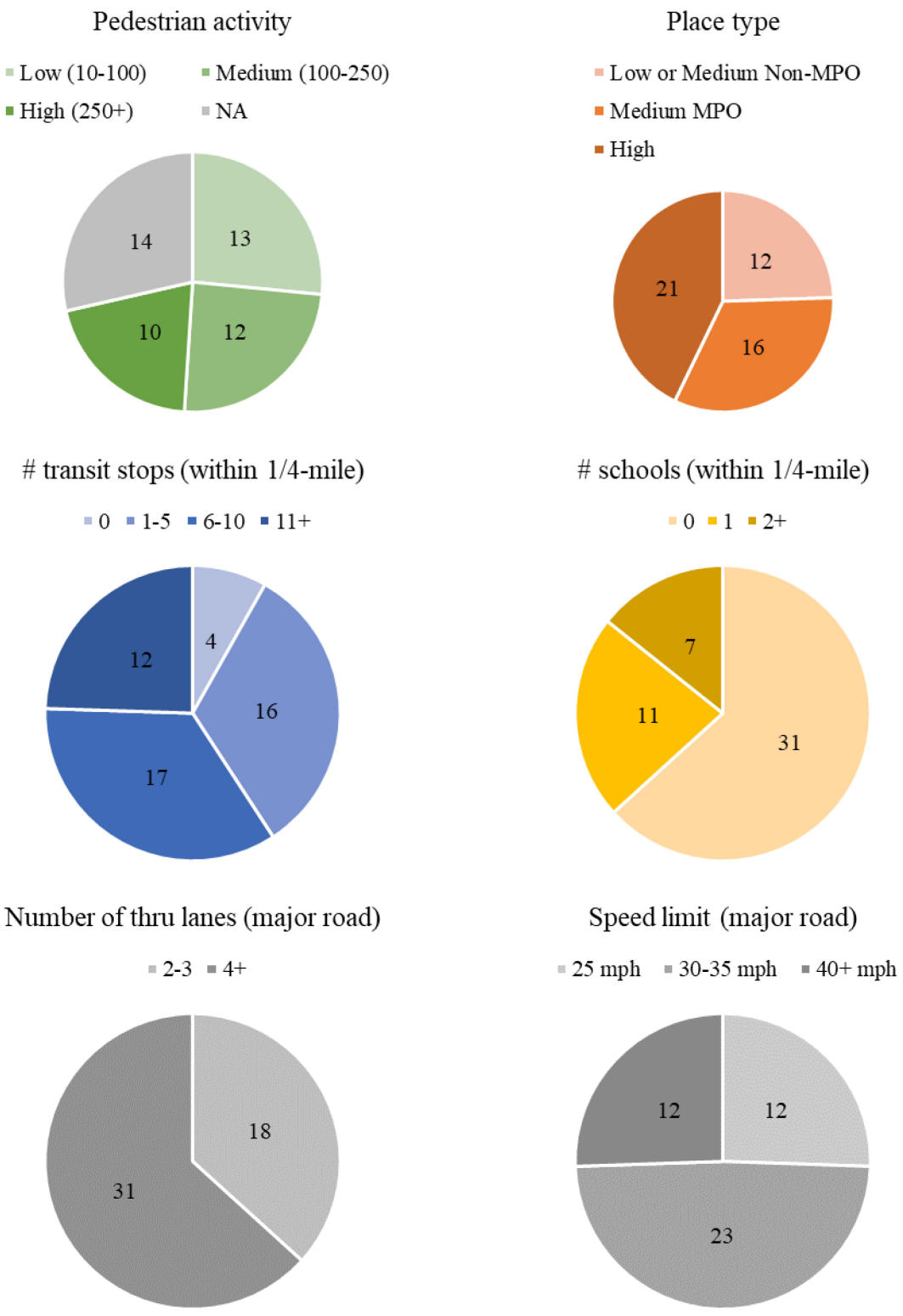


Figure 3.1: Study locations by various characteristics

The selected locations covered a variety of conditions, as shown in Figure 3.3. Locations were from all levels of pedestrian activity and place types, and included sites near multiple schools and transit stops and with a variety of major road through lanes and speed limits.

After the start of the data collection, another data source became available that added additional study locations to the project. Through separate ODOT efforts, 16-and 32-hour pedestrian volumes were collected (via videos) at several dozen intersections throughout eastern Oregon. Several of these counts occurred at signalized intersections where pedestrian push-button data were available. Therefore, 16 additional ODOT signals were added, as shown in Figure 3.4 and Table 3.2.

Regarding the distributions of location characteristics shown in Figure 3.3, the 16 extra locations had the following characteristics. All 16 additional signals had "low" pedestrian activity (< 100), were in "low or medium non-MPO" place types, and had 0 schools within a quarter-mile. For transit stops within a quarter mile, six locations had 0, eight had 1-5, one had 6-10, and one had 10+ transit stops nearby. The locations were evenly split between those with 2-3 thru lanes (eight) and 4+ thru lanes (eight) on the major road. Regarding speed limit, three locations had 25 mph, 11 locations had 30-35 mph, and 2 locations had 40+ mph.

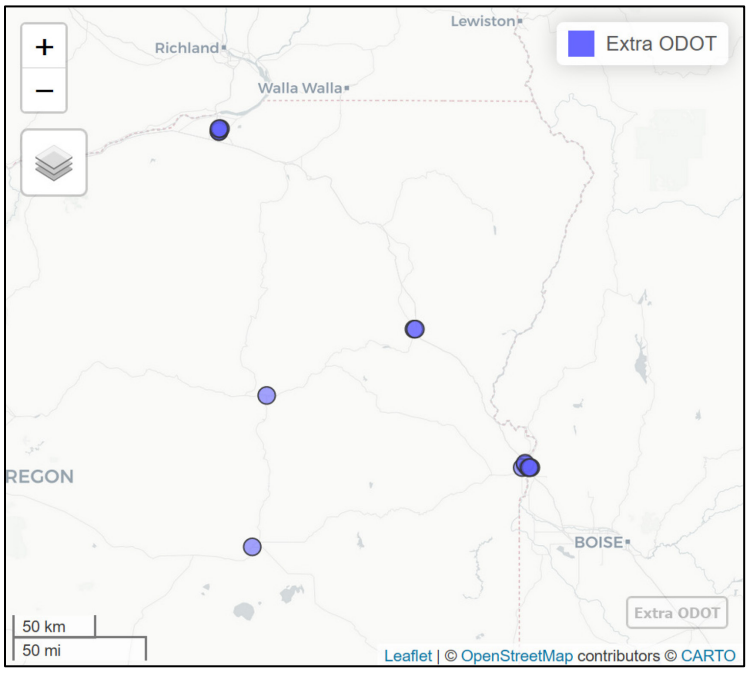


Figure 3.4: Map of extra study locations

Table 3.2: List of extra study locations

Agency	Local ID	Location	Dates	Hours of Video	
				Collected	Used
ODOT Extra	405 / 12071	OR207 (Elm Ave) @ 11th	4/20/22	16	16
	406 / 12072	OR207 (11th) @ Orchard Ave	4/20/22	16	16
	521 / 12070	OR207 (11th) @ Hermiston Ave	4/20/22	16	16
	717 / 12049	W Elm Avenue @ N 1st. Place	4/18/22	16	16
	718 / 12050	SW 11th. Street @ W. Highland Avenue	4/20/22	16	16
	753 / 14015	Central Oregon Hwy. @ Egan Avenue	5/16-17/22	32	32
	756 / 14020	US 26 - Main Street @ US 395 - Canyon Blvd.	5/24-25/22	32	32
	762 / 14034	E. Idaho Avenue @ E. 4th. Street	6/20-21/22	32	32
	763 / 14035	E. Idaho @ E.B. Ramp I-84	6/21-22/22	32	32
	764 / 14036	E. Idaho Avenue @ W.B. Ramp I-84	6/21-22/22	32	32
	765 / 14037	E. Idaho Avenue @ Goodfellow Lane	6/21-22/22	32	32
	766 / 14075	Olds Ferry Hwy. @ SW 4th. Avenue	7/18-19/22	32	32
	767 / 14076	Olds Ferry Hwy. @ Verde Drive	7/18-19/22	32	32
	770 / 14077	Olds Ferry Hwy. @ Washington Street	7/18-19/22	32	32
	781 / 13021	Campbell @ Cedar/Clark	3/29/22	16	16
	783 / 13020	Main Street @ Campbell	3/29/22	16	16
			Total	400	400

3.3 DATA COLLECTION PROTOCOL

At most of the selected sites (excluding the extra study locations), video data were collected for approximately 48 continuous hours, which captured both peak and off-peak hours, for pedestrian volumes at crosswalks for each location. Field data were collected using one video camera at each location, usually on a typical weekday during clear weather conditions. Sometimes, a weekend day was also captured.

A vendor (All Traffic Data) was used to assist with the video data collection. The research team worked with the vendor to ensure that camera angles and placement facilitated clear views of pedestrian crossings at each location. Once the video recordings were obtained, relevant data (e.g., pedestrian counts by time of day) were extracted from the videos as described in the following section.

3.4 DATA EXTRACTION

Once the videos were obtained from the vendor, researchers developed a plan to manually watch the videos and transcribe timestamped pedestrian crossing events, in order to obtain pedestrian crossing volumes to compare against pedestrian signal data.

A standardized data collection form was developed in Excel using Visual Basic macros for the transcription of the crossing events. For each location, a custom figure containing a screenshot of the intersection was developed and crosswalk name and pedestrian crossing directions were added to the figure to aid the researchers collecting the crossing data with noting the direction that the pedestrian was crossing (see Figure 3.5 below). Additionally, the data collection form had the intersection ID and the intersection name already coded in for each location. For each crossing event the coders had to manually input the following information:

- Crosswalk name (N, S, E, W)
- Direction of travel (N→S, S→N, E→W, W→E)
- Crosswalk user type (pedestrian, bicyclist, wheelchair user, skateboard, scooter, other)
- Number of users
- Date
- Timestamp 1 (time when pedestrian started crossing after stepping off the curb)
- Timestamp 2 (time when pedestrian finished crossing by stepping onto the curb)
- Notes (any possible significant observations about the crossing)

These options were included in the dropdown lists for each categorial variable, so that the coders could easily select the appropriate option. Once the coders finished entering all the information for one crossing event, they were instructed to press the submit button, which saved the crossing details in a new row in another sheet (named stored data).

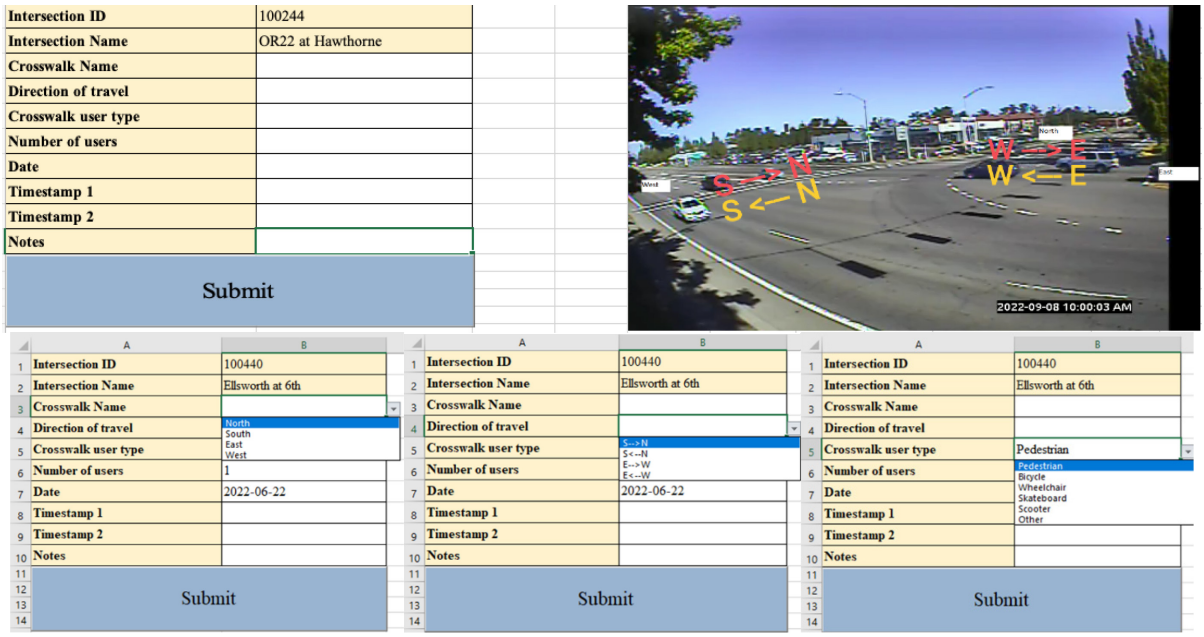


Figure 3.5: Video data extraction form

High-resolution traffic signal controller log data (Li et al., 2020) were also obtained, for the same locations and days as the videos were recorded. These were obtained from ODOT, PBOT, and Washington County staff. Each record contained the signal ID, the timestamp of the event, the event code, and the event parameter. In most cases, only the following event codes were obtained:

- 0: Phase On
- 21: Pedestrian Begin Walk
- 22: Pedestrian Begin Clearance
- 23: Pedestrian Begin Solid Don't Walk
- 45: Pedestrian Call Registered
- 90: Pedestrian Detector On
- 89: Pedestrian Detector Off

For event codes 0, 21, 22, 23, and 45, the event parameter is the corresponding phase number. For event codes 89 and 90, the event parameter is the pedestrian detector channel number, which in most cases is set to be the same as the phase number for the corresponding pedestrian indication.

The most relevant event for this study was event code 90, which represents an active detection (push-button press). These events will be discussed in more detail in the Data Integration section below.

3.5 DATA VALIDATION

After the video data were completely transcribed for each location, they were validated by a trained researcher (someone other than the person who collected the data), who randomly selected at least 5% (a minimum of 10) of all crossing events recorded for that location and crosschecked the entries with the videos. If systematic issues were found, they were corrected, and additional entries were checked. Figure 3.6 shows an example of the video validation form. The entries highlighted in green are the ones that were checked, and the entries with red text are examples where the timestamp and the number of users was incorrect or missing.

	B	C	D	E	F	G	H	I	J
1	Intersection ID	Intersection Name	Crosswalk Name	Direction of travel	Crosswalk user type	Number of users	Date	Timestamp 1	Timestamp 2
212	100408	OR99W at Main	South	E-->W	Pedestrian	1	6/7/2022	10:45:42 AM	10:45:57 AM
213	100408	OR99W at Main	South	E<--W	Bicycle	1	6/7/2022	10:49:29 AM	10:49:41 AM
214	100408	OR99W at Main	South	E<--W	Pedestrian	1	6/7/2022	10:49:29 AM	10:49:41 AM
215	100408	OR99W at Main	South	E<--W	Pedestrian	2	6/7/2022	10:53:10 AM	10:53:25 AM
216	100408	OR99W at Main	North	E-->W	Pedestrian	1	6/7/2022	10:53:14 AM	10:53:25 AM
217	100408	OR99W at Main	South	E<--W	Pedestrian	2	6/7/2022	11:12:03 AM	11:12:15 AM
218	100408	OR99W at Main	North	E-->W	Pedestrian	1	6/7/2022	11:33:02 AM	11:33:13 AM
219	100408	OR99W at Main	East	S<--N	Bicycle	1	6/7/2022	11:33:51 AM	11:34:00 AM
220	100408	OR99W at Main	North	E<--W	Pedestrian	2	6/7/2022	11:36:33 AM	11:36:46 AM
221	100408	OR99W at Main	East	S<--N	Pedestrian	1	6/7/2022	11:43:10 AM	11:43:23 AM
222	100408	OR99W at Main	North	E<--W	Bicycle	1	6/7/2022	11:44:00 AM	11:44:05 AM
223	100408	OR99W at Main	South	E-->W	Pedestrian	1	6/7/2022	11:44:17 AM	11:44:33 AM
224	100408	OR99W at Main	West	S<--N	Pedestrian	1	6/7/2022	11:50:30 AM	11:50:38 AM
225	100408	OR99W at Main	North	E<--W	Pedestrian	1	6/7/2022	11:51:48 AM	11:52:00 AM
226	100408	OR99W at Main	North	E<--W	Bicycle	2	6/7/2022	11:51:48 AM	11:52:00 AM
227	100408	OR99W at Main	South	E-->W	Pedestrian	1	6/7/2022	12:01:12 PM	12:01:25 PM
228	100408	OR99W at Main	South	E<--W	Pedestrian	1	6/7/2022	12:01:12 PM	12:01:25 PM
229	100408	OR99W at Main	South	E-->W	Pedestrian	1	6/7/2022	12:11:13 PM	12:11:24 PM
230	100408	OR99W at Main	North	E-->W	Pedestrian	1	6/7/2022	12:16:28 PM	12:16:41 PM
231	100408	OR99W at Main	South	E<--W	Pedestrian	1	6/7/2022	12:23:49 PM	12:24:03 PM
232	100408	OR99W at Main	West	S<--N	Pedestrian	1	6/7/2022	12:24:20 PM	12:24:29 PM
233	100408	OR99W at Main	East	S-->N	Pedestrian	1	6/7/2022	12:24:33 PM	12:24:46 PM
234	100408	OR99W at Main	South	E<--W	Pedestrian	1	6/7/2022	12:25:50 PM	12:26:01 PM
235	100408	OR99W at Main	West	S-->N	Pedestrian	1	6/7/2022	12:30:44 PM	12:30:51 PM
236	100408	OR99W at Main	East	S<--N	Pedestrian	1	6/7/2022	12:37:02 PM	12:37:14 PM
237	100408	OR99W at Main	North	E-->W	Scooter	1	6/7/2022	12:44:37 PM	12:44:41 PM

Figure 3.2: Video data validation

When extracting information from the videos, coders could note any strange or unusual behavior they observed that was not otherwise accounted for in the data collection form. The research team manually reviewed all of these notes and, if necessary, took action accordingly. In some cases, reviewing these notes led to edits, while in other cases, notes were made to remove that entry. Some examples of these instances include:

- Editing a record:
- If people using multiple modes were crossing at the same time, some coders marked that as one combined entry: Pedestrian/Bicycle. These were changed to be two separate entries, one as Pedestrian, one as Bicycle.
- Removing a record:

- Some coders recorded crossing events for people using a crosswalk across a channelized (but unsignalized) right-turn lane. Because such a crossing is not signalized, there is no traffic signal data to link to it, these entries were flagged for removal.

- Some coders recorded crossing events for people crossing outside of the crosswalk, such as a midblock crossing. While this information could be useful, it was not fair to include it in the final dataset, because some video views did not allow for observing any pedestrians crossing mid-block. Therefore, these events were only retained if the crosswalk user started from or ended on the sidewalk. Also excluded (flagged for removal) were any events where the crossing happened more than about a car's length away from the crosswalk.

In addition to these manual quality control procedures, the research team also performed several more semi-automated checks for other potential erroneous or missing data:

- Checking for missing data on key columns:

  - Identified and corrected any missing data for the crosswalk name, crosswalk user type, number of users, date, and the two timestamps.

- Checking for potentially erroneous dates and timestamps:

  - Checked the dates and timestamps against the start and end time of each video, and corrected any records that were outside of the video range.

  - Checked the timestamps against each other, corrected any records that had a zero or negative crossing time (time difference), and checked (and corrected as needed) any records with crossing times greater than 2 minutes.

3.6 DATA INTEGRATION

As data extraction and validation were completed for each location, count information obtained from the videos was merged and connected with data from high-resolution traffic signal controller event logs. These signal data include information on pedestrian push-button actuations, calls registered, and phasing information (see Data Extraction section above).

Automated scripts in R were developed to process these two datasets and compile information into one combined dataset for subsequent analysis. This process produced a combined dataset that included observations, by hour, for each signal (ID) and crosswalk (phase number). It matched the two datasets based on the signal ID, crosswalk / phase number, and times (Timestamp 1). The following information was produced by this process:

- General information

  - SIGNAL Signal ID for the intersection

- This was a PSU/USU internal adjustment of agency signal IDs.
- TIME1 Start time.
- TIME2 End time.
- TDIFF Duration in minutes (= TIME2 – TIME1)
- DIR Leg of crossing (usually one of: N, E, S, W)
- P Phase number associated with crossing (usually one of: 2, 4, 6, 8)

- Data collected from videos.
  - PED # people walking under their own power.
  - BIKE # people bicycling.
  - SCOOT # people using (e-)scooter.
  - SKATE # people using skateboard.
  - WHEEL # people in wheelchair
  - OTHER # crosswalk users not in above categories (e.g., golf cart, horse)
- Data assembled from traffic signal controller logs.
  - A00 # phase on (event code 0)
  - A21 # pedestrian begin walk (event code 21)
  - A45 # pedestrian call registered (event code 45)
  - A90 # pedestrian detector on (event code 90)
  - A45A # pedestrian actuations (imputed; # times 90 after 0 or 22)
  - A45B # pedestrian actuations (imputed; # times 90 after 0 or 21)
  - A45C # pedestrian actuations (imputed, # times 90 after 0)
  - A90A # unique pedestrian detections (imputed; # 90s at least 5 sec apart)
  - A90B # unique pedestrian detections (imputed; # 90s at least 10 sec apart)
  - A90C # unique pedestrian detections (imputed; # 90s at least 15 sec apart)

In prior similar research effort conducted in Utah (Singleton, Runa, and Humagain, 2020), these same data were extracted and integrated in the same way.

The "imputed" pedestrian actuations (A45A/B/C) were calculated to account for potential differences in controller behavior/programming regarding when a pedestrian call is registered (event code 45) and how it could affect pedestrian signal phasing. Most of the time, event code 45 happens immediately after the first time (during a cycle) when a push button is pressed (event code 90). However, in the Utah project, discussions with traffic signal engineering staff and indicated that some types of controllers behaved differently; for instance, only showing event code 45 at the time when coordination along a corridor started for the day. Also, when inspecting the raw controller log data from various signals in Oregon, some controllers did not log event code 45 when the phase was on pedestrian recall. Thus, there was a potential need to impute pedestrian calls registered from the corresponding push-button and pedestrian indication event codes (90, 21, 22, 23).

A related explanation is that these imputed pedestrian actuations also allow for different representations of how a pedestrian detection might be treated by the controller; see Figure 3.7 and Table 3.3. Let us assume phase 2 at signal 99 is operating without pedestrian recall, and with a long-enough green time that when the first pedestrian push-button press (event code 90) occurs at time 12:01:05, the walk indication immediately appears (event code 21). Let us also assume that the same (or a different) pedestrian presses the button at least once while the walk indication shows walk (21), flashing don't walk (22), and steady don't walk (23), prior to the start of phase 2 during the next cycle (event code 0). The imputed pedestrian actuations (A45A, A45B, A45C) would be calculated as follows. All three measures would count the first pedestrian detection (event code 90 at 12:01:05) that causes the walk indication to appear. A45A would ignore any 90s that occur while walk is showing (between event codes 21 and 22), under the assumption that the pedestrian has time to start crossing now and does not need a walk indication during the next cycle. But A45A would count the first 90 that occurs after flashing don't walk (22), assuming the pedestrian has to wait for the next cycle for a chance to start crossing. The difference for A45B is that it counts the first pedestrian detection (90) after the start of walk (21) and ignores any subsequent 90s until the next event codes 0 or 21. For A90C, the difference is that it only counts the first pedestrian detection (90) after the phase begins (0) and ignores all subsequent 90s until the phase begins again during the next cycle. In summary, this situation depicted in Table 3.3 would produce values of 2 for A45A, 2 for A45B, and 1 for A45C.

One additional note about the difference between A45B and A45C. Usually, the walk indication (event code 21) starts at the same time as the phase (0) parallel motor vehicle green (1), and thus A45B and A45C would be equivalent. However, this may not happen sometimes. One example is shown in Table 3.3: when 0 occurs before 21, due to the walk indication being called after the start of the phase. Another example might happen during low-volume conditions, when there is no demand from the side streets. In that situation, two or more walk sequences could occur (21-22-23) between subsequent event 0s for the main street phases. The pedestrian activity metric A45B considers these uncommon situations, whereas the metrics A45C effectively ignores them.

A final reason for doing these same calculations in this Oregon study is that the imputed A45B measure was better correlated with pedestrian volumes in the Utah study (than A45 itself) under

certain conditions: when the crosswalk was on pedestrian recall, rest-in-walk (Singleton and Runa, 2021).

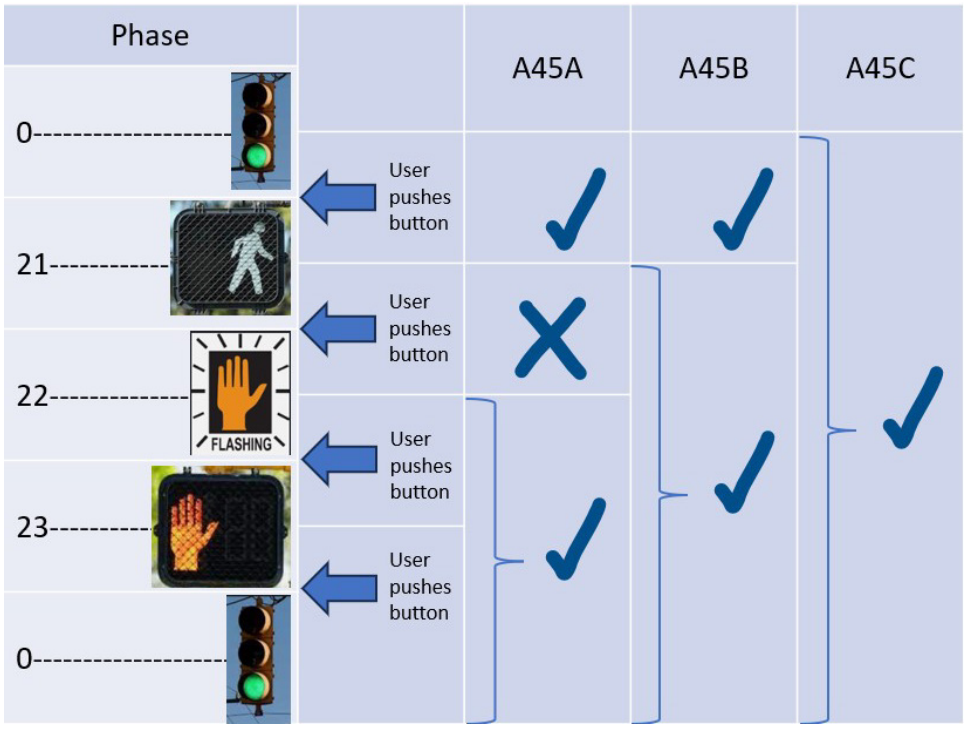


Figure 3.7: Meaning of imputed pedestrian actuations (A45A, A45B, A45C)

By using 5-, 10-, and 15-second time filters, the unique pedestrian detections (especially A90C) were calculated to account for the fact that some people press the push-button multiple times in quick succession, or sometimes the push-button can get stuck on for a few seconds. Using a time filter (e.g., 15 seconds = A90C) was shown to better predict pedestrian volumes than the raw number of push-button presses (A90) in the Oregon study (Singleton and Runa, 2021).

In the example of Table 3.3, all three metrics (A90A, A90B, A90C) would count the first detection (event code 90 at 12:01:05), assuming there haven't been any button-presses recently. The second detection occurs only 1 second later (at 12:01:06), so none of the metrics would count it. The time between the second and third button presses (12:01:18) is 12 seconds, so it would count for A90A and A90B, but not for A90C. The fourth button press counts for all three metrics, because it occurred 47 seconds later (12:02:05). The fifth button press occurs 6 seconds later (at 12:02:11), so it counts for A90A but not A90B or A90C. Finally, the time from the fifth to the sixth detection (at 12:02:24) is 13 seconds, so it counts for A90A and A90B but not A90C. Note that it has been more than 15 seconds since the last count for A90C (12:02:05 to 12:02:24), but the 15-second time threshold resets at each button press (event code 90), even if it does not count for that metric. So, overall, this pattern of signal controller events would give a value of 5 for A90A, 4 for A90B, and 2 for A90C.

Table 3.1: Example calculations of imputed pedestrian signal activity measures

Signal	Timestamp	Event Code	Event Parameter	A45A	A45B	A45C	A90A	A90B	A90C
99	01/01/2023 12:01:00.000	0	2	--	--	--	--	--	--
99	01/01/2023 12:01:05.000	90	2	1	1	1	1	1	1
99	01/01/2023 12:01:05.000	21	2	--	--	--	--	--	--
99	01/01/2023 12:01:05.500	89	2	--	--	--	--	--	--
99	01/01/2023 12:01:06.000	90	2	0	1	0	0	0	0
99	01/01/2023 12:01:06.500	89	2	--	--	--	--	--	--
99	01/01/2023 12:01:12.000	22	2	--	--	--	--	--	--
99	01/01/2023 12:01:18.000	90	2	1	0	0	1	1	0
99	01/01/2023 12:01:18.500	89	2	--	--	--	--	--	--
99	01/01/2023 12:01:30.000	23	2	--	--	--	--	--	--
99	01/01/2023 12:02:05.000	90	2	0	0	0	1	1	1
99	01/01/2023 12:02:05.500	89	2	--	--	--	--	--	--
99	01/01/2023 12:02:11.000	90	2	0	0	0	1	0	0
99	01/01/2023 12:02:11.500	89	2	--	--	--	--	--	--
99	01/01/2023 12:02:24.000	90	2	0	0	0	1	1	0
99	01/01/2023 12:02:24.500	89	2	--	--	--	--	--	--
99	01/01/2023 12:03:00.000	0	2	--	--	--	--	--	--
99	01/01/2023 12:03:00.000	21	2	--	--	--	--	--	--

At the same time as or just after completing data integration, the research team also performed additional quality control checks.

- The research team checked for missing traffic signal data during the periods when videos were recorded. This included tabulating the frequency of various signal event codes (e.g., 21 and 90) by phase and hour. Visual inspections did not indicate any systematic

missingness that would have been due to communication issues. This process identified a few signals where some of the phases were not recording push-button data. For example, signal 204044 (SE Cesar Chaves Blvd. & SE Hawthorne Blvd. in Portland, OR) only had push-button data associated with phase 4, not phases 2, 6, or 8.

- On four occasions, data had to be excluded for a period of time within the range of when the video was recorded. In two instances, this was because the video cut out (went dark) or was missing for a period of time: once for 4 minutes, once for 3 hours, and once for almost 4 hours. In a fourth instance, data was excluded for a 3-hour period around when a parade was happening that blocked traffic at the intersection being studied. In all of these instances, the research team removed or truncated the hourly records contained in the combined dataset.

- The research team also visually checked the relationship between the video-obtained crosswalk volume counts with the measures of pedestrian activity obtained from the traffic signal data. This was done overall and by signal and crossing/phase. Analysts were looking for events that stood out as being different from the rest, notably: events with high video volumes but low signal activity, or events with high signal activity but no video volumes. For events flagged using this visual inspection method, the completeness of the video data extraction and traffic signal controller log data was checked again, and corrections were made if necessary.

    - This process allowed for correcting some mis-matched crosswalks and phase numbers. It also identified that data for one video (100942) was actually from a different intersection, so it had to be excluded.

- Table 3.1 and Table 3.2 (shown earlier) report the number of hours of video data used at each signal: a total of 2,679 hours of video.

A few other details of this data integration process should be noted.

- At a few signals, multiple crossings were assigned the same (pedestrian) phase number. For example, on a one-way street, the two crossings parallel to the one-way street were sometimes given the same phase number. In this situation, there is no way from the signal data to know which pedestrian push-button press (event code 90) was for which of the two crosswalks. In these situations, the research team combined (added) the total video-observed crossing volumes for the two crosswalks, so the records in the combined dataset are technically organized by phase number (corresponding to usually one but sometimes two crosswalks).

- It is possible that the timestamps from the two sources (videos, signal controllers) did not line up perfectly. To confirm this, a sample of five events from each video were checked, and the video timestamps and the signal timestamps were compared. In many cases, the differences were negligible: just a couple of seconds, plus or minus. But for a few videos, the timestamps were anywhere from 1–3 minutes off. The implications of this were expected to be small, but there could be some difference for events happening near the start/end of an hour. Luckily, the time difference was consistent across each video

recording. Thus, the timestamps were adjusted accordingly before integrating the two data sources into the combined dataset.

3.7 SUMMARY

This chapter summarized the data collection process including assembling an inventory, criteria for selecting sites, and methods for data collection, extraction, validation and integration. Overall, video data was extracted from 49 locations where video data was collected for this study. Data from an additional 16 sites in eastern Oregon was combined with the data from the 49 sites to generate a combined data set for analysis. The next chapter details the data analysis process.

4.0 DATA ANALYSIS

This chapter describes the data analysis process, including the data analysis principles, approach, and sequential search process.

4.1 DATA ANALYSIS PRINCIPLES

The data analysis framework that the research team established adheres to the following four key principles. Each principle is accompanied by a set of defined metrics, which are described below.

4.1.1 Accuracy

The modeling framework should possess the capability to predict pedestrian volumes with low error and minimal bias, avoiding both underestimation and overestimation. Metrics to quantitatively measure the modeling framework's accuracy include correlation, root-mean-square error (RMSE), mean absolute error (MAE), symmetric mean absolute percentage error (SMAPE), and mean absolute scaled error (MASE).

4.1.2 Generalizability

The developed models should exhibit consistent performance on out-of-sample data, from different places and/or times. To achieve this, a hold-out sample and K-fold cross-validation approach will be used, dividing the data into training and testing datasets multiple times. Additionally, the results obtained will be compared to the findings from the Utah study (Singleton and Runa, 2021).

4.1.3 Simplicity

The developed models should use a limited number of inputs that are readily available, and both easily and consistently calculated. These models will utilize only one independent variable and test a few model segmentations (situations in which a different equation applies).

4.1.4 Intuitiveness

The variables and the relationships between them represented by the models should possess a clear and logical interpretation. The results obtained from the models must be easily comprehensible.

4.2 DATA ANALYSIS APPROACH

4.2.1 Dependent Variables (DVs)

The analysis used an hour as the time unit of analysis, and the pedestrian phase as the space unit of analysis. So, each observation represented one hour at a specific crosswalk within a given intersection. These definitions of an observation in this study are consistent with those used in the Utah study (Singleton and Runa, 2021). Future research could explore alternative

approaches, such as using different time intervals (e.g., 15-minute periods) or considering pedestrian volumes for the entire intersection.

The dependent variables (DVs) in this study comprised measures of observed pedestrian volumes. However, as noted in Section 3.4, data collection incorporated all crosswalk users, including those utilizing a variety of active transportation and micromobility modes. Therefore, assorted definitions of a "pedestrian" could be applicable. In this study, three distinct DVs were considered:

- PED: This variable includes only pedestrians, referring to individuals who are walking under their own power.
- UPED: This alternative definition includes pedestrians, skateboard users, and wheelchair users. It aligns with the approach used in the Utah study (Singleton and Runa, 2021).
- TOTAL: This outcome encompasses all sidewalk and crosswalk users observed, comprising pedestrians, bicyclists, scooter users, skateboard users, wheelchair users, and other users (e.g., OneWheel).

4.2.2 Independent Variables (IVs)

To maintain the generalizability of the analysis, only measures of pedestrian activity taken from traffic signal controller log data were considered as independent variables (IVs). For simplicity, each model focused on a single variable at a time. Initially, eight distinct IVs were evaluated, as previously defined:

- A45: # pedestrian call registered (event code 45)
- A45A: # pedestrian actuations (imputed; # times 90 after 0 or 22)
- A45B: # pedestrian actuations (imputed; # times 90 after 0 or 21)
- A45C: # pedestrian actuations (imputed, # times 90 after 0)
- A90: # pedestrian detector on (event code 90)
- A90A: # unique pedestrian detections (imputed; # 90s at least 5 sec apart)
- A90B: # unique pedestrian detections (imputed; # 90s at least 10 sec apart)
- A90C: # unique pedestrian detections (imputed; # 90s at least 15 sec apart)

4.2.3 Model Segmentation

The research team also investigated whether the relationships between the DV and IV(s) varied by context. To accomplish this, additional variables were employed to segment the data, and separate models were estimated for each segment. Statistically, this segmentation approach is equivalent to incorporating these variables as interaction terms in the models, where the intercept and IV terms are multiplied by these variables. The segmentation variables used in this analysis included signal status parameters (pedestrian recall and rest-in-walk, cycle length), temporal elements (time-of-day, day-of-week), and spatial characteristics (region, place type, proximity to transit stops and schools). Table 4.1 lists the segmentation variables considered and their respective descriptions.

Table 4.1: Segmentation variables in the modeling process

Variable	Description
RECALL	If pedestrian phase is imputed to be on pedestrian recall
CYCLE090[a]	If average cycle length is longer than 90 seconds
CYCLE120[a]	If average cycle length is longer than 120 seconds
PEAKAM	If hour-of-day is AM peak hour (7, 8)
PEAKPM	If hour-of-day is PM peak hour (16, 17)
PEAKAMPM	If hour-of-day is either AM or PM peak hours
OVERNIGHT	If hour-of-day is overnight (21, 22, 23, 0, 1, 2, 3, 4, 5)
WEEKEND	If day-of-week is Saturday or Sunday
ODOTREG	ODOT Region: 1: Portland Metro; 2: Willamette Valley and North Coast; 3: Southwestern Oregon; 4: Central Oregon; and 5: Eastern Oregon.
PLACETYPE3	Oregon Department of Land Conservation and Development (DLCD) place type[b], combined into three categories: Low or non-MPO (Rural, Rural Near Major Center, MPO Low Density, Isolated City, City near Major Center); Medium MPO (MPO Residential, MPO Employment); and High MPO (MPO Mixed Use, MPO TOD)
TRANSIT4C	If any transit stop is present within 400 meters
TRANSIT8C	If any transit stop is present within 800 meters
EDTOTAL4C	If any educational institution is present within 400 meters
EDTOTAL8C	If any educational institution is present within 800 meters

[a] Cycle length was an approximate average, calculated by dividing the time interval duration (usually one hour) by the number of event code 0s (phase on) that occurred within that time period.
[b] https://www.oregon.gov/lcd/cl/pages/place-types.aspx

4.2.4 Functional Forms (FFs)

To adequately capture the underlying relationship between the DV(s) and IV(s) in the dataset, it is necessary to determine the most applicable structure of the model, commonly referred to as the model's functional form (FF). Alongside linear relationships, numerous common non-linear relationships that can be modeled by incorporating polynomial terms, logarithmic transformations, or other mathematical functions. In many cases, non-linear relationships can be modeled as if they were linear relationships, but include transformations of the DV and/or IV(s).

Furthermore, given the nature of the DV (non-negative integers), relationships appropriate for count data outcomes are also possible. As a result, eight different FFs were examined:

- Linear
- Piecewise linear
- Quadratic
- Cubic
- Exponential
- Power
- Poisson
- Negative binomial

In all instances, the models included an intercept term. This decision enabled the models to predict some (non-zero) level of pedestrian volume, even when there was zero pedestrian signal activity for a particular observation. This choice differed from the approach adopted in the Utah study, where an intercept term was not included. Consequently, in the Utah model, zero pedestrian signal activity would always result in a prediction of zero pedestrian volume. The reason for including the intercept term in this study was to account for observed scenarios, typically with low pedestrian volumes, where some people crossed during the hour but no one pressed the pedestrian push-button. Such situations reflect real-world occurrences that may be encountered in the field.

Each of the eight FFs used in the models is described in more detail in the following paragraphs.

Linear: Linear regression is a commonly-used functional form to represent a linear relationship. The magnitude of the relationship between an independent variable (X) and a dependent variable (Y) is represented by an estimated slope coefficient (b). Excluding the error term, but including the intercept (a), a linear model is described with the following equation:

$$\boldsymbol{Y = a + bX}$$

Piecewise linear: A piecewise linear function is composed of multiple distinct linear segments, each defined over a specific interval. It can be visualized as a series of linear regressions linked together at specific points. As a result, this functional form is able to represent non-linear relationships by having different slopes for specific ranges of the independent variable. A piecewise linear model with two different slopes (b_1, b_2) meeting at a single breakpoint (c) is represented in the following equation:

$$\boldsymbol{Y = \begin{cases} a + b_1 X & if\ X \leq c \\ a + b_1 X + (b_2 - b_1)(X - c) & if\ X > c \end{cases}}$$

Quadratic: A quadratic functional form depicts non-linear relationships that follow a parabolic curve. In simple terms, it extends the linear functional form by incorporating a squared or quadratic term (X^2) of the independent variable to the model:

$$\boldsymbol{Y = a + bX + cX^2}$$

Cubic: More advanced polynomial functions can describe even more complex relationships between dependent and independent variables. The cubic form adds a third cubed term (X^3) to the quadratic model and is described by the following equation:

$$\boldsymbol{Y = a + bX + cX^2 + dX^3}$$

Exponential: An exponential function represents a non-linear relationship in which the dependent variable increases exponentially as an independent variable increases. While the exponential form itself is not a linear model, it can be transformed into one and estimated using the same procedures as linear regression (ordinary least squares):

$$\boldsymbol{Y = ae^{bX} \rightarrow \ln Y = \ln a + bX}$$

Power: A power function is another type of non-linear model. When evaluating a power function model, the exponent (or power) of the independent variable is also an estimated coefficient. Just like the exponential model, the power functional form can also be transformed into a linear equivalent, as described mathematically below:

$$\boldsymbol{Y = aX^b \rightarrow \ln Y = \ln a + b \ln X}$$

Poisson: Poisson regression is a frequently used method to model count data (non-negative integers). This functional form assumes that the count dependent variable (Y_i) follows a Poisson distribution:

$$\boldsymbol{P(Y_i) = e^{(-\lambda_i)} {\lambda_i}^{Y_i} \div Y_i!}$$

The expected value (mean) of this distribution is lambda (λ_i) and can be predicted by an independent variable (X_i) according to the following equation:

$$\boldsymbol{\lambda_i = e^{a+bX_i}}$$

Negative binomial: Poisson regression is occasionally restrictive due to the assumption of the Poisson distribution, where the mean and variance are equal (both to λ_i). However, in real-world situations, count data frequently exhibit overdispersion, meaning that the variance is greater than the mean. To account for overdispersion, the count dependent variable can be assumed to follow a negative binomial distribution (too complicated to describe in detail here). This causes a minor change to the relationship between the expected value of the dependent variable (Y_i) and an independent variable (X_i) as described below:

$$\boldsymbol{\lambda_i = e^{a+bX_i+\varepsilon_i}}$$

Note the inclusion of a new error term, e^{ε_i} , which is Gamma-distributed with a mean of 1 and variance of α. Thus, the negative binomial model estimates an additional overdispersion parameter which allows the mean (λ_i) to be less than the variance ($\lambda_i + \alpha{\lambda_i}^2$). When $\alpha = 0$, the negative binomial and Poisson models are equivalent.

4.2.5 Model Validation

To assess the goodness of fit of the models and the accuracy of their predictions compared to the observed data, various statistics were calculated. To avoid overfitting and assess the models' generalizability, the research team used a robust K-fold (with $K = 10$) cross-validation approach against hold-out data (Browne, 2000). In this cross-validation method, the data were randomly divided into 10 groups or folds. Ten separate models were then estimated, each using a different 90% of the data for training and the remaining 10% of the data for testing. The folds were randomly sampled from the total 65 study locations, resulting in each fold containing an unequal number of observations from either six or seven signals. After estimating each fold's model on its respective 90% training data, it was applied to predict the DV values of the remaining 10% testing data. Validation statistics were then calculated using the results for the testing data, comparing predicted and observed outcomes (out-of-sample validation). The same statistics were also calculated for the within-sample condition, allowing for a comparison. Finally, the validation statistics were averaged across all 10 folds' models to obtain overall measures of model performance. Five validation statistics were considered in this analysis:

- Correlation.
- Root-mean-square error.
- Mean absolute error.
- Symmetric mean absolute percentage error.
- Mean absolute scaled error.

Each of the five validation statistics is defined below.

Correlation (COR): The Pearson (product-moment) correlation coefficient assesses the strength of the linear association between two variables. It is a standardized measure, ranging from −1 (indicating a perfect negative relationship) to +1 (representing a perfect positive relationship). In the context of the present study, a correlation closer to +1 is more favorable. The correlation between the observed dependent variable (Y_i) and the predicted dependent variable ($\hat{Y}_i$) is:

$$\boldsymbol{r_{Y_i,\hat{Y}_i}} = \boldsymbol{\sum_{i=1}^{n}\left[(Y_i - \overline{Y})\left(\hat{Y}_i - \overline{\hat{Y}}\right)\right] \div \sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \sum_{i=1}^{n}\left(\hat{Y}_i - \overline{\hat{Y}}\right)^2}}$$

Root-mean-square error (RMSE): RMSE is a widely used measure to evaluate the accuracy of a models' predictions. As the name suggests, it is the square root of the average of the squared errors, which are the differences between the observed and predicted dependent variables ($\hat{Y}_i - Y_i$). Lower values of RMSE indicate less error and greater accuracy, although the specific values

depend on the scale of the data. It is important to note that RMSE is sensitive to outliers, since it depends on the squared errors. In this study, RMSE can be calculated using the following equation:

$$\boldsymbol{RMSE} = \sqrt{\sum_{i=1}^{n}\left(\widehat{\boldsymbol{Y}}_i - \boldsymbol{Y}_i\right)^2 \div \boldsymbol{n}}$$

Mean absolute error (MAE): MAE is another typically-used measure of model prediction accuracy. As the name implies, it represents the average of the absolute value of the errors, making it less sensitive to outliers compared to RMSE. Similar to RMSE, lower values of MAE imply less error and greater accuracy. One advantage of MAE is that its value is easily interpretable in the units of the original variable. Here, MAE is calculated as:

$$\boldsymbol{MAE} = \sum_{i=1}^{n}\left|\widehat{\boldsymbol{Y}}_i - \boldsymbol{Y}_i\right| \div \boldsymbol{n}$$

Symmetric mean absolute percentage error (SMAPE): SMAPE is a model prediction accuracy statistic that aims to consider the relative magnitude of the errors. It is essentially the average of the absolute values of the errors, expressed as a proportion. However, to avoid calculation issues when the denominator (observed value) is zero, the proportion calculation instead divides by the average of the absolute values of the observed and predicted values. This makes SMAPE "symmetric" to error, as it can be expressed as a proportion of either the observed value or the predicted value. Values of SMAPE range from 0 to 2, with values closer to 0 indicating greater accuracy. Values less than 1 imply that the amount of error is usually less than the average value of the (observed and predicted) dependent variable. SMAPE can be calculated as follows:

$$\boldsymbol{SMAPE} = \sum_{i=1}^{n}\left(\left|\widehat{\boldsymbol{Y}}_i - \boldsymbol{Y}_i\right| \div \left[\left(|\boldsymbol{Y}_i| + \left|\widehat{\boldsymbol{Y}}_i\right|\right) \div \boldsymbol{2}\right]\right) \div \boldsymbol{n}$$

Mean absolute scaled error (MASE): As an additional measure of a model's accuracy, MASE builds upon MAE and scales it relative to a baseline, which (in this study) is the mean absolute deviation (MAD) of the data (i.e., the average of the absolute values of the data around its mean). This scaling ensures that MASE values are scale independent and easily interpretable. For instance, a value less than 1 indicates that the model's predictions have less error than the deviations in the original data. Furthermore, MASE is less sensitive than SMAPE to observations with a value of zero. In this study, MASE can be calculated using the following equation:

$$\boldsymbol{MASE} = \boldsymbol{MAE} \div \boldsymbol{MAD} = \left(\sum_{i=1}^{n}\left|\widehat{\boldsymbol{Y}}_i - \boldsymbol{Y}_i\right| \div \boldsymbol{n}\right) \div \left(\sum_{i=1}^{n}|\boldsymbol{Y}_i - \overline{\boldsymbol{Y}}| \div \boldsymbol{n}\right)$$

4.3 SEQUENTIAL SEARCH PROCESS

In this study, a vast number of possible models existed due to the three dependent variables, eight independent variables, eight functional forms, and 14 segmentation variables considered (as described in Sections 4.2.1, 4.2.2, and 4.2.4), as well as the 10 folds used in the K-fold cross-validation process. To simplify the search process and identify the most promising model or set of models for further examination, a sequential search process was employed, as follows.

Step 1 involved narrowing down the list of potential dependent variables, independent variables, and functional forms. K-fold cross-validation was performed on the 192 unique combinations (3 DVs × 8 IVs × 8 FFs). Validation statistics were calculated for each set of models and averaged across models for each DV, IV, and FF. Based on the validation statistics, the best performing variables (3 DVs, 2 IVs) and functional forms (3 FFs) were selected to proceed to the next step.

In Step 2, the 14 segmentation variables were added to the models (as interaction terms) and tested (using K-fold cross-validation) with each combination of models that passed the first step. Validation statistics were again calculated for each of the 252 sets of models (3 DVs × 2 IVs × 3 FFs × 14 segmentation variables). This time, performance was assessed using both the validation statistics themselves and the improvements in the validation statistics resulting from the segmentation. This process further filtered the options to 3 DVs, 1 IV, 1 FF, and 4 segmentation variables (with an additional option of no segmentation).

In the final Step 3, the remaining 15 options (3 DVs × 1 IV × 1 FF × 5 (non-)segmentation options) were more closely examined across several criteria: the performance of the validation statistics, any practical differences resulting from segmentation, the intuitiveness of the results, the ability of the variables to be used across contexts and into the future, and model complexity. Ultimately, one model was selected with variations for the three different dependent variables. The recommended models were estimated using the full (100%) dataset.

4.4 SUMMARY

This chapter provided an overview of the data analysis process, guided by the principles of accuracy, generalizability, simplicity, and intuitiveness. The data analysis process involved exploring three dependent variables, eight independent variables, eight functional forms, and 14 segmentation variables. To ensure reliable model performance, 10-fold cross-validation was employed, and validation statistics were used to help identify the best-performing models. In the next chapter, the results of this data analysis process will be presented, including the recommended model(s). Additionally, the chapter will discuss the key findings derived from the analysis.

5.0 RESULTS

This chapter presents descriptive statistics obtained from the data collection, as well as the results of the sequential search process applied during data analysis, culminating in the final recommended models.

5.1 DESCRIPTIVE STATISTICS

Using the reduced video data, each crosswalk location was coded for pedestrian, bicyclist, scooter, wheelchair, and skateboard volumes according to the criteria described earlier (see Section 3.4), and the results are presented in Table 5.1. Note that only pedestrian count information was received from the additional ODOT locations in Eastern Oregon.

Table 5.1: Observed User Volumes

Local ID	Location	Pedestrian	Bicyclist	Scooter	Skateboard	Wheelchair
Oregon Department of Transportation						
231	OR8_Adair @ 10th_W	986	63	7	9	5
244	OR22 at Hawthorne	142	85	7	4	3
408	OR99W at Main St	593	82	12	5	7
412	OR224 @ Springwater Rd	62	5	0	0	0
437	I-84 EB Ramp @ 238th	260	23	2	0	2
440	Ellsworth at 6th	282	50	7	8	10
445	OR99E @ Harrison/17th Street	388	152	8	1	4
449	OR99E @ Ivy St	406	108	1	0	3
455	OR212@ 135th	190	34	1	0	1
462	OR212@ Richey Rd	81	65	0	0	0
470	99W @ SunsetBlvd/ElwertRd	185	19	17	8	0
505	OR34 at 35th	322	555	5	12	0
579	OR126 at Territorial Hwy	237	60	9	9	16
585	US26 EB @ Meining	1,907	44	7	8	3
619	I-84 EB Ramp at 257th/S. Frontage Rd	450	115	4	0	0
792	OR8WB_Baseline @ 1st Ave (OR219)	623	34	6	19	10
798	I-84WB at 2nd Street	1,403	38	12	10	5
827	OR99E at Young	256	46	0	0	2
862	US26EB @ Bethany	134	65	4	4	0
873	OR213_82nd @ Otty	910	153	10	11	19
875	5th at Adams	633	86	9	5	1
883	OR212 @ 82nd Drive	668	145	16	3	0
893	OR141 @ 95th/Commerce Circle	61	26	0	0	0
895	US26 at 148th	294	53	0	0	7
901	I-205 SB at Stark Street	451	223	18	6	3
902	US26 at 92nd	1,627	172	12	19	15
914	OR224 at Oak Street	380	86	13	3	5
926	2B471 US26EB@Cornelius Pass	97	92	3	6	0
943	OR141 LBF@72nd/Bridgeport	350	36	6	8	0

Local ID	Location	Pedestrian	Bicyclist	Scooter	Skateboard	Wheelchair
953	OR213 at Sunnyside/Harmony	243	41	0	2	2
957	OR213 at Causey	1,153	104	27	4	31
971	1st at College	2,471	40	4	3	1
978	OR99W at Villa	457	95	5	8	6
1040	OR213 @ Molalla	289	36	9	8	7
Portland Bureau of Transportation						
1026	N Interstate @ Going	2,241	356	22	17	9
1095	N Anchor @ Channel	147	35	0	0	0
2156	NE Halsey and NE 102	1,491	133	45	24	23
3048	NW Vaughn @ 23rd	1,455	44	17	0	10
3060	NW Naito Ave at 15th Ave	997	55	12	0	1
4025	SE Foster @ 72nd	1,968	228	15	14	4
4044	SE Cesar Chavez @ Hawthorne	1,414	37	7	6	4
4055	SE McLoughlin @ 17th	243	244	8	5	0
4113	SE 82nd @ Foster	1,781	130	19	7	10
4127	SE Division @ 76th	304	136	7	2	0
5288	SW Naito @ Lincoln	647	219	43	6	3
Washington County						
38	Durham @ 92nd, Tigard	330	33	3	0	1
32	Evergreen @ Rock Creek Trail, Hillsboro	497	287	6	1	0
40	Glencoe @ North HS Driveway, Hillsboro	1,274	64	4	13	0
ODOT – Eastern Oregon						
405	OR207 (Elm Ave) @ 11th	10	Non-pedestrian data not collected.			
406	OR207 (11th) @ Orchard Ave	22	Non-pedestrian data not collected.			
521	OR207 (11th) @ Hermiston Ave	19	Non-pedestrian data not collected.			
717	W Elm Avenue @ N 1st. Place	12	Non-pedestrian data not collected.			
718	SW 11th. Street @ W. Highland Avenue	49	Non-pedestrian data not collected.			
753	Central Oregon Hwy. @ Egan Avenue	181	Non-pedestrian data not collected.			
756	US 26 Main St. @ US 395 Canyon Blvd.	604	Non-pedestrian data not collected.			
762	E. Idaho Avenue @ E. 4th. Street	231	Non-pedestrian data not collected.			

Local ID	Location	Pedestrian	Bicyclist	Scooter	Skateboard	Wheelchair
763	E. Idaho @ E.B. Ramp I-84	179	Non-pedestrian data not collected.			
764	E. Idaho Avenue @ W.B. Ramp I-84	188	Non-pedestrian data not collected.			
765	E. Idaho Avenue @ Goodfellow Lane	341	Non-pedestrian data not collected.			
766	Olds Ferry Hwy. @ SW 4th. Avenue	19	Non-pedestrian data not collected.			
767	Olds Ferry Hwy. @ Verde Drive	8	Non-pedestrian data not collected.			
770	Olds Ferry Hwy. @ Washington Street	5	Non-pedestrian data not collected.			
781	Campbell @ Cedar/Clark	96	Non-pedestrian data not collected.			
783	Main Street @ Campbell	23	Non-pedestrian data not collected.			
	Total	**35,767**	**5,032**	**449**	**278**	**233**

In total, 41,843 users were observed, with the following categorization: 35,767 pedestrians, 5,032 bicyclists, 449 scooter users, 278 skateboard users, 233 wheelchair users, and 84 other users (e.g., OneWheel), as presented in Table 5.1.

Figure 5.1 illustrates the time-of-day distributions at each signal (light grey lines), and the overall mean time-of-day distribution across all signals (thicker red line) for all crosswalk users. The data suggests that crosswalk activity tends to peak during the afternoon and early evening hours, specifically between 3pm and 6pm.

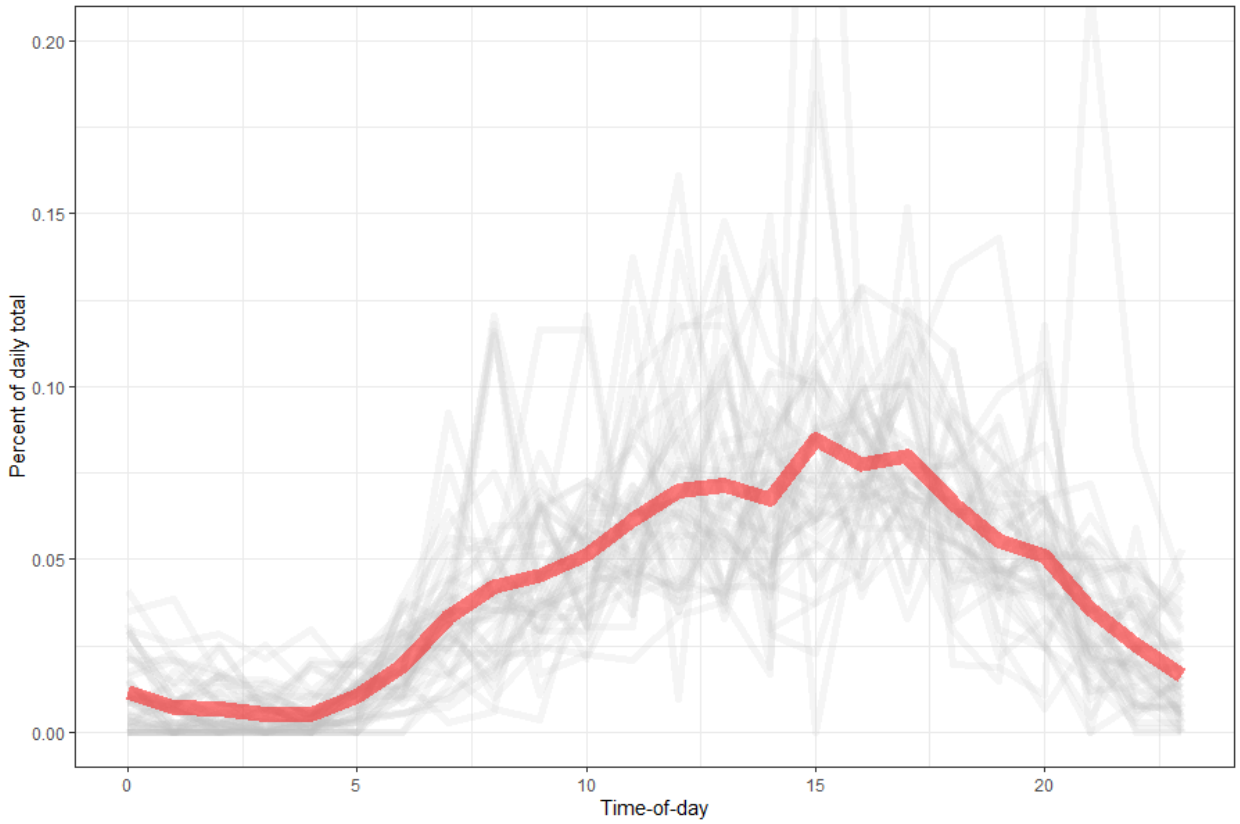


Figure 5.1: Time-of-day distributions across study locations

The scatterplots below depict the relationships between pedestrian push-button activity from traffic signal data (x-axis) and the number of observed crosswalk users from the videos (y-axis). Each point on the figures corresponds to one hour at one crosswalk at a specific intersection. Figure 5.2 displays the data for pedestrians (individuals walking or running under their own power). Figure 5.3 includes data for people using skateboards and wheelchairs. Lastly, Figure 5.4 represents the total count of all crosswalk users.

Across all figures, a consistent trend emerges: an increase in push-button activity is associated with a rise in pedestrian and crosswalk user volume. Additionally, there seems to be a somewhat non-linear pattern, resembling a quadratic or exponential relationship; however, this observation could be influenced by a few locations with exceptionally high activity and volume.

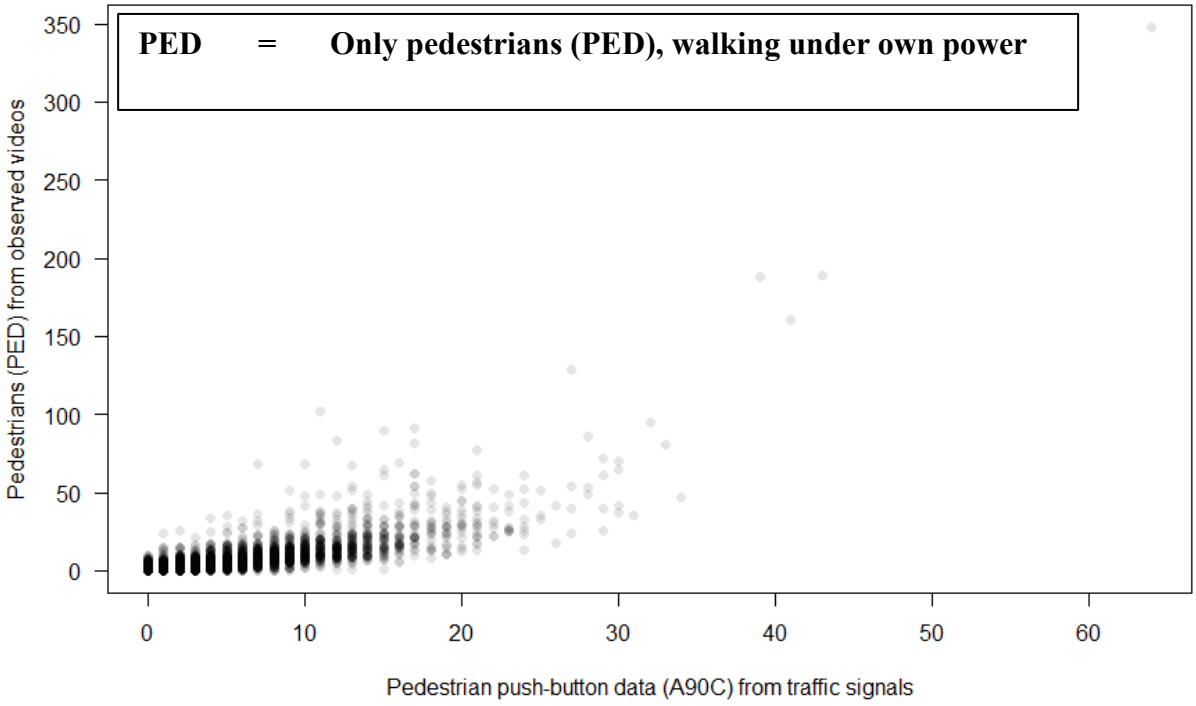


Figure 5.2: Pedestrian signal activity vs. only pedestrian volumes

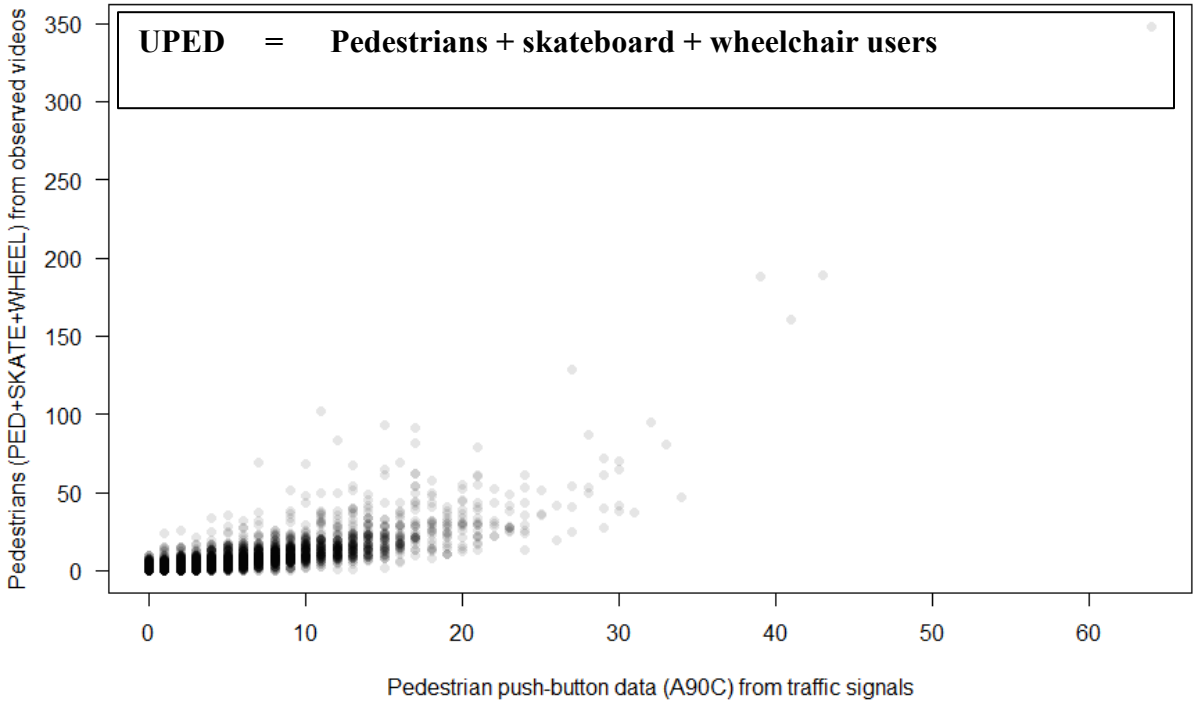


Figure 5.3: Pedestrian signal activity vs. pedestrian, skateboard, and wheelchair user volumes

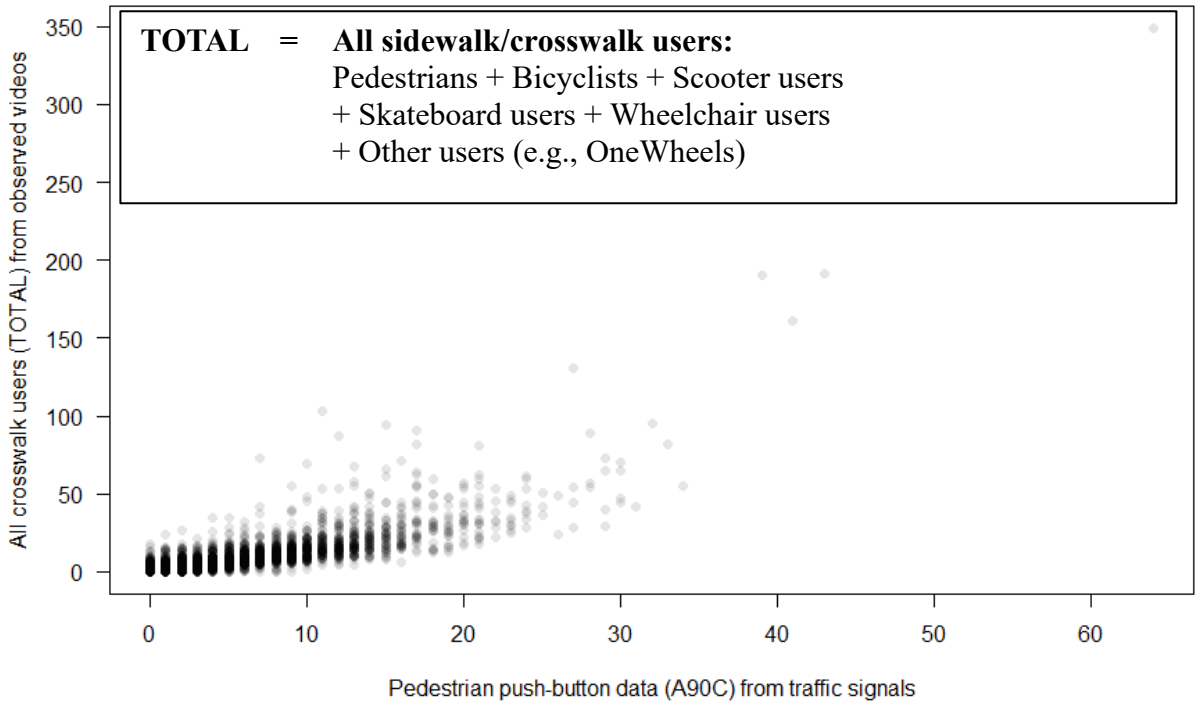


Figure 5.4: Pedestrian signal activity vs. all sidewalk/crosswalk user volumes

5.2 RESULTS, STEP 1: DEPENDENT/INDEPENDENT VARIABLES AND FUNCTIONAL FORMS

As mentioned earlier, the initial step in the data analysis process involved evaluating all 192 combinations of three dependent variables, eight dependent variables, and eight functional forms, using 10-fold cross-validation and calculating several validation statistics. However, upon careful examination, the research team decided to exclude the exponential, Poisson, and negative binomial FFs due to their substantially poorer fits and validation statistic results. Therefore, Table 5.2 presents the averages of the validation statistics for the remaining 120 sets of models (3 DVs × 8 IVs × 5 FFs) under consideration.

Table 5.1: Average validation statistics for models in Step 1

		Using estimation data					Using validation data				
Option		**COR**	**RMSE**	**MAE**	**SMAPE**	**MASE**	**COR**	**RMSE**	**MAE**	**SMAPE**	**MASE**
DV	**PED**	0.783	5.638	2.351	1.106	0.876	0.779	5.430	2.462	1.100	0.906
	UPED	0.786	5.641	2.360	1.100	0.871	0.782	5.432	2.470	1.094	0.899
	TOTAL	0.804	5.690	2.468	1.025	0.831	0.799	5.516	2.571	1.020	0.860
IV	**A45**	0.765	6.047	2.654	1.090	0.953	0.698	6.083	2.794	1.090	1.005
	A45A	0.830	5.254	2.241	1.073	0.805	0.807	5.052	2.339	1.065	0.834
	A45B	0.837	5.157	2.204	1.067	0.792	0.810	4.937	2.292	1.060	0.819
	A45C	0.816	5.436	2.222	1.076	0.798	0.805	5.205	2.321	1.069	0.826
	A90	0.608	7.407	3.093	1.098	1.110	0.657	7.037	3.220	1.093	1.145
	A90A	0.822	5.351	2.266	1.072	0.814	0.835	5.175	2.370	1.065	0.835
	A90B	0.825	5.314	2.241	1.070	0.805	0.839	5.124	2.344	1.064	0.826
	A90C	0.827	5.285	2.223	1.070	0.799	0.843	5.059	2.326	1.064	0.819
FF	**Linear**	0.756	6.021	2.579	1.157	0.927	0.783	5.719	2.682	1.152	0.961
	Piecewise	0.801	5.454	2.373	1.053	0.852	0.784	5.397	2.485	1.046	0.887
	Quadratic	0.818	5.186	2.417	1.018	0.868	0.784	5.268	2.531	1.011	0.907
	Cubic	0.829	5.041	2.321	1.054	0.834	0.797	5.106	2.430	1.047	0.868
	Power	0.751	6.581	2.274	1.103	0.817	0.785	5.805	2.375	1.100	0.820

Results for the dependent variables showed similar performance. The TOTAL DV exhibited superior performance for the validation data in scale-independent statistics, with the best COR of 0.799, the best SMAPE of 1.020, and the best MASE of 0.860. However, it had higher RMSE and MAE values, where PED performed better (RMSE = 5.430, MAE = 4.262). These findings make intuitive sense, as TOTAL includes all crosswalk users, leading to higher absolute errors compared to PED and UPED, which represent subsets of all crosswalk users. Yet, TOTAL is likely to have less relative error and stronger correlation with observations, since it includes all push-button users. While the performance statistics were not vastly different across the DVs, the research team decided to retain all three DVs (PED, UPED, and TOTAL) for various applications in which different definitions of a "pedestrian" may be desired.

Among the independent variables, the unadjusted values A45 and A90 showed significantly poorer performance—had lower COR and higher RMSE, MAE, SMAPE, and MASE values—

than their respective imputed versions, leading to their exclusion from further consideration. For A45A/B/C and A90A/B/C, the results were fairly similar, but A45B and A90C slightly outperformed the others in all validation statistics (A45B: COR = 0.810, RMSE = 4.937, MAE = 2.292, SMAPE = 1.060, MASE = 0.819; A90C: COR = 0.843, RMSE = 5.059, MAE = 2.326, SMAPE = 1.064, MASE = 0.819). Hence, the research team decided to proceed with these two IVs: A45B and A90C. Notably, these were also found to be the best performing IVs in the Utah study (Singleton and Runa, 2021).

The functional forms exhibited varied performance when applied to the validation dataset. The linear model performed the worst and was eliminated from consideration. Among the remaining four FFs, there was no clear "best" or "worst" form across all statistics. Piecewise linear was average for all statistics; quadratic performed well on SMAPE but worse on MAE and MASE; cubic had the highest COR and lowest RMSE; and power had the best MAE and MASE but the worst RMSE. To make a decision, the team examined the performance of statistics on the estimation data. The power FF showed inconsistent performance, performing especially poorly on COR and RMSE but performing slightly better on MAE and MASE. Moreover, it did not allow a non-zero predicted value when the IV was zero, leading to its removal from consideration. Eventually, the research team decided to proceed with testing three FFs: piecewise linear, quadratic, and cubic. It is worth noting that both piecewise linear and quadratic FFs were utilized in the Utah study (Singleton and Runa, 2021).

5.3 RESULTS, STEP 2: MODEL SEGMENTATION

In Step 2 of the analysis, the research team tested the performance of 14 different segmentation variables. Using 10-fold cross-validation, the team examined all 252 combinations of models with and without segmentation/interaction terms, calculating and comparing validation statistics. Table 5.3 presents the average values of these validation statistics for models without and with segmentation, organized by dependent variable, independent variable, functional form, and segmentation variable. (Note that the statistics for the "validation data, without segmentation" case are the same for all of the segmentation rows, because they average over the same sets of 18 models (3 DVs × 2 IVs × 3 FFs) without segmentation variables.)

Table 5.2: Average validation statistics for segmented models in Step 2

		Validation data, without segmentation					Validation data, with segmentation				
	Type	**COR**	**RMSE**	**MAE**	**SMAPE**	**MASE**	**COR**	**RMSE**	**MAE**	**SMAPE**	**MASE**
DV	**PED**	0.821	4.689	2.229	1.052	0.824	0.815	4.796	2.227	1.060	0.820
	UPED	0.824	4.684	2.233	1.046	0.816	0.818	4.794	2.231	1.054	0.812
	TOTAL	0.842	4.764	2.327	0.965	0.781	0.837	4.883	2.333	0.973	0.781
IV	**A45B**	0.814	4.579	2.223	1.016	0.798	0.807	4.758	2.237	1.024	0.801
	A90C	0.844	4.846	2.303	1.026	0.816	0.840	4.890	2.290	1.034	0.808
FF	**Piecewise**	0.820	5.046	2.288	1.037	0.813	0.821	5.027	2.284	1.043	0.809
	Quadratic	0.829	4.625	2.299	0.992	0.823	0.829	4.654	2.283	1.006	0.815
	Cubic	0.837	4.467	2.202	1.035	0.785	0.820	4.792	2.225	1.038	0.790
Seg	**RECALL**	0.829	4.712	2.263	1.021	0.807	0.795	5.509	2.293	1.026	0.813
	CYCLE090	0.829	4.712	2.263	1.021	0.807	0.831	4.619	2.215	1.034	0.787
	CYCLE120	0.829	4.712	2.263	1.021	0.807	0.832	4.609	2.238	1.026	0.792
	PEAKAM	0.829	4.712	2.263	1.021	0.807	0.828	4.722	2.262	1.022	0.807
	PEAKPM	0.829	4.712	2.263	1.021	0.807	0.828	4.756	2.267	1.024	0.808
	PEAKAMPM	0.829	4.712	2.263	1.021	0.807	0.828	4.751	2.263	1.025	0.806
	OVERNIGHT	0.829	4.712	2.263	1.021	0.807	0.820	5.050	2.258	1.041	0.800
	WEEKEND	0.829	4.712	2.263	1.021	0.807	0.829	4.711	2.260	1.025	0.806
	ODOTREG	0.829	4.712	2.263	1.021	0.807	0.792	5.623	2.394	1.033	0.846
	PLACETYPE3	0.829	4.712	2.263	1.021	0.807	0.840	4.371	2.196	1.041	0.781
	TRANSIT4C	0.829	4.712	2.263	1.021	0.807	0.826	4.754	2.259	1.025	0.806
	TRANSIT8C	0.829	4.712	2.263	1.021	0.807	0.818	4.796	2.253	1.018	0.802
	EDTOTAL4C	0.829	4.712	2.263	1.021	0.807	0.835	4.416	2.211	1.039	0.782
	EDTOTAL8C	0.829	4.712	2.263	1.021	0.807	0.825	4.852	2.324	1.029	0.829

To be considered useful, the addition of a segmentation variable (as an interaction term) had to improve the model's performance, according to the validation statistics. In other words, the research team looked for instances where the segmentation variable increased the correlation, and/or decreased the RMSE, MAE, SMAPE, and/or MASE, compared to a non-segmented model. The research team identified four segmentation variables (CYCLE090, CYCLE120, PLACETYPE3, and EDTOTAL4C) that increased correlation and notably decreased RMSE. WEEKEND also slightly reduced RMSE. Many variables reduced MAE and MASE, with CYCLE090, CYCLE120, PLACETYPE3, and EDTOTAL4C showing the biggest improvements. Only TRANSIT8C reduced SMAPE, although the changes were minor. Based on

these findings, the research team proceeded with considering four segmentation variables: RECALL, CYCLE090, PLACETYPE3, and EDTOTAL4C. Despite RECALL's poor performance in the Oregon data segmentation tests, it was chosen for additional consideration due to its importance in the Utah study (Singleton and Runa, 2021). Similarly, CYCLE90 was selected over CYCLE120 because a similar variable was found beneficial in the Utah study.

The results for dependent variables in Step 2 closely matched those from Step 1. Model segmentation generally led to a slight worsening of COR, RMSE, and SMAPE, while resulting in slight improvements in MAE and MASE. Due to the reasons mentioned in Step 1, the research team decided to retain all three DVs: PED, UPED, and TOTAL.

Shifting focus to the two remaining independent variables, the inclusion of segmentation variables resulted in poorer performance for models with A45B across all validation statistics. In comparison, for A90C, performance varied: COR decreased and RMSE/SMAPE increased, but MAE/MASE decreased. However, the decline in performance was less pronounced for A90C than for A45B. Additionally, A90C exhibited noticeably better correlation than A45B, despite A45B showing slightly better accuracy based on other validation statistics. An important discovery upon detailed inspection was that A45B did not function effectively in certain situations. Specifically, it produced only zeros when there was no parallel motor vehicle movement generating events with code 0 (phase on), such as at a midblock crossing. Consequently, the research team decided to proceed with just a single IV: A90C.

Finally, when segmentation variables or interaction terms were utilized, performance results varied across functional forms. Cubic models displayed weaker performance for segmented models across all validation statistics. The quadratic FF improved some statistics (MAE, MASE), but did not show improvements on others (RMSE, SMAPE), and the correlation did not change significantly. The piecewise linear model, on the other hand, resulted in improvements for four out of the five validation statistics. However, among segmented models, quadratic performed best for three statistics. Further examination of the cubic models, which performed the best for the other two test statistics, indicated possible issues with overfitting and excessive complexity. Moreover, the piecewise linear models were more challenging to estimate compared to the quadratic models. As a result, the research team chose to proceed with the quadratic model as the sole FF.

5.4 RESULTS, STEP 3: DETAILED INSPECTION

After completing Step 2, the research team had three DVs (PED, UPED, TOTAL), one IV (A90C), one FF (quadratic), and four segmentation variables (RECALL, CYCLE090, PLACETYPE3, EDTOTAL4C), along with the option of no segmentation, remaining for continued consideration. In Step 3, the team further inspected the resulting models, not only considering validation statistics but also examining specific model coefficients and predictions visually. This was done to determine whether the added segmentation was justified in terms of increased complexity in modeling and application. The following results are presented for the dependent variable (DV) of TOTAL, but findings were similar for the PED and UPED models.

Table 5.4 displays the validation statistics for the considered models (DV = TOTAL, IV = A90C, FF = quadratic) without any segmentation and with each segmentation variable. These statistics

are averaged over the 10 folds used in the K-fold cross-validation process. All four segmentation models exhibited superior within-sample performance (using estimation data) for four of the five statistics (not SMAPE) compared to the models without segmentation. Among these, PLACETYPE3 performed the best. When applied to out-of-sample (validation) data, the results were not as clear, but three segmentation variables (CYCLE090, PLACETYPE3, EDTOTAL4C) exhibited higher correlation and lower error, with PLACETYPE3 again being the best performer (highest COR, lowest RMSE, MAE, and MASE). It is worth noting that the model without segmentation had the lowest SMAPE value.

Table 5.3: Validation statistics for models considered in Step 3

		Using estimation data					Using validation data				
Type		**COR**	**RMSE**	**MAE**	**SMAPE**	**MASE**	**COR**	**RMSE**	**MAE**	**SMAPE**	**MASE**
Seg	**None**	0.871	4.732	2.300	0.948	0.775	0.863	4.757	2.399	0.941	0.803
	RECALL	0.879	4.596	2.231	0.948	0.752	0.863	5.082	2.379	0.943	0.792
	CYCLE090	0.882	4.529	2.185	0.959	0.736	0.871	4.436	2.291	0.955	0.764
	PLACETYPE3	0.890	4.378	2.162	0.983	0.728	0.877	4.224	2.273	0.979	0.757
	EDTOTAL4C	0.887	4.433	2.178	0.993	0.734	0.874	4.303	2.289	0.989	0.757

However, recall from Section 4.1 that model prediction accuracy and generalizability are not the sole factors to consider. Simplicity and interpretability are also important considerations. In Figure 5.5 the predicted values (lines) are plotted over the observed values (points) for models with each of the four segmentation variables. For ease of interpreting the final results, these models include 100% of the data. These plots offer valuable insights, particularly when compared to the findings of the Utah study (Singleton and Runa, 2021).

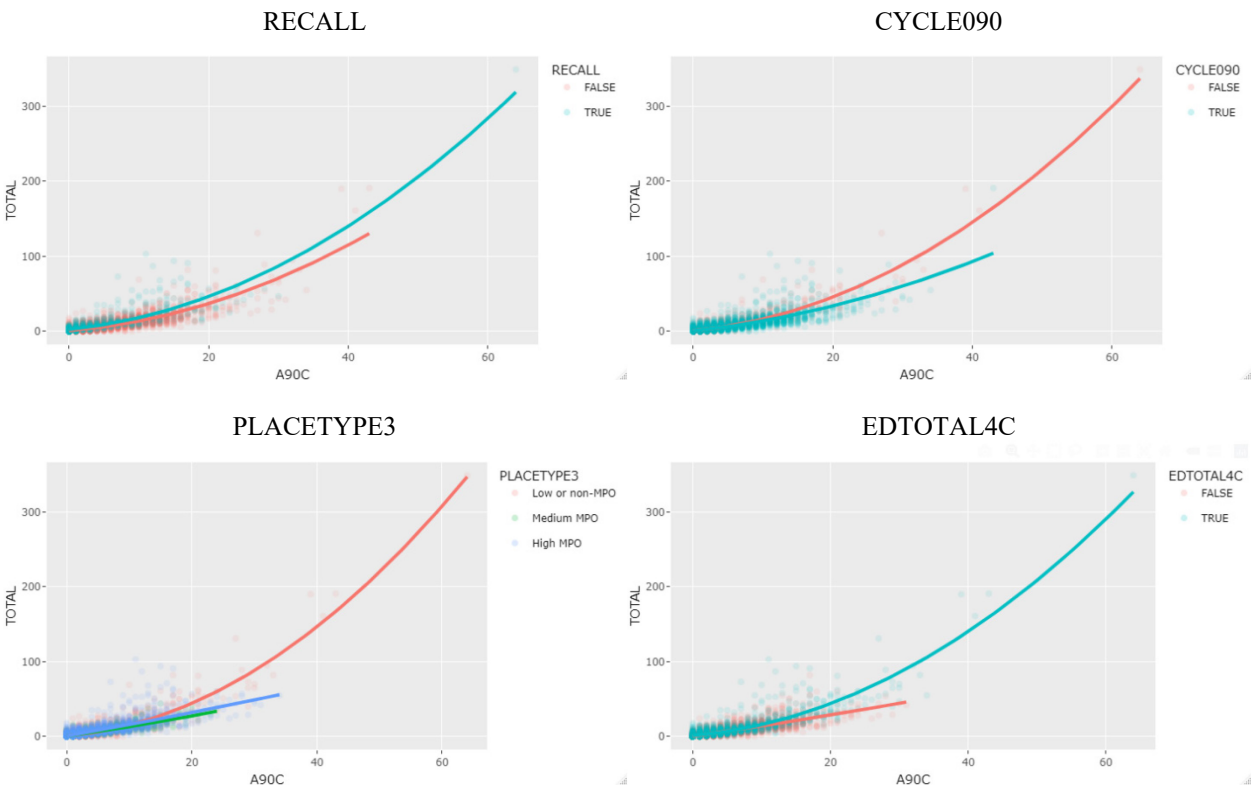


Figure 5.5: Segmented model results (hourly data) considered in Step 3

For RECALL, the two fit lines were relatively close, although crossings on pedestrian recall and rest-in-walk had slightly larger predicted values for the same IV value. This aligns with general findings from the Utah study (Singleton and Runa, 2021). Pedestrians at crossings on pedestrian recall do not need to press the push-button to receive a walk indication, so the same number of button presses (processed as A90C) represents more people. However, it should be noted that the importance of pedestrian recall status in predicting pedestrian volumes was diminished in Oregon (worse validation statistics, similar prediction lines) compared to Utah, where this was an important segmentation variable. The lower frequency of pedestrian recall in the Oregon dataset (9%) compared to Utah (22%) could explain this difference. Given the inferior statistical performance and similar prediction lines, the research team does not recommend using pedestrian recall (RECALL) as a segmentation variable in Oregon, at least at this time.

Regarding CYCLE090 (cycle length is longer than 90 seconds), the plot in Figure 5.5 conflicts with the results from the Utah study. In Oregon, it appears that locations with shorter cycle lengths had higher predicted pedestrian volumes (for the same IV value) than locations with longer cycle lengths. This is in opposition to the findings from Utah, where the model prediction line for longer cycle lengths was above (not below) the line for shorter cycle lengths. In Utah, the explanation was that, with longer cycle lengths, larger groups of pedestrians could form (waiting longer to cross), and some may not press the push-button because other people were already present. The reason for the different findings in Oregon is uncertain. Perhaps, with a longer cycle length, pedestrians may become impatient and press the push-button several times (with at least a 15-second gap, given the construction of A90C). It could also be that traffic engineers in Portland and other Oregon study locations use shorter cycle lengths in areas with high pedestrian

activity. Due to the unexpected result and no large difference in the segmented model's predictions for most of the range of the observed IV, the research team decided to not recommend using cycle length (CYCLE090) for model segmentation.

Considering PLACETYPE3, the validation statistics showed promise, but the plot in Figure 5.5 revealed some issues. First, the prediction lines for the medium MPO and large MPO categories were nearly identical, with few distinctions. Second, the fact that the low or non-MPO category had higher predicted pedestrian volumes (for a given IV value) than locations in the other two more urbanized categories was unexpected and potentially contrary to the results of the Utah study (Singleton and Runa, 2021). In Utah, higher activity signals (in more dense parts of urbanized areas) had a steeper slope than lower activity signals, explained by likely larger pedestrian group sizes and fewer push-button presses per person. However, the results for Oregon imply the opposite: in the least-dense locations, pedestrians are less likely to press the button, at least on a per-person basis. A final issue is the longevity and transferability of the PLACETYPE3 variable itself. While currently based on the well-defined place type measure used by ODOT, there is no guarantee that this metric's construction will remain the same in future years. It needs to be updated as development patterns change, and is not easily transferrable to other jurisdictions. Given these challenges, the research team recommended against using place type (PLACETYPE3) as a segmentation variable.

Concerning EDTOTAL4C, the variable showed some promise, with the second-best validation statistics in Table 5.4, and reasonably distinct lines in Figure 5.5 There are plausible explanations for why signals within 400 meters of a school would have more pedestrians for a given IV value. Due to class schedules, pedestrian activity near schools tends to be more peaked and concentrated at certain times-of-day, when larger groups are likely to form, and when a smaller proportion of pedestrians would need to press the button to cross. However, there was concern about the lack of a comparable IV range of observations for locations without nearby schools (defined up to 30, not 60+), causing questions about the appropriateness of predictions above IV values in the observed range. Moreover, there were similar concerns about the need for periodically re-calculating this variable from updated data sources and the somewhat arbitrary threshold of 400 meters. Consequently, the research team recommended against using education (EDTOTAL4C) as a segmentation variable for the pedestrian models.

In conclusion, after careful consideration, the research team decided not to use any segmentation variables in the final model, as described in detail in the following section.

5.5 FINAL RESULTS: RECOMMENDED MODEL(S)

Ultimately, the research team recommends using one model to predict pedestrian crossing volumes at signalized intersections, utilizing pedestrian activity information derived from high-resolution traffic signal controller log data. The recommended model implements a dependent variable of all crosswalk users (TOTAL: pedestrians, bicyclists, and users of scooters, skateboards, wheelchairs, and other modes); an independent variable equal to an imputed version of pedestrian detections (A90C: the number of push-button presses, calculated using a 15-second filter); the functional form as characterized by a quadratic relationship; and no segmentation variables. For situations where a more specific definition of "pedestrian" or "crosswalk user" is

desired, the research team also offers two alternative models that utilize different DVs: one with just pedestrians (PED) and one including pedestrians, wheelchair, and skateboard users (UPED).

Table 5.5 presents the coefficient estimates of the recommended models, estimated using 100% of the data. Table 5.6 displays several measures of model goodness-of-fit, including the R^2 values of the models in Table 5.5, along with various validation statistics (COR, RMSE, MAE, SMAPE, and MASE). These statistics were calculated as averages by applying the models to hold-out samples, which consist of 10% of signals, using K-fold cross-validation. Figure 5.6 provides plots of the model prediction lines over observed point data for all three recommended models; these plots appear similar due to the modest differences in the DVs (y-axes). The three final quadratic model equations are presented, for a given hour and crossing, as follows:

$$\textbf{TOTAL (all crosswalk users)} = \mathbf{1.1063 + 0.7167(A90C) + 0.0599(A90C)^2}$$

$$\textbf{UPED (pedestrians, skateboard, wheelchair)} = \mathbf{0.9953 + 0.5000(A90C) + 0.0633(A90C)^2}$$

$$\textbf{PED (only pedestrians)} = \mathbf{0.9917 + 0.4778(A90C) + 0.0636(A90C)^2}$$

Table 5.4: Recommended models: Estimation results

DV	IV	B	SE	t	p
TOTAL	Intercept	1.1063	0.0659	16.80	<0.001
	A90C	0.7167	0.0211	33.95	<0.001
	$A90C^2$	0.0599	0.0010	59.56	<0.001
UPED	Intercept	0.9953	0.0659	15.10	<0.001
	A90C	0.5000	0.0211	23.65	<0.001
	$A90C^2$	0.0633	0.0010	62.90	<0.001
PED	Intercept	0.9917	0.0659	15.05	<0.001
	A90C	0.4778	0.0211	22.63	<0.001
	$A90C^2$	0.0636	0.0010	63.20	<0.001

Table 5.5: Recommended models: Goodness-of-fit statistics

DV	N	R^2	COR	RMSE	MAE	SMAPE	MASE
TOTAL	8,546	0.7580	0.8628	4.7566	2.3990	0.9412	0.8027
UPED	8,546	0.7308	0.8339	4.7628	2.3328	1.0165	0.8517
PED	8,546	0.7278	0.8311	4.7697	2.3288	1.0229	0.8602

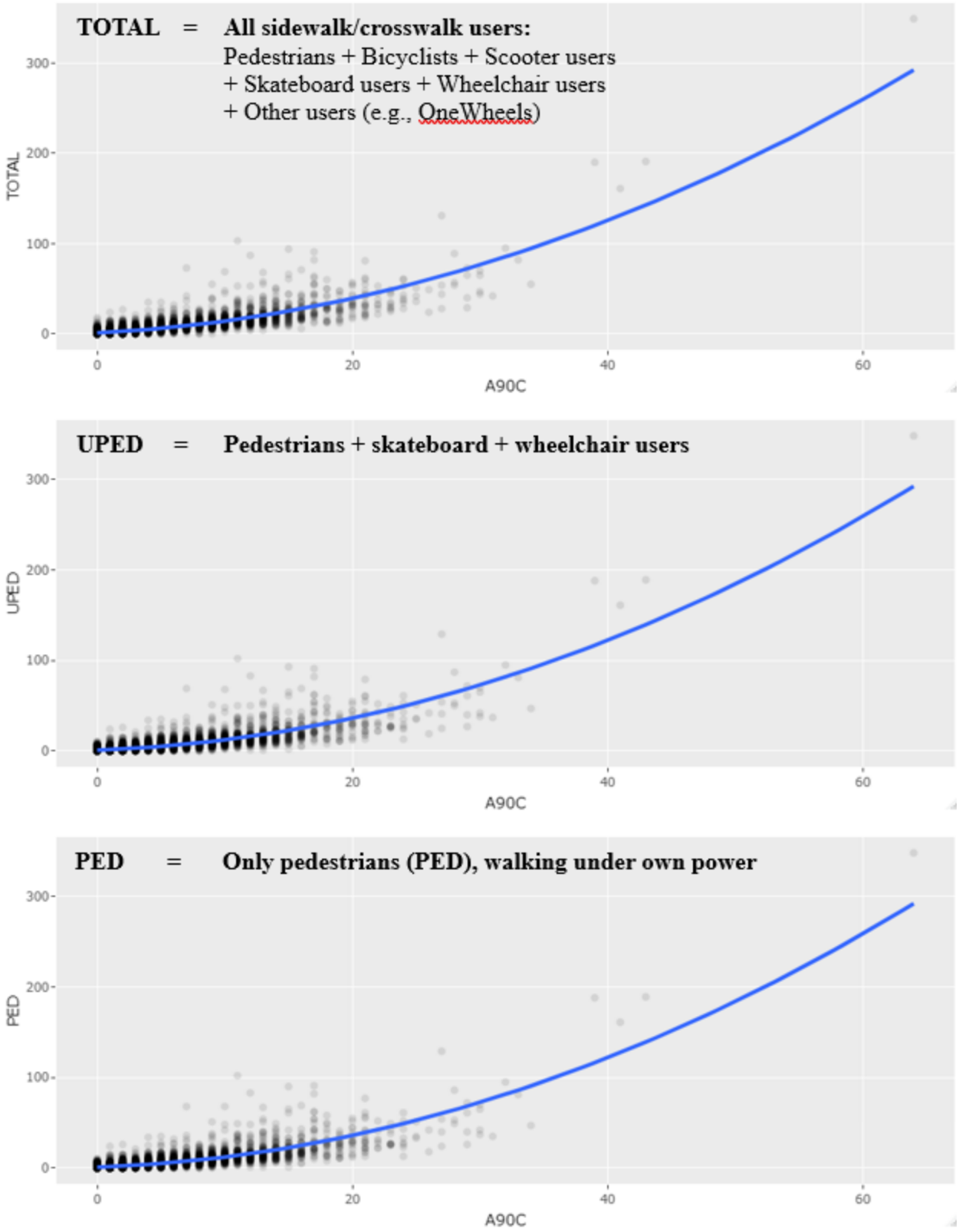


Figure 5.6: Recommended models: Plots

The recommended model reveals a noticeable non-linear (quadratic) relationship between pedestrian signal activity (A90C) and total crosswalk volumes (TOTAL). The estimated intercept term (1.11) means that the model predicts approximately one pedestrian per hour per crossing, even in the absence of any time-filtered push-button presses. This accounts for situations in which all pedestrians in that hour do not activate the push-button. The estimated linear coefficient (0.72) is significantly larger than the estimated quadratic term (0.06), but as the values of the independent variable increase, the squared term contributes a larger portion of the predicted value of the dependent variable. This enables the model to better predict the observations with the highest volumes.

Overall, the recommended model (TOTAL) demonstrates remarkable accuracy and generalizability, considering its simplicity: it utilizes just one independent variable (A90C) and estimates only three coefficients in a quadratic model. The R^2 value was 0.76, and the predictions to 10% hold-out samples showed a strong correlation of 0.86 with the observed values. Furthermore, the accuracy statistics indicate that the model has an average (absolute) error (MAE) of ±2.4 pedestrians per hour. The SMAPE value (0.94) suggests errors are 6% smaller than the average value of the dependent variable, and the MASE value (0.80) indicates that the model's predictions have 20% less error than the deviations in the original data. Overall, while not perfect, the recommended model performs admirably in predicting pedestrian crosswalk volumes in a variety of situations.

5.5.1 Additional Testing: Errors

To test the recommended model, the research team performed a first set of additional evaluations to assess its performance (in terms of prediction error) in different situations.

First, the model's performance was examined across the range of predicted values (TOTAL), with no aggregation (i.e., for each hour and phase and each signal). For this test, the model was estimated using the entire (100%) dataset and then applied to the same (100%) estimation dataset. (Results were fairly similar when using some cross-validation or hold-out sampling method.) Observations were then split into five groups based on TOTAL values, at 50%, 75%, 90%, and 99%; bounds were 1.10, 1.88, 4.93, 12.41, 39.40, and 292.35. Accuracy statistics were calculated for observations within each group; results are shown in the top half of Table 5.7 (full results are available upon request). An approximate 90% prediction interval (PI) was also calculated, as ± the standard deviation of the error multiplied by 1.645 (the z-score for a two-tailed 90% interval). This PI indicates the bounds around the predicted values within which one would expect the true values to fall, roughly 90% of the time.

These results showed that both correlation and relative accuracy (SMAPE, MASE) tended to improve for the highest values; although these quantiles also had the worst values of absolute error (RMSE, MAE). For the lowest 75% of the data, the absolute error (RMSE, MAE) was lower than average. Even in the highest 1% of observations, the model's percentage error was reasonable (SMAPE = 0.35), better than lowest 75% of the data (SMAPE = 0.50 and 1.36). For almost 95% of the observations (94.65%), the observed pedestrian volumes fell within ±7.8 pedestrians per hour of the predicted pedestrian volumes. Overall, the research team found that the model performed reasonably well across a range of predicted values.

Table 5.7: Validation statistics for final model, by quantiles of predicted values

Aggregation	Quantile	N	Mean	COR	RMSE	MAE	SMAPE	MASE	90%PI
None	None	8546	4.90	0.8706	4.7477	2.3028	0.9456	0.7760	±7.81
(hour, phase)	0-50%	4757	1.32	0.3745	1.6698	1.1855	1.3636	1.2220	±2.68
	50-75%	1810	3.63	0.3347	3.0707	1.9859	0.4981	0.8701	±4.95
	75-90%	1252	8.58	0.3739	5.2379	3.2823	0.3506	0.8366	±8.49
	90-99%	656	21.18	0.4208	11.5971	7.6523	0.3492	0.9955	±19.09
	99-100%	71	61.68	0.9079	23.6546	18.5364	0.3469	0.7449	±37.58
Day, Signal	None	121	345.72	0.9293	127.3874	90.7733	0.3921	0.4472	±210.40
	0-25%	31	85.78	0.7175	42.5462	34.6840	0.6717	1.0553	±52.00
	25-50%	30	200.27	0.4262	121.9894	81.3669	0.3752	0.7137	±204.02
	50-75%	30	329.72	0.3578	137.0798	106.6107	0.3185	0.7318	±229.29
	75-100%	30	775.77	0.8654	172.9401	142.3012	0.1937	0.4617	±286.46

Second, the model's performance was assessed when pedestrian volumes (observed and predicted) were summed to a daily total for each signal (i.e., aggregated across all hours in a day, and all phases at a signal). The totals are for consecutive 24 hours periods from the start of the count. This time, observations (121 signal-days) were then split into four groups based on total daily predicted pedestrian volumes, this time into quartiles: bounds were 25.75, 155.92, 246.42, 444.30, and 1309.54. Once again, accuracy statistics were calculated for observations within each group; results are shown in the bottom half of Table 5.7 (full results are available upon request). The calculation of prediction intervals for sums of model results is quite complicated, considering both the variance and covariance of the errors involved in the sums. Therefore, approximate 90% prediction intervals were again calculated as ± 1.645 times the standard deviation of the error in each group.

A similar pattern emerges when investigating results aggregated to daily intersection-level pedestrian volumes. Relative accuracy (SMAPE, MASE) is better and absolute error (RMSE, MAE) is worse at higher-volume signals. The correlation was best for the highest-volume quartile (0.87), followed by the lowest-volume quartile (0.72). The observed daily pedestrian volume was with ±210 pedestrians per day of the predicted volume for 91% of the 121 signal-day observations. Overall, the biggest differences were between the first quartile (the lowest quarter of location–days) versus all instances of higher pedestrian volumes. For instance, 74% of the 31 observations in the first quartile had their observed total pedestrian volumes within ±210 of predicted values. In summary, the model still performs reasonably well across a range of pedestrian activity, even when aggregated to daily volumes. In fact, the relative measures of accuracy (SMAPE, MASE) tend to be better (lower) for daily than for hourly predictions.

Figure 5.7 plots the predicted daily pedestrian volumes against both the observed pedestrian volumes (upper) and the prediction error (lower). Overall, the predictions closely match the observed pedestrian volumes (correlation 0.93), when model predictions are aggregated to daily intersection totals.

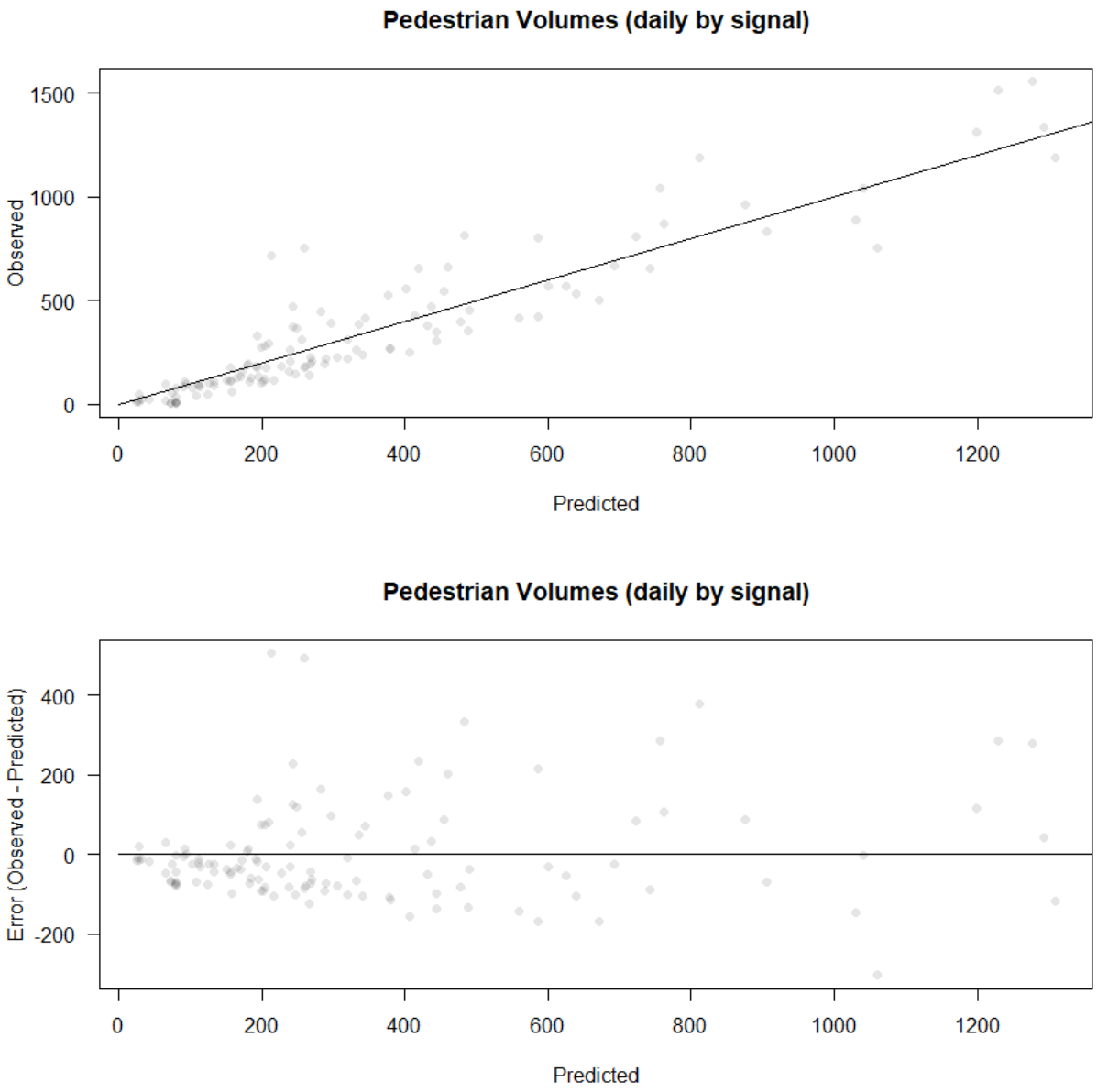


Figure 5.7: Predicted vs. observed (and error) pedestrian volumes (daily by signal)

A final note about the prediction intervals shown in Table 5.7: It has already been mentioned that the calculated PIs are approximate. More critically, their calculation assumes that errors are normally distributed with homogenous variance within groups. As Figure 5.1 shows, this assumption is likely not valid. For instance, a prediction of 2.00±2.68 pedestrians per hour includes negative volumes within the PI, which is impossible. Thus, the PIs should be considered with caution in light of this important limitation.

5.5.2 Additional Testing: Oregon versus Utah

A second set of tests involved comparing the Oregon results and the Utah results. Two comparisons were conducted: one comparing the accuracy statistics from this Oregon study (using UPED as the DV, for comparability) to those reported in the Utah study (Singleton and Runa, 2021), and the second replicating the Oregon model's specification but estimating it on the Utah data.

For the first comparison, the results showed that this study's simpler Oregon model was a better and more accurate fit to the Oregon data ($R^2 = 0.731$, COR = 0.834, RMSE = 4.763, MAE = 2.333) than the more complex (and segmented) suite of Utah models was to the Utah data ($R^2 = 0.724$, COR = 0.839, RMSE = 12.247, MAE = 2.961). The superior within-sample performance of the Oregon model may be attributed to several factors. One reason may be the difference in dataset size. The Utah dataset was notably larger (N = 22,630) than the Oregon dataset (8,546), allowing for data collection at a greater variety of locations. As a result, the model may not have performed as well in all locations in the Utah dataset due to the broader range of situations covered. A second and more likely explanation for the disparity in model performance is the difference in signal operational strategies between the two states. In Utah, a larger portion of the data was observed at signals with crossings on pedestrian recall for part of the time (22%), and very few signals throughout the state operate without pushbuttons (mostly in downtown Salt Lake City). In contrast, in Oregon, pedestrian recall made up a much smaller share of the dataset (9%), and many locations with very high pedestrian volumes (e.g., downtown Portland) do not have pushbuttons and were not included in this study. As a result, the Utah study could include locations with higher pedestrian volumes—which the previous paragraph noted to have higher error—while the Oregon study did not have access to such locations to the same extent. Thus, the Oregon model was a better fit to (and more accurate predictor of) its observed data.

A second comparison between Oregon and Utah was conducted by replicating the Oregon model specification and estimating it on the Utah data (Singleton, Runa, & Humagain, 2021). This involved using a quadratic model with TOTAL as the DV, A90C as the IV, and no segmentation variable. The estimated coefficients, prediction line, and goodness-of-fit of the two models were then compared. Additionally, accuracy statistics were calculated when each state's model was applied to the other state's dataset, allowing an assessment of out-of-sample accuracy (transferability) when applying Oregon's model to Utah's data and vice-versa. Full model results are available upon request, but the new Utah models' estimated coefficients had a smaller intercept (0.383), a larger linear term (1.138) and a similar quadratic term (0.056), compared to the Oregon model. The goodness-of-fit was poorer for the Utah model ($R^2 = 0.591$) than the Oregon model, as supported by explanations in the previous paragraph.

Figure 5.7 displays the observations (points) and model predictions (lines) for Oregon (teal) superimposed on top of Utah (red). The consistency of the two quadratic fit lines across the entire range of data is remarkable, considering that the data were collected in two different states and three years apart. The figure also allays concerns about the Oregon data being sparse in the region of higher signal activity (above A90C = 20), as the "outliers" in the Oregon data fall reasonably within the range of observations from Utah. This gives confidence that if more data from Oregon were collected in this IV range, they might still fit well with the estimated model.

Table 5.8 shows the results of validation statistics for models when applied to their own and the other state's data. While the results illustrate the superior performance of the Oregon model on its own dataset compared to the Utah model on its own dataset, the more relevant findings are when applying models estimated in one state to data from the other state. Utah's model predicts data for Oregon slightly worse than Oregon's model, but the difference is not substantial. When the Oregon model is applied to Utah data, it slightly degrades COR and RMSE, but it results in improved (lower) MAE, SMAPE, and MASE.

Overall, these results offer an important and positive finding about model transferability: simple models like these (one IV = A90C, quadratic FF) may be applied in locations (even other states) where they were not developed and still yield approximately the same degree of accuracy as a similar locally developed model. In other words, pedestrian signal models may be transferrable across states with no significant loss in accuracy. Although promising, further research is needed to see if this finding holds in contexts outside of Western US states, especially in the Eastern and Southern US.

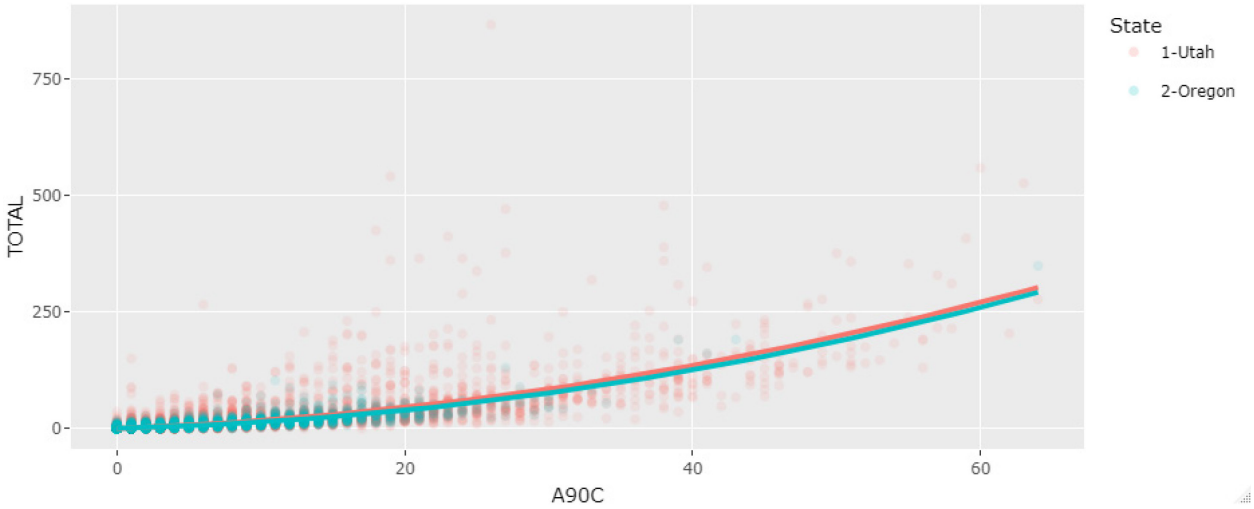


Figure 5.8: Plots of models estimated on Oregon and Utah data

Table 5.8: Validation statistics for models estimated on Oregon and Utah data

Data	Model	COR	RMSE	MAE	SMAPE	MASE
Oregon	Oregon	0.8706	4.7477	2.3028	0.9456	0.7760
	Utah	0.8680	5.0195	2.3746	1.0321	0.8002
Utah	Utah	0.7685	14.8584	3.7847	1.1557	0.9500
	Oregon	0.7673	14.9890	3.7442	1.0916	0.9398

5.6 SUMMARY

This chapter presented the results of the multi-step data analysis process used in this study, which included descriptive statistics, and the inspection and selection of dependent variables, independent variables, functional forms, and segmentation variables. Based on the findings, the research team recommends using a relatively simple model (or set of models), with TOTAL (or

UPED or PED) as the DV, A90C as the IV, a quadratic FF, and no segmentation variables. This model performs well in terms of validation statistics in K-fold cross-validation and also when applied to data from Utah. Notably, the simpler Oregon model fits the data better than the more complex Utah models. The next chapter will discuss key findings and recommendations resulting from this study.

6.0 CONCLUSIONS AND RECOMMENDATIONS

This research explored the feasibility of deriving pedestrian volumes from existing on-street infrastructure (like pedestrian pushbuttons) at a large scale, by developing adjustment factors to convert pedestrian data obtained from such infrastructure to pedestrian volumes. Additionally, this study examined the transferability of the developed methods.

To accomplish these goals, the study selected 55 signalized intersections in Oregon, representing locations with diverse pedestrian volumes and land uses. Videos were recorded at these sites; however, pedestrian push-button data was unavailable at six locations due to data loss and retrieval issues. Consequently, the analysis focused on the data from 49 sites. The videos from these sites were coded to extract the counts of pedestrians, bicyclists, scooter users, skateboarders, wheelchair users, and other crosswalk users. Additionally, pedestrian volumes and actuation data from 16 signals in eastern Oregon were obtained. In total, 41,843 users were observed, including 35,767 pedestrians, 5,032 bicyclists, 449 scooter users, 278 skateboarders, 233 wheelchair users, and 84 other crosswalk users across the 65 sites.

Next, the study proceeded with model estimation in three steps. In Step 1, 192 different models were constructed, representing combinations of three dependent variables (PED, UPED, and TOTAL), eight independent variables (A45, A45A, A45B, A45C, A90, A90A, A90B, A90C), and eight functional forms (linear, piecewise linear, quadratic, cubic, exponential, power, Poisson, negative binomial), as defined previously. These models underwent 10-fold cross-validation, and validation statistics (correlation, RMSE, MAE, SMAPE, MASE) were calculated to identify potential model combinations for further testing. Models passing Step 1 included combinations of all three DVs (PED, UPED, and TOTAL), two IVs (45B and A90C), and three FFs (piecewise linear, quadratic, and cubic).

In Step 2, the performance of 14 different segmentation variables was tested. The models were subjected to 10-fold cross-validation, and validation statistics were calculated and compared for models with and without the segmentation/interaction terms. Based on the validation statistics at this step, the research team selected four segmentation variables for further consideration: RECALL, CYCLE090, PLACETYPE3, and EDUCTOT4C. The research team continued to include all three DVs (PED, UPED, and TOTAL) but only one IV (A90C) and a sole FF (quadratic).

In Step 3, the research team conducted a more detailed examination of the models with and without segmentation. However, the validation statistics from the models with segmentation did not show a significant improvement; in fact, they were either similar or worse in performance compared to the models without segmentation. Moreover, some of these results did not align with the research team's intuitive expectations. As a result, the decision was made to abandon model segmentation at this point, and the team proceeded with the simpler models without any segmentation variables.

Ultimately, the research team recommends using a single model for predicting pedestrian crossing volumes at signalized intersections. This model utilizes pedestrian activity data derived from high-resolution traffic signal controller logs. The dependent variable is defined as all

crosswalk users (TOTAL), including pedestrians, bicyclists, scooter users, skateboarders, wheelchair users, and others. The independent variable is an imputed version of pedestrian detections (A90C), calculated using a 15-second filter. The functional form of the model is represented by a quadratic relationship, and no segmentation variables are included. For situations where a different definition of "pedestrian" or "crosswalk user" is desired, the research team also provides two alternative models: one with only pedestrians (PED) and another including pedestrians, wheelchair users, and skateboarders (UPED).

In summary, the recommended model (TOTAL) demonstrated remarkable accuracy and generalizability, considering its simplicity. With just one IV (A90C) and three estimated coefficients in a quadratic model, it achieved an R^2 value of 0.76. The model's predictions were strongly correlated (0.86) with observations in 10% hold-out samples. The accuracy statistics indicate that the model's average absolute error (MAE) was ±2.4 pedestrians per hour.

The research team conducted additional testing to assess the model's performance in various situations. First, they examined the model's performance across the range of the IV (A90C) by comparing observed values to predictions from a model estimated using the entire (100%) dataset. Observations were split into five groups based on A90C values, at 50%, 75%, 90%, and 99%; bounds were 0, 1, 4, 9, 20, and 64. The results showed that the third quartile (50–75%) or second group (A90C = 2–4) had the best accuracy, with the lowest values of most validation statistics. For 95% of the data, the errors were below average.

Next, the research team compared the Oregon data and recommended model to the Utah data in two ways. First, the accuracy statistics from the Oregon model were compared to those reported in the Utah study (Singleton and Runa, 2021). The simpler model recommended in this study showed better accuracy when fit to the Oregon data (R^2 = 0.731, COR = 0.834, MAE = 2.333) than the more complex (and segmented) suite of Utah models when fit to the Utah data (R^2 = 0.724, COR = 0.839, MAE = 2.961). Secondly, the Oregon model was replicated and estimated on the Utah data (Singleton, Runa, & Humagain, 2021). The estimated coefficients of the new Utah model had a smaller intercept, a larger linear term, and a similar quadratic term, as compared to the Oregon model. The goodness-of-fit for the Utah model was poorer (R^2 = 0.591), possibly due to differences between the Oregon and Utah datasets. However, the remarkable consistency of the two state's quadratic fit lines across the entire data range—collected in two different states and three years apart—gives some expectation of transferability for this model. When applying the Oregon model to Utah data, it slightly reduced the COR and RMSE, but resulted in improved (lower) MAE, SMAPE, and MASE. Overall, these promising results indicate that the model may be transferable to locations without similar data available to estimate a local model.

6.1 LIMITATIONS

Pedestrian push-button actuation data holds great potential for estimating pedestrian volumes at signalized intersections. As demonstrated in this study, scaling factors and models can be developed to convert actuations into reliable pedestrian volume estimates. However, it is important to address potential challenges with push-button data accuracy, as data loss or corruption arising from equipment malfunction or communication issues may occur. Construction, utility work, collisions, and adverse weather can all affect the ability of data to be

transmitted between pedestrian detectors and controllers, and how controllers send event logs to be archived in a central system. Data quality may also be affected due to erratic issues with how the push-button actuation data is being recorded and/or stored, which may be hard to detect and distinguish from both low-volume time periods or high-activity special events.

Furthermore, it is worth noting that push-button actuation data are only available at intersections equipped with pushbuttons, which are often found outside downtown areas. In contrast, signals in downtown areas often operate on recall, providing the WALK display indication each cycle. This difference in signal operation limits the availability of push-button data in downtown areas and in other locations with high pedestrian volumes.

Although the model has shown promise in its ability to work well at various locations, the higher error observed when estimating high pedestrian volumes suggests that the model could benefit from more data collected at such locations to improve its accuracy.

The models were developed based on correlations between pedestrian push-button actuations and 48-hour pedestrian counts collected during the summer months. As a result, seasonal differences in actuation and pedestrian volumes have not been explored yet.

6.2 RECOMMENDATIONS

ODOT's mission is to provide a safe and reliable multimodal transportation system that connects people and helps Oregon's communities and economy thrive. Currently, there is no systemic record of pedestrian traffic data throughout the transportation network. Having measures of walking and pedestrian travel activity is crucial for comprehensive crash analysis and project prioritization. Furthermore, ODOT aims to establish a new key performance measure for reporting to the Oregon Transportation Commission (OTC) regarding pedestrian travel, which will necessitate pedestrian traffic counts.

Pedestrian push-button actuations have proven to be effective in estimating pedestrian volumes at locations where counts are not available. To regularly estimate pedestrian volumes, there is a need to routinely archive these actuation data as part of an automated traffic signal performance measurement (ATSPM) system, and to continuously validate the data. In case the raw data becomes too large to store, careful aggregation into larger time bins (e.g., 15-minute periods) can help preserve essential information while mitigating data storage constraints. The research team has developed a memo outlining the critical variables necessary for pedestrian volume estimation using actuation data, which will need to be aggregated to maintain the data. Storing this data also enables agencies to generate additional performance measures that can be valuable for operations and traffic management.

By employing the recommended model(s) for pedestrian volume estimation, agencies can gain access to a valuable longitudinal dataset of pedestrian data that can serve as a foundation for various analyses, such as traffic safety assessments, pedestrian monitoring, equity evaluations, and health-related studies. This research presented an opportunity to validate the methods and modeling framework utilized in the Utah study (Singleton, Runa, and Humagain, 2020).

To maintain the accuracy of the estimates provided by the model, it is advisable to perform regular validation at a few locations. The validation process will help ensure that the model's estimates remain reliable and accurate over time.

6.3 FUTURE WORK

This study has laid the groundwork for several promising avenues of future research. Firstly, the successful development of models to estimate pedestrian volumes in both Utah and Oregon provides a foundation for further investigation. The initial tests of transferability indicated that the models estimated in Oregon can be effectively transferred to Utah and vice versa. However, during this analysis, it became evident that underlying differences in the data used for model estimation may introduce errors. Hence, further research is highly recommended to identify the sources of these errors and to devise effective methods to mitigate them.

Furthermore, replicating studies like these in other regions of the US, particularly the East, South, and Midwest, would be valuable. This approach can shed light on potential regional variations in the relationships between pedestrian push-button actuations and actual pedestrian volumes, allowing for a more comprehensive understanding of pedestrian behavior in different contexts. By exploring diverse geographical areas, researchers can gain insights into the generalizability and adaptability of the developed models beyond the specific study locations.

All the models estimated in this study were based on data primarily collected during the summer months. For comprehensive insights, future research should investigate the effects of data collected in other months to understand the impact of seasonality on pedestrian volume estimation. Additionally, while this study explored several factors (such as segmentation variables and measures of traffic signal activity) that can influence the prediction of pedestrian volumes using traffic signal data, there are still other potential factors to be considered. Characteristics of land uses, the built environment, roadway design, and weather could play significant roles in different locations. Moreover, alternative methods of processing traffic signal controller log events may lead to more accurate predictions than the versions used in this study (e.g., A90C). Thus, future research could delve into testing a broader range of independent and segmentation variables to enhance modeling accuracy and applicability to diverse contexts.

When implementing the models recommended in this study on pedestrian traffic signal data spanning multiple months or years, it is important to anticipate the presence of missing data (e.g., construction, loss of connectivity to the signal) and invalid records (e.g., a stuck-on push-button). To handle these instances, it will be crucial to employ outlier detection techniques and missing data imputation methods. In future research, a range of statistical and machine learning methods should be explored to effectively identify and address these issues in pedestrian signal data. By adopting such approaches, more robust and accurate pedestrian volume estimations can be achieved, enhancing the overall reliability of the models when dealing with real-world data scenarios.

The models developed in this study represent empirical relationships between pedestrian traffic signal data and pedestrian crossing volumes, capturing the average behaviors of pedestrians and other sidewalk users during street crossings. These behaviors encompass group formation, push-button press patterns, crossing behaviors, and platooning. As previously mentioned, these

behaviors might vary across different cities, states, regions, etc., highlighting the importance of future research in investigating the transferability of these pedestrian volume models to other jurisdictions.

Moreover, as the transportation system evolves over time, these behaviors and relationships may change due to factors such as increased usage of micromobility modes on sidewalks, advancements in pedestrian detection methods, or modifications in intersection control strategies to accommodate connected and autonomous vehicles. Hence, it is imperative for future research to explore how frequently these models should be re-estimated and under which circumstances. By addressing these aspects, the pedestrian volume models can remain relevant and accurate in dynamic and evolving transportation environments, ensuring their continued utility for pedestrian traffic estimation and planning purposes.

7.0 REFERENCES

Abedi, N., Bhaskar, A., Chung, E., & Miska, M. (2015). Assessment of antenna characteristic effects on pedestrian and cyclists travel-time estimation based on Bluetooth and WiFi MAC addresses. Transportation Research Part C: Emerging Technologies, 60, 124–141. doi: 10.1016/j.trc.2015.08.010

Akamatsu, S., Shimaji, N., & Tomizawa, T. (2014). Development of a person counting system using a 3D laser scanner. 2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014), 1983–1988. Bali, Indonesia: IEEE. doi: 10.1109/ROBIO.2014.7090627

Atkins North America. (2020). Automated Traffic Signal Performance Measures Reporting Details. Atlanta, GA: Georgia Department of Transportation.

Bai, L., Ireson, N., Mazumdar, S., & Ciravegna, F. (2017). Lessons learned using wi-fi and Bluetooth as means to monitor public service usage. Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers, 432–440. Maui Hawaii: ACM. doi: 10.1145/3123024.3124417

Balsys, K., Valinevičius, A., & Eidukas, D. (2009). Urban Traffic Control using IR video detection technology. Elektronika Ir Elektrotechnika, 96(8), 81–84. Retrieved from https://eejournal.ktu.lt/index.php/elt/article/view/9958

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(1), 69–79. doi: 10.3141/2527-08

Brandon Nevers, Tom Urbanik, Kevin Lee, Burak Cesme, Jennifer Musselman, Laura Zhao, Kittelson & Associates, Inc., University, D. B. H. L. P., Tanaka, C. O. P. A., Day, I. S. U. C., Kimley-Horn & Associates Lucy Richardson, National Cooperative Highway Research Program, … National Academies of Sciences, Engineering, and Medicine. (2020). Performance-Based Management of Traffic Signals (p. 25875). Washington, D.C.: Transportation Research Board. doi: 10.17226/25875

Browne, M. W. (2000). Cross-Validation Methods. Journal of Mathematical Psychology, 44(1), 108–132. doi: 10.1006/jmps.1999.1279

Bullock, D. M., Haseman, R., Wasson, J. S., & Spitler, R. (2010). Automated Measurement of Wait Times at Airport Security: Deployment at Indianapolis International Airport, Indiana. Transportation Research Record, 2177(1), 60–68. doi: 10.3141/2177-08

Clifton, K. J., Singleton, P. A., Muhs, C. D., & Schneider, R. J. (2016a). Development of destination choice models for pedestrian travel. Transportation Research Part A: Policy and Practice, 94, 255–265. doi: 10.1016/j.tra.2016.09.017

Clifton, K. J., Singleton, P. A., Muhs, C. D., & Schneider, R. J. (2016b). Representing pedestrian activity in travel demand models: Framework and application. Journal of Transport Geography, 52, 111–122. doi: 10.1016/j.jtrangeo.2016.03.009

Dai, C., Zheng, Y., & Li, X. (2007). Pedestrian detection and tracking in infrared imagery using shape and appearance. Computer Vision and Image Understanding, 106(2–3), 288–299. doi: 10.1016/j.cviu.2006.08.009

Danalet, A., Farooq, B., & Bierlaire, M. (2014). A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures. Transportation Research Part C: Emerging Technologies, 44, 146–170. doi: 10.1016/j.trc.2014.03.015

Day, C., Bullock, D., Li, H., Remias, S., Hainen, A., Freije, R., … Brennan, T. (2014). Performance Measures for Traffic Signal Systems: An Outcome-Oriented Approach. Purdue University. doi: 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).

Day, C., Premachandra, H., & Bullock, D. (2016). Rate of Pedestrian Signal Phase Actuation as a Proxy Measurement of Pedestrian Demand. Transportation Research Record. Retrieved from https://docs.lib.purdue.edu/civeng/24

De Silva, V., Roche, J., & Kondoz, A. (2018, January 1). Fusion of LiDAR and camera sensor data for environment sensing in driverless vehicles. Loughborough University. Retrieved from https://hdl.handle.net/2134/33170

Diogenes, M. C., Greene-Roesel, R., Arnold, L. S., & Ragland, D. R. (2007). Pedestrian Counting Methods at Intersections: A Comparative Study. Transportation Research Record, 2002(1), 26–30. doi: 10.3141/2002-04

Du, Y., Yue, J., Ji, Y., & Sun, L. (2017). Exploration of optimal Wi-Fi probes layout and estimation model of real-time pedestrian volume detection. International Journal of Distributed Sensor Networks, 13(11), 155014771774185. doi: 10.1177/1550147717741857

Ewing, R., & Cervero, R. (2010). Travel and the Built Environment: A Meta-Analysis. Journal of the American Planning Association, 76(3), 265–294. doi: 10.1080/01944361003766766

Federal Highway Administration. (2016). Traffic Monitoring Guide (Tech Report No. FHWA-PL-17-003; p. 473). US Department of Transportation. Retrieved from US Department of Transportation website: https://www.fhwa.dot.gov/policyinformation/tmguide/

Foster, N., Monsere, C. M., & Carlos, K. (2014). Evaluating Driver and Pedestrian Behaviors at Enhanced, Multilane, Midblock Pedestrian Crossings: Case Study in Portland, Oregon. Transportation Research Record: Journal of the Transportation Research Board, 2464(1), 59–66. doi: 10.3141/2464-08

Fu, T., Stipancic, J., Zangenehpour, S., Miranda-Moreno, L., & Saunier, N. (2017). Automatic Traffic Data Collection under Varying Lighting and Temperature Conditions in Multimodal Environments: Thermal versus Visible Spectrum Video-Based Systems. Journal of Advanced Transportation, 2017, 1–15. doi: 10.1155/2017/5142732

Garcia, J., Gardel, A., Bravo, I., Lazaro, J. L., Martinez, M., & Rodriguez, D. (2013). Directional People Counter Based on Head Tracking. IEEE Transactions on Industrial Electronics, 60(9), 3991–4000. doi: 10.1109/TIE.2012.2206330

Gomez, A., Conti, F., & Benini, L. (2018). Thermal image-based CNN's for ultra-low power people recognition. Proceedings of the 15th ACM International Conference on Computing Frontiers, 326–331. Ischia Italy: ACM. doi: 10.1145/3203217.3204465

Greene-Roesel, R., Diogenes, M. C., Ragland, D. R., & Lindau, L. A. (2008, April). 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. Washington, DC. Retrieved from https://escholarship.org/uc/item/2n83w1q8

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. Transportation Research Record: Journal of the Transportation Research Board, 2673(4), 941–950. doi: 10.1177/0361198119837235

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. doi: 10.1177/0361198118793006

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. Journal of Transport & Health, 7, 149–159. doi: 10.1016/j.jth.2017.08.009

Humagain, P., & Singleton, P. (2021). Advances in pedestrian travel monitoring: Temporal patterns and spatial characteristics using pedestrian push-button data from Utah traffic signals. Journal of Transport and Land Use, 14(1). doi: 10.5198/jtlu.2021.2112

Jahangiri, A., Hasani, M., Sener, I. N., Munira, S., Owens, J., Appleyard, B. S., … Virginia Tech Transportation Institute. (2019). Use of Direct-Demand Modeling in Estimating Nonmotorized Activity: A Meta-analysis (No. 01–003). Safety through Disruption University Transportation Center. Retrieved from Safety through Disruption University Transportation Center website: https://safed.vtti.vt.edu/wp-content/uploads/2020/07/UTC-Safe-D_Direct-Demand-Model-for-PedBike_TTI-Report_12Oct17_Final.pdf

Jones, M. G., Ryan, S., Donlon, J., Ledbetter, L., Ragland, D. R., & Arnold, L. 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 Berkeley Traffic Safety Center, 120. Retrieved from https://trid.trb.org/View/919880

Jung, J., Olsen, M. J., Hurwitz, D. S., Kashani, A. G., & Buker, K. (2018). 3D virtual intersection sight distance analysis using lidar data. Transportation Research Part C: Emerging Technologies, 86, 563–579. doi: 10.1016/j.trc.2017.12.004

Kim, S., Park, S., & Jang, K. (2019). Spatially-varying effects of built environment determinants on walking. Transportation Research Part A: Policy and Practice, 123, 188–199. doi: 10.1016/j.tra.2019.02.003

Kocamaz, M. K., Gong, J., & Pires, B. R. (2016). Vision-based counting of pedestrians and cyclists. 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), 1–8. Lake Placid, NY, USA: IEEE. doi: 10.1109/WACV.2016.7477685

Kothuri, S. (2014). Exploring Pedestrian Responsive Traffic Signal Timing Strategies in Urban Areas (Doctoral Dissertation, Portland State University). Portland State University. doi: 10.15760/etd.1933

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(1), 11–18. doi: 10.3141/2644-02

Krotosky, S. J., & Trivedi, M. M. (2007). On Color-, Infrared-, and Multimodal-Stereo Approaches to Pedestrian Detection. IEEE Transactions on Intelligent Transportation Systems, 8(4), 619–629. doi: 10.1109/TITS.2007.908722

Kurkcu, A., & Ozbay, K. (2017). Estimating Pedestrian Densities, Wait Times, and Flows with Wi-Fi and Bluetooth Sensors. Transportation Research Record: Journal of the Transportation Research Board, 2644(1), 72–82. doi: 10.3141/2644-09

Kutela, B., & Teng, H. (2020). Evaluating the influential factors for pushbutton utilization at signalized midblock crosswalks. Safety Science, 122, 104533. doi: 10.1016/j.ssci.2019.104533

Kuzmyak, J. R., Walters, J., Bradley, M., Kockelman, K. M., National Cooperative Highway Research Program, Transportation Research Board, & National Academies of Sciences, Engineering, and Medicine. (2014). Estimating Bicycling and Walking for Planning and Project Development: A Guidebook (p. 22330). Washington, D.C.: Transportation Research Board. doi: 10.17226/22330

Lee, K., & Sener, I. N. (2020). Emerging data for pedestrian and bicycle monitoring: Sources and applications. Transportation Research Interdisciplinary Perspectives, 4, 100095. doi: 10.1016/j.trip.2020.100095

Lefebvre-Ropars, G., Morency, C., Singleton, P. A., & Clifton, K. J. (2017). Spatial transferability assessment of a composite walkability index: The Pedestrian Index of the

Environment (PIE). Transportation Research Part D: Transport and Environment, 57, 378–391. doi: 10.1016/j.trd.2017.08.018

Lesani, A., & Miranda-Moreno, L. (2019). Development and Testing of a Real-Time WiFi-Bluetooth System for Pedestrian Network Monitoring, Classification, and Data Extrapolation. IEEE Transactions on Intelligent Transportation Systems, 20(4), 1484–1496. doi: 10.1109/TITS.2018.2854895

Lesani, A., Nateghinia, E., & Miranda-Moreno, L. F. (2020). Development and evaluation of a real-time pedestrian counting system for high-volume conditions based on 2D LiDAR. Transportation Research Part C: Emerging Technologies, 114, 20–35. doi: 10.1016/j.trc.2020.01.018

Li, H., Hainen, A. M., Sturdevant, J. R., Atkison, T., Talukder, S., Mathew, J. K., … Stiles, T. (2019). Indiana Traffic Signal Hi Resolution Data Logger Enumerations. Purdue University. doi: 10.5703/1288284316998

Li, S., Sayed, T., Zaki, M. H., Mori, G., Stefanus, F., Khanloo, B., & Saunier, N. (2012). Automated Collection of Pedestrian Data through Computer Vision Techniques. Transportation Research Record: Journal of the Transportation Research Board, 2299(1), 121–127. doi: 10.3141/2299-13

Li, X., & Wu, Y.-J. (2021). Real-time estimation of pedestrian volume at button-activated midblock crosswalks using traffic controller event-based data. Transportation Research Part C: Emerging Technologies, 122, 102876. doi: 10.1016/j.trc.2020.102876

Li, X., Xu, P., & Wu, Y.-J. (2021). Pedestrian crossing volume estimation at signalized intersections using Bayesian additive regression trees. Journal of Intelligent Transportation Systems, 26(5), 557–571. doi: 10.1080/15472450.2021.1933471

Liu, G., Yin, Z., Jia, Y., & Xie, Y. (2017). Passenger flow estimation based on convolutional neural network in public transportation system. Knowledge-Based Systems, 123, 102–115. doi: 10.1016/j.knosys.2017.02.016

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. Transportation Research Part D: Transport and Environment, 63, 244–260. doi: 10.1016/j.trd.2018.05.011

MacCarley, C. A., Hemme, B. M., & Klein, L. (2000). Evaluation of Infrared and Millimeter-wave Imaging Technologies Applied to Traffic Management. SAE Transactions, 109, 373–382. Retrieved from https://www.jstor.org/stable/44699150

Malinovskiy, Y., Saunier, N., & Wang, Y. (2012). Analysis of Pedestrian Travel with Static Bluetooth Sensors. Transportation Research Record: Journal of the Transportation Research Board, 2299(1), 137–149. doi: 10.3141/2299-15

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. doi: 10.1177/0361198119838266

Miranda-Moreno, L. F., & Fernandes, D. (2011). Modeling of Pedestrian Activity at Signalized Intersections: Land Use, Urban Form, Weather, and Spatiotemporal Patterns. Transportation Research Record: Journal of the Transportation Research Board, 2264(1), 74–82. doi: 10.3141/2264-09

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. doi: 10.1016/j.trd.2013.02.008

Musa, A. B. M., & Eriksson, J. (2012). Tracking unmodified smartphones using wi-fi monitors. Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems, 281–294. Toronto Ontario Canada: ACM. doi: 10.1145/2426656.2426685

Nordback, K., Kothuri, S., Petritsch, T., McLeod, P., Rose, E., & Twaddell, H. (2016). Exploring Pedestrian Counting Procedures (Covered Research Result and Conclusion No. FHWA-HPL-16-026). Washington, DC: Office of Highway Policy Information.

Noyce, D. A., & Bentzen, B. L. (2005). Determination of Pedestrian Push-Button Activation Duration at Typical Signalized Intersections. Transportation Research Record: Journal of the Transportation Research Board, 1939(1), 63–68. doi: 10.1177/0361198105193900108

Pacala, A., & Eldada, L. (2018). United States Patent No. US10036803B2. Retrieved from https://patents.google.com/patent/US10036803B2/en

Premebida, C., Ludwig, O., & Nunes, U. (2009). Exploiting LIDAR-based features on pedestrian detection in urban scenarios. 2009 12th International IEEE Conference on Intelligent Transportation Systems, 1–6. St. Louis: IEEE. doi: 10.1109/ITSC.2009.5309697

Replica. (2020). Replica Methodology. SACOG. Retrieved from https://www.sacog.org/sites/main/files/file-attachments/replica_methodology_2020.pdf?1602683559

Ryus, P., Ferguson, E., Laustsen, K. M., Schneider, R. J., Proulx, F. R., Hull, T., … National Academies of Sciences, Engineering, and Medicine. (2014a). Guidebook on Pedestrian and Bicycle Volume Data Collection (p. 22223). Washington, D.C.: Transportation Research Board. doi: 10.17226/22223

Ryus, P., Ferguson, E., Laustsen, K. M., Schneider, R. J., Proulx, F. R., Hull, T., … National Academies of Sciences, Engineering, and Medicine. (2014b). Methods and Technologies for Pedestrian and Bicycle Volume Data Collection (p. 23429). Washington, D.C.: Transportation Research Board. doi: 10.17226/23429

Saelens, B. E., & Handy, S. L. (2008). Built Environment Correlates of Walking: A Review. Medicine & Science in Sports & Exercise, 40(7), S550–S566. doi: 10.1249/MSS.0b013e31817c67a4

Savtchenko, C. (2011). Pedestrian Tracking With A Low-Cost 3D LIDAR System On A Mobile Platform (Master's Thesis, Lehigh University). Lehigh University. Retrieved from https://core.ac.uk/download/pdf/228643945.pdf

Schauer, L., Werner, M., & Marcus, P. (2014). Estimating Crowd Densities and Pedestrian Flows Using Wi-Fi and Bluetooth. Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services. Presented at the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services, London, Great Britain. London, Great Britain: ICST. doi: 10.4108/icst.mobiquitous.2014.257870

Schneider, R. J., Arnold, L. S., & Ragland, D. R. (2009a). 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. doi: 10.3141/2140-01

Schneider, R. J., Arnold, L. S., & Ragland, D. R. (2009b). Pilot Model for Estimating Pedestrian Intersection Crossing Volumes. Transportation Research Record: Journal of the Transportation Research Board, 2140(1), 13–26. doi: 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. Transportation Research Record: Journal of the Transportation Research Board, 2299(1), 65–78. doi: 10.3141/2299-08

Schneider, R. J., Schmitz, A., & Qin, X. (2021). Development and Validation of a Seven-County Regional Pedestrian Volume Model. Transportation Research Record: Journal of the Transportation Research Board, 2675(6), 352–368. doi: 10.1177/0361198121992360

Sindagi, V. A., & Patel, V. M. (2017). CNN-Based cascaded multi-task learning of high-level prior and density estimation for crowd counting. 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 1–6. Lecce, Italy: IEEE. doi: 10.1109/AVSS.2017.8078491

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. doi: 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. doi: 10.1177/0361198121994126

Singleton, P. A., Runa, F., Humagain, P., & Utah State University. Department of Civil and Environmental Engineering. (2020). Utilizing Archived Traffic Signal Performance Measures for Pedestrian Planning and Analysis (No. UT-20.17). Retrieved from https://rosap.ntl.bts.gov/view/dot/54924

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. Transportation Research Record: Journal of the Transportation Research Board, 2672(35), 58–68. doi: 10.1177/0361198118773555

Singleton, P., Ferdousy Runa, & Prasanna Humagain. (2021). singletonpa/ped-signal-data: First release of pedestrian traffic signal data project (Version v1.0) [Data set]. Zenodo. doi: 10.5281/ZENODO.4759089

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. doi: 10.3141/2035-11

Spielberg, F. & Vanasse Hangen Brustlin, Incorporated. (2007). Determination of the State of the Practice in Metropolitan Area Travel Forecasting: Findings of the Surveys of Metropolitan Planning Organizations. Transportation Research Board. Retrieved from Transportation Research Board website: https://trid.trb.org/View/1599347

StreetLight Data. (2020). Bike and Pedestrian Metrics Methodology, Data Sources, and Validation. Retrieved from https://learn.streetlightdata.com/bike-and-pedestrian-methodology-validation

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 [Data set]. Purdue University. doi: 10.4231/K4RN35SH

Sun, S., Akhtar, N., Song, H., Zhang, C., Li, J., & Mian, A. (2019). Benchmark Data and Method for Real-Time People Counting in Cluttered Scenes Using Depth Sensors. IEEE Transactions on Intelligent Transportation Systems, 20(10), 3599–3612. doi: 10.1109/TITS.2019.2911128

Tarko, A., Romero, M., Ariyur, K., Bandaru, V., & Lizarazo, C. (2018). Detecting and tracking vehicles, pedestrians, and bicyclists at intersections with a stationary LiDAR. Presented at the 18th International Conference Road Safety on Five Continents (RS5C 2018), Jeju Island, South Korea, May 16-18, 2018. Statens väg- och transportforskningsinstitut. Retrieved from http://urn.kb.se/resolve?urn=urn%3Anbn%3Ase%3Avti%3Adiva-12915

Transportation Research Board. (2007). Metropolitan Travel Forecasting: Current Practice and Future Direction -- Special Report 288. Washington, D.C.: National Academies Press. doi: 10.17226/11981

Utsch, P., & Liebig, T. (2012). Monitoring Microscopic Pedestrian Mobility Using Bluetooth. 2012 Eighth International Conference on Intelligent Environments, 173–177. Guanajuato, Mexico: IEEE. doi: 10.1109/IE.2012.32

Van Houten, R., Ellis, R., Sanda, J., & Kim, J.-L. (2006). Pedestrian Push-Button Confirmation Increases Call Button Usage and Compliance. Transportation Research Record: Journal of the Transportation Research Board, 1982(1), 99–103. doi: 10.1177/0361198106198200113

Vu, L., Nahrstedt, K., Retika, S., & Gupta, I. (2010). Joint bluetooth/wifi scanning framework for characterizing and leveraging people movement in university campus. Proceedings of the 13th ACM International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems, 257–265. Bodrum Turkey: ACM. doi: 10.1145/1868521.1868563

Xu, F., Liu, X., & Fujimura, K. (2005). Pedestrian Detection and Tracking With Night Vision. IEEE Transactions on Intelligent Transportation Systems, 6(1), 63–71. doi: 10.1109/TITS.2004.838222

Yoichiro Iwasaki. (2008). A method of robust moving vehicle detection for bad weather using an infrared thermography camera. 2008 International Conference on Wavelet Analysis and Pattern Recognition, 86–90. Hong Kong, China: IEEE. doi: 10.1109/ICWAPR.2008.4635755

Yoneyama, A., Yeh, C.-H., & Kuo, C.-C. J. (2005). Robust Vehicle and Traffic Information Extraction for Highway Surveillance. EURASIP Journal on Advances in Signal Processing, 2005(14), 912501. doi: 10.1155/ASP.2005.2305

Zhang, D., Xu, X., Lin, H., Gui, R., Cao, M., & He, L. (2019). Automatic road-marking detection and measurement from laser-scanning 3D profile data. Automation in Construction, 108, 102957. doi: 10.1016/j.autcon.2019.102957

Zhang, Y., Zhou, D., Chen, S., Gao, S., & Ma, Y. (2016). Single-Image Crowd Counting via Multi-Column Convolutional Neural Network. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 589–597. Las Vegas, NV, USA: IEEE. doi: 10.1109/CVPR.2016.70