GEORGIA DOT RESEARCH PROJECT 21-09

Final Report

NON-MOTORIZED COUNT PROGRAM AND RISK FACTORS ASSESSMENT

Office of Performance-based Management and Research

600 West Peachtree Street NW — Atlanta, GA, 30308

March 2025

TECHNICAL REPORT DOCUMENTATION PAGE

1. Report No. Enter: FHWA-GA-24-2109	2. Government Accession No. N/A	3. Recipient's Catalog No. N/A
4. Title and Subtitle Non-Motorized Count Program and Risk Factors Assessment		5. Report Date March 2025
		6. Performing Organization Code N/A
7. Author(s) Iris Tien, Ph.D. https://orcid.org/0000-0002-1410-632X Kari E. Watkins, P.E., Ph.D. https://orcid.org/0000-0002-3824-2027 Rachael Thompson Panik, AICP, Ph.D. https://orcid.org/0000-0001-7669-5409		8. Performing Organization Report No. N/A
9. Performing Organization Name and Address Georgia Tech Research Corporation North Avenue, Atlanta, GA 30332		10. Work Unit No. N/A
		11. Contract or Grant No. RP 21-09
12. Sponsoring Agency Name and Address Georgia Department of Transportation (SPR) Office of Performance-based Management and Research 600 W. Peachtree St. NW Atlanta, GA 30308		13. Type of Report and Period Covered Final [March 2022 - March 2025]
		14. Sponsoring Agency Code N/A
15. Supplementary Notes Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.		
16. Abstract Non-motorized volumes at different locations in a transportation network are important data for understanding safety and travel needs for people who bike and walk. In Georgia, however, there are limited data about non-motorized volumes in transportation systems given, and in general, these modes volumes are difficult to predict due to high variability spatially and temporally. It is of interest to Georgia DOT and other transportation agencies how to leverage existing proxy (i.e., indirect) data to better understand biking and walking travel. The research in this technical report explores several facets of non-motorized volumes, proxy data, and safety: (1) the use of proxy data in non-motorized counting through literature review, (2) time series classification of proxy data using machine learning, (3) Bayesian time series models and proxy data for predicting biking and walking volumes at a high resolution, (4) and the relationship among non-motorized volumes and other bicycle and pedestrian risk factors. This work found that some proxy data, specifically Strava data for bicyclists and Automated Traffic Signal Performance Metrics data for pedestrians, are useful for predicting high-resolution biking and walking volumes, but that sufficient observed data are needed in places where traffic volumes vary greatly hourly or daily. Transportation agencies and researchers should use this work to direct future non-motorized counting and modeling.		
17. Key Words Non-motorized monitoring; bicycle safety; pedestrian safety; travel trends		18. Distribution Statement No restrictions.

19. Security Classification (of this report)	20. Security Classification (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	203	N/A

Georgia DOT Research Project 21-09

Final Report

NON-MOTORIZED COUNT PROGRAM AND RISK FACTORS ASSESSMENT

By

Iris Tien, Associate Professor and Principal Investigator
Kari E. Watkins, PE, PhD
Rachael Thompson Panik, AICP, PhD
Julie Shorey, EIT

Georgia Institute of Technology

Contract with
Georgia Department of Transportation

In cooperation with
U.S. Department of Transportation, Federal Highway Administration

Georgia Institute of Technology

March 2025

METRIC CONVERSION CHART

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	square millimeters	mm^2
ft^2	square feet	0.093	square meters	m^2
yd^2	square yard	0.836	square meters	m^2
ac	acres	0.405	hectares	ha
mi^2	square miles	2.59	square kilometers	km^2
		VOLUME		
fl oz	fluid ounces	29.57	milliliters	mL
gal	gallons	3.785	liters	L
ft^3	cubic feet	0.028	cubic meters	m^3
yd^3	cubic yards	0.765	cubic meters	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 (or "metric ton")	Mg (or "t")
		TEMPERATURE (exact degrees)		
°F	Fahrenheit	5 (F-32)/9 or (F-32)/1.8	Celsius	°C
		ILLUMINATION		
fc	foot-candles	10.76	lux	lx
fl	foot-Lamberts	3.426	$candela/m^2$	cd/m^2
		FORCE and PRESSURE or STRESS		
lbf	poundforce	4.45	newtons	N
lbf/in^2	poundforce per square inch	6.89	kilopascals	kPa

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	square millimeters	0.0016	square inches	in^2
m^2	square meters	10.764	square feet	ft^2
m^2	square meters	1.195	square yards	yd^2
ha	hectares	2.47	acres	ac
km^2	square kilometers	0.386	square miles	mi^2
		VOLUME		
mL	milliliters	0.034	fluid ounces	fl oz
L	liters	0.264	gallons	gal
m^3	cubic meters	35.314	cubic feet	ft^3
m^3	cubic meters	1.307	cubic yards	yd^3
		MASS		
g	grams	0.035	ounces	oz
kg	kilograms	2.202	pounds	lb
Mg (or "t")	megagrams (or "metric ton")	1.103	short tons (2000 lb)	T
		TEMPERATURE (exact degrees)		
°C	Celsius	1.8C+32	Fahrenheit	°F
		ILLUMINATION		
lx	lux	0.0929	foot-candles	fc
cd/m^2	$candela/m^2$	0.2919	foot-Lamberts	fl
		FORCE and PRESSURE or STRESS		
N	newtons	0.225	poundforce	lbf
kPa	kilopascals	0.145	poundforce per square inch	lbf/in^2

* SI is the symbol for the International System of Units. Appropriate rounding should be made to comply with Section 4 of ASTM E380. (Revised March 2003)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

AADB Annual Average Daily Bicyclists

AADNT Annual Average Daily Non-Motorized Traffic

AADP Annual Average Daily Pedestrians

AADT Annual Average Daily Traffic

AANT Annual Average Non-Motorized Traffic

ACS American Community Survey

ATSPM Automated Traffic Signal Performance Measures

BAMs Bike Activity Metrics

CART Classification and Regression Trees

CSV Comma-Separated Values

CVIs Cluster Validity Indices

DBI Davies-Bouldin index

DI Dunn Index

DTW Dynamic Time Warping

ED Euclidean Distance

GDOT Georgia Department of Transportation

GIS Geographic Information System

LPPD Log Point-Wise Predictive Density Leave-One-Out Cross Validation

LPPD LLog Point-Wise Predictive Density

MAE Mean Absolute Error

MARTA Metropolitan Atlanta Rapid Transit Authority

MCMC Markov Chain Monte Carlo

NCHRP National Cooperative Highway Research Program

PAMs Pedestrian Activity Metrics

PCA Principal Components Analysis

QGIS Quantum Geographic Information System

RMSE Root Mean Squared Error

SI Silhouette Index

SSE Sum of Squared Error

SSP Safe Systems Pyramid

TMAS Travel Monitoring Analysis System

TMG *Travel Monitoring Guide*

U.S. United States

VRU Vulnerable Roadway User

WAIC Watanabe-Akaike Information Criterion

EXECUTIVE SUMMARY

This research report is the result of investments of the Georgia Department of Transportation's (GDOT's) Office of Performance-based Management and Research. It summarizes findings from research on non-motorized (i.e., bike and pedestrian) hourly volume estimation using proxy data sources, including self-reported bike rides via crowd-sourced applications and pedestrian actuations at intersections. It explores the possibility of using these proxy data sources to estimate volumes at the hourly level at intersections and roadway segments in Georgia, and it discusses the value of such high-resolution estimates in project selection and safety analyses. Specifically, this research uses Bayesian time series models and two cross-validation approaches (spatial and temporal) to explore the feasibility of using these models to estimate bike and pedestrian volumes in Georgia. This work found that proxy data sources can be used to estimate hourly volumes when calibrated with local counts, and that it is possible for them to be used to estimate volumes for locations where data has not been collected, but more data is needed than was collected for this work. Specifically, more long-term counts and longer short term counts are needed to further calibrate the models more effectively, creating more accurate out-of-sample estimations. The final sections of the report discuss recommendations for GDOT in future non-motorized modeling and counting based on the findings of this research.

INTRODUCTION

Traffic safety is a crisis in the United States, especially for people biking and walking. While less than 1% of all trips are made by bicycle, more than 2% of the total reported traffic deaths are bicyclists, indicating an over-representation of cyclists in crash statistics [1]. Similarly, pedestrian fatalities are also high and have been increasing steadily for a decade [2]. Despite the many bicycle and pedestrian crashes, injuries, and fatalities occurring on publicly owned assets, transportation agencies and local governments have been unsuccessful at reducing crashes. While traffic safety is a complex problem with many facets, a common problem arises around decision-making to improve safety outcomes. It is often not clear to decision-makers where or how to direct safety improvements. While there are many reasons for this depending on the agency and locality – from inadequate political will to limited funds for new infrastructure to poor quality data to poor safety to ineffective safety cultures among transportation organizations – all of which limit what they can know and how they can react to transportation safety problems.

First, this research explores activity patterns in non-motorized travel data. Transportation organizations, like metropolitan planning organizations and departments of transportation, are the primary monitors of traffic in the United States (U.S.), but many agencies (including Georgia) do not systematically monitor biking and walking. This gap of information creates a self-defeating cycle: limited information about biking and walking means that agencies who may want to begin monitoring may not have a good means to begin counting, as system-wide counting requires some understanding of travel patterns in the system. This cycle is compounded by the reality of biking and walking travel – it is highly variable depending on time of day, seasons, and events. While many agencies struggle to know how to monitor biking and walking effectively, many agencies also have access to high-resolution proxy (i.e., indirect) data that represent a portion of non-motorized traffic in the system. These data have yet to be fully explored for their potential to extract travel

patterns. This research uses unsupervised machine learning to extract discrete categories of activity from two datasets – one that captures pedestrian "button pushes" at intersections, and one that captures self-reported bike rides.

Second, this research uses novel methods of activity modeling to create bicycle and pedestrian volume estimates from proxy data, which are important for quantifying risk and exposure. One of the more prominent challenges facing transportation agencies and decision-makers is that they have sparse information about how and when people are walking and biking in the transportation systems. Effective responses to transportation safety problems require realistic measures of exposure to inform safety assessments and prioritize interventions that can reduce system-wide risk. Because the number of bicycle and pedestrian crashes is confounded with the number of people biking and walking, simply addressing safety problems where crashes are most prevalent is not an adequate approach to addressing safety problems through the lens of crash risk. Although crashes, injuries, and fatalities are never negligible, transportation agencies must often funnel limited resources towards the most critical safety problems – which requires at least relative measures of exposure (e.g., volumes of cyclists and pedestrians, volumes of vehicles, etc.), even though true measures of exposure are more complex [3]. This research applies these activity models in two ways – in-sample predictions, where predictions of biking and walking volumes are calculated at places where the model has observe rations on which to calibrate; and out-of-sample predictions, where the predicted volumes are made for locations that have not been developed with data from those locations (i.e., testing data).

While vehicular travel is measured universally and is required for state agencies to receive federal dollars, measures of pedestrian and bicycle activity are often not collected at all, which limits practitioners' understanding of infrastructure needs and safety problems. The State of Georgia, one of the major funding sources for this work, specifically does not systematically monitor biking and walking. While state and local agencies collect non-motorized traffic counts for warrants and traffic impact analyses that are useful in themselves, opportunistic counting does not capture enough information to truly direct safety projects; they are not regularly used to justify safety projects, nor

are they used to inform network-level understanding of non-motorized activity. There is a clear need for better methods for estimating pedestrian and bicycle volumes in a transportation system – preferably methods that make use of existing resources and minimize data collection needs. This report seeks to help address that problem by harnessing data sources already frequently used by transportation agencies and a Bayesian framework to estimate biking and walking volumes in the Georgia context.

The remainder of this report is organized as follows:

- *Chapter 1: Literature Review* – First, a summary of relevant literature is presented, beginning with the state of biking and walking safety in the United States and Georgia, followed by a synthesis of roadway safety theories as they relate to biking, walking, and exposure to crashes. Then, an overview of typical systematic counting methods is synthesized, concluding with a summary of activity modeling using alternative sources of data, like crowdsourced data.

- *Chapter 2: Clustering of Non-Motorized Count Data* – This chapter explores the first aim of this research. It contains the methods and results of several time series clustering approaches applied to non-motorized proxy data. Categories of non-motorized activity can help agencies who have yet to create a count program begin counting or direct existing non-motorized counting efforts. This chapter assesses the potential of using unsupervised machine learning techniques to create these categories of activity.

- *Chapter 3: Predicting Non-Motorized Volumes In-Sample* – Next, this chapter presents novel methods for estimating non-motorized traffic volumes at intersections and roadway segments for pedestrians and bicyclists, respectively. Using a new (to civil engineering) Bayesian framework for volume estimation of bicycle and pedestrian volumes, this chapter demonstrates that even biased and indirect measures of biking and walking can be leveraged to make high-resolution estimates of biking and walking in some contexts, focusing on making predictions at places where observed data was collected. The methodology developed in this

chapter could be used by agencies to estimate hourly rates of biking and walking so that they can better measure risk per person as a prioritization metric for safety projects.

- *Chapter 4: Risk Factors Assessment* – Exposure (measured via bike and pedestrian volumes) is not the only factor that influences bicycle and pedestrian safety outcomes. This chapter synthesizes the relevant bike and pedestrian safety literature and then summarizes risk factors that are relevant to GDOT, describing each risk factors' relationship to exposure.

- *Chapter 5: Spatially Out-of-Sample Predictions* – Chapter 3 developed models to predict bike and pedestrian volumes at places where observed data can calibrate predictions using Bayesian time series models. This chapter explores the potential of these models to make predictions at locations where observed data has not calibrated the models through a training-testing approach, focusing only pedestrian proxy and observed data.

- *Conclusion* – Finally, this chapter summarizes the research presented in this report, including recommendations for GDOT to operationalize the findings from this research.

CHAPTER 1: LITERATURE REVIEW

This literature review is situated at the intersections of transportation safety, non-motorized traffic monitoring, and crowd-sourced data. The review covers the following sections: Vulnerable Roadway User (VRU) safety trends in the US and in Georgia, roadway safety theories, the relationship between traffic volumes and safety (i.e., exposure), and the state of non-motorized traffic counting in the practice and in research.

VULNERABLE ROADWAY USER SAFETY

Vulnerable Roadway User Safety in Georgia

In the United States, although less than 1% of all trips are made by bicycle, more than 2% of the total reported traffic deaths are bicyclists, indicating an over-representation of cyclists in crash statistics [1]. Similarly, pedestrian fatalities are also high and have been increasing steadily for a decade [2]. In Georgia, in 2019, there were 236 pedestrian and 21 bicyclist fatalities [4]. This represents a 22% increase over 2015 for pedestrians, and a similar trend in bicycle crashes over the same time period. Georgia, however, is not unique in this increasing trend. The number of crashes in the U.S. has also increased in recent years; even though overall traffic fatalities declined between 2004 and 2019, pedestrian and bicyclist fatalities have increased, both in terms of the absolute numbers of deaths and as a share of all traffic deaths [5, 2].

One of the many reasons that this safety problem persists in Georgia and in the United States is a lack of data to completely define the safety problem. Agencies that address transportation safety problems have limited resources, and limited data about the state of the infrastructure within the system and about when and where people use it. The following subsection defines the theoretical constructs and units of measurement related to safety in the transportation system.

Road Safety Theories

It is important to situate this research into theories of traffic crashes, injuries, and fatalities. This is important, as "one of the major problems of roadway safety research is that most of it does not have a strong theoretical basis" [6, p. 331]. Poor theoretical foundations often preclude researchers and practitioners from interpreting findings and building appropriately off of past work in the literature. A clear conceptual model is important in creating models that describe traffic safety because the conceptual model informs the mathematical models.

There are many possible approaches to safety within a system that could be used as scaffolding for this work, but this research is viewed through the lens of the "Safe Systems Pyramid," developed by Ederer et al. [3]. Informed by established and demonstrably effective approaches to injury prevention developed in the public health and injury prevention, as well as the hierarchy of controls, the Safe Systems Pyramid (Figure) is a framework that correlates the effectiveness of safety measures in the transportation system with both the amount of individual effort required for the safety measure to have effect, as well as the potential effect size of the safety measure. For example, this approach puts safety interventions related to education at the "top," or least effective part, of the pyramid. Interventions in this category would include efforts like roadway safety campaigns. To have any effect on safety, educational interventions require people to be able to access the education *and* for them to correctly interpret the meaning of the messages and then apply that knowledge to their own behavior (i.e., high levels of individual effort). Since not all people in a population will have access to a campaign, and an even smaller number of those people will integrate that knowledge, educational campaigns impact a smaller portion of the population exposed to that intervention.

When applied in transportation practice, it is a way for transportation professionals to evaluate the effectiveness of safety interventions. In a research setting, however, the Safe Systems Pyramid informs one main aspect of this work. It identifies that the risk factors of speed, closeness of roadway users in time and space, socioeconomic factors, and infrastructure are to be considered in safety models, as they are the context in which traffic events occur – as should some kind of

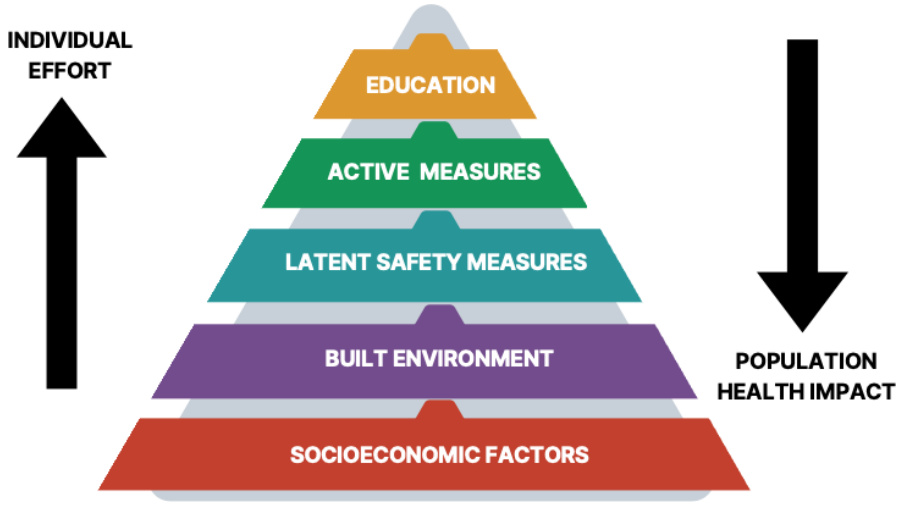


Figure 1: Illustration. Safe Systems Pyramid from Ederer et al. (2023)

quantification of how many people are potentially impacted by an intervention. Some of the risk factors are often translated into infrastructure-related proxies: design speed, bicycle- or pedestrian-designated infrastructure, crossing distances, sociodemographic variables etc. But it also states that the effect of an intervention is directly related to the number of people who are potentially impacted by that intervention, which requires some kind of estimate of how many people are exposed to the risks and are effected by the interventions.

Building from this lens, the comprehensive roadway safety framework proposed by Schepers et al. [6], formally delineates the relationships among risk factors and contextual factors that influence roadways' safety. The authors present a thorough conceptual model linking road safety, exposure, and risk, shown in Figure . This framework departs from other older, widely accepted models, such as Heinrich's Domino Theory [7], which assumes a linearly causal chain of events that leads to crashes so that eliminating one of the causes prevents unsafe outcomes. But this is an overly simplistic approach to conceptualizing traffic safety, as crashes more often are resultant of interactions of many variables [8] – which requires more complex, non-linear models [6].

To this end, Schepers et al.'s model of roadway safety shown in Figure is multifaceted and far-reaching. Each element of the framework is summarized as follows.

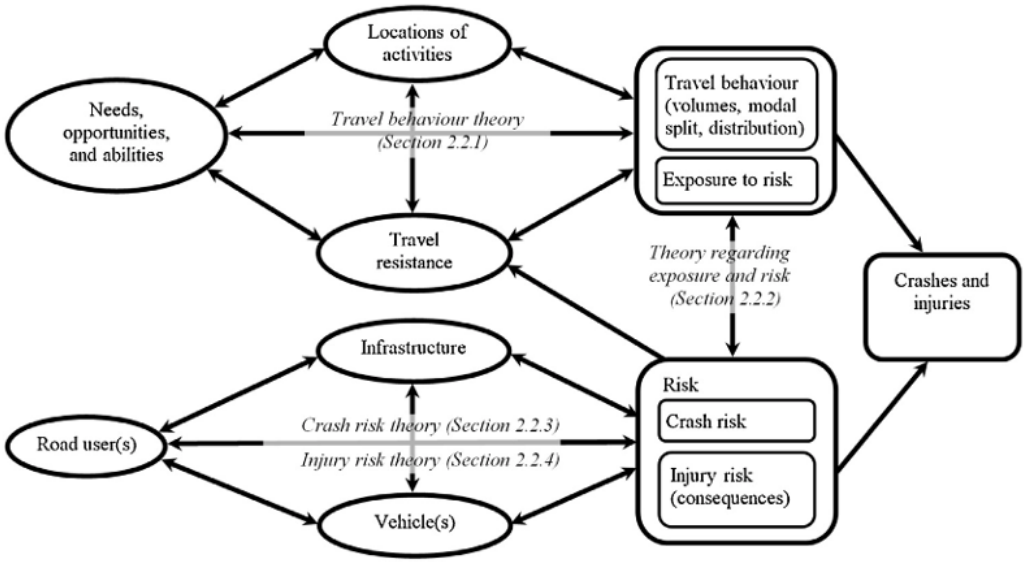


Fig. 1. Conceptual framework for road safety, including exposure and risk (sections describing the theories are referred to in parenthesis).

Figure 2: Illustration. Safety Framework from Schepers et al.

- *Travel behavior theory*: Travel is influenced by human behavior on many levels of consideration, including and choices surrounding behavior are modeled by the location of activities, the reason of travel, the cost/friction of travel, abilities and resources, and subjective norms, values, and attitudes.

- *Crash risk theory*: Risk of a crash results from interactions among the "three traffic safety pillars" – road users, vehicles, and infrastructure [9]. This is comparable to the Haddon Matrix approach to injury prevention, as noted by Ederer et al. – host, agent, and environment [10, 3].

- *Injury risk theory*: Crashes will be fatal when the energy and forces transferred to the person in the crash exceeds a physiological tolerance, which depends on many factors, including age of roadway users, preexisting conditions, and the speed at which the crash occurred [3].

- ***Theory of exposure and risk***: Most relevant to this report is the relationship between exposure and risk. This is captured in the conceptual model, indicating that people who travel are exposed to only the risks that are present, but that traveling is required for many (although

not all) types of exposure to accrue. Exposure to risk in this model also influences travel resistance.

In Scheper et al.'s framework, exposure is required for measuring crash risk *and* injury risk – both of which are required for directing infrastructure adaptations. Given the importance of exposure in the understanding of traffic crashes, it is paramount to understand baseline exposure; without exposure, it is not possible to determine whether an increase in the number of crashes is due to poorly designed facilities (or some other problem with the transportation system), or due to an increased number of people being exposed to crashes [5, 6]. The following section further defines exposure and common units of analysis.

Exposure to Crashes

Formally, exposure is defined in public health and in transportation engineering as proximity to and/or contact with a risk agent such that transmission of harmful effects of the agent may occur, usually in a specific period of time (adapted from Schepers et al. [6]). Applied to the transportation system, exposure is a measure of opportunity for a crash to occur [11]. This is related to both risk – which is the probability of a crash occurring (in a given time frame) – and injury probability – which is the likelihood of an injury given a crash (in a given time frame). Total traffic injury is the multiplicative result of exposure, risk, and injury probability. This conceptualization of risk and exposure means that risk cannot be understood without a measure of exposure, and that predicting risk for decision-making is dubious due to possible confounding of risk and exposure.

Since exposure is essential to understanding safety and informing safety interventions, it is useful to summarize what kinds of metrics are used to measure exposure to crashes. There are many ways to quantify exposure to risk for non-motorized travelers, and there is debate about which metrics are the most informative and useful [12]. The usefulness of each metric in general, however, depends on whether geographic scale and the corresponding unit of analysis match the scale of safety intervention of interest, or the unit at which crashes actually occur [5].

The largest units of analyses are on the area-wide scale with commensurately large time units

(e.g., number of trips by mode per year, distance traveled by mode per year, number of travel- ers within a census tract, average time spent traveling in a region). Area-wide exposure metrics correlate to area-wide levels of risk measurements (e.g., crashes per number of trips in a county, crashes per share of commute bicycle trips in a census tract). While these may be helpful metrics in directing regional or municipal policies, they lack the specificity needed for infrastructure interventions; crashes occur at intersections and on roadway segments, so much of the intervention effort to reduce crash risk and total crashes must also act at that level. Research in Hong Kong found that these large-level estimates of exposure, like population, population density, and mode share of cyclists, are not high-quality measures of exposure and can produce misleading findings about high-risk areas [13].

Given the challenges of large-scale measurements, there is need for risk factors associated with [non-motorized] crashes, including measures of exposure, to be measured at the scale at which [bike and pedestrian] crashes occur [14]. A more granular unit of analysis is exposure metrics on the facility-level scale. These can be measured in a variety of ways, including the number of pedestrian crossings at an intersection, Annual Average Daily Non-Motorized Traffic (AADNT), or Annual Average Daily Pedestrians (AADP) or Annual Average Daily Bicyclists (AADB), respectively, per segment, and daily vehicular volumes per corridor [12]. These measures are most useful for informing infrastructure-level safety interventions, as they share the same units of analyses as the infrastructure improvements. These data are collected as a part of count programs or through proxy measures, such as cell phone data or crowd-sourced data. Systematic data collection methods are summarized in the following section.

While these more granular estimates of exposure are an improvement over area-wide estimates, there are still limitations to these units. First, many still use larger time units of analysis. Annual Average Non-Motorized Traffic (AANT) is of course reported at a yearly level, and even AADNT, which is a daily measure, still lacks nuances that are relevant in measuring the impact of certain risk factors. For example, low lighting conditions are risk factors for both bicyclists and pedestrians [15, 16], which obviously varies temporally within 24-hours.

Second, while AADNT and other measures of biking and walking volumes are frequently based on large quantities of data, these measures are reduced to point estimates. Single values of biking and walking activity are useful for their simplicity, but they are often problematically reductive. They can imply more accuracy than they actually should, especially when the estimates are annualized over 365 days. They also do not capture the realities of biking/walking travel variability; averaging counts over 365 days both reduces an information-rich time series to a single value, but the average yearly value may not convey much physical meaning given highly seasonal and weekly variation on a given segment or intersection.

Finally, *surrogate measures of safety* are also used to measure exposure. These measures do not come from crash records, but instead they are safety metrics that intend to capture the conditions in which roadway crashes occur. Most surrogate measures are calculations based on the nearness in time and space of travelers, often at conflict points [17, 18, 19, 20]. The metrics are justified conceptually, as (multi-user) crashes can only occur when travelers are near each other physically and temporally.

Said another way, surrogate safety metrics are designed to capture "critical safety events." Traffic events are defined as the "interaction between roadway users" [21, p. XII]. As per Johnsson et al., traffic events occur constantly in the transportation system [18], but some traffic events are riskier than others. The surrogate measures of safety, then, attempt to capture those traffic events that carry meaningful risk of an unsafe outcome, or "critical safety events," (i.e., nearness of a pedestrian and a driver in an intersection).

The benefits of using surrogate measures of safety are three-fold. First, surrogate safety measures are inherently more proactive than safety metrics derived from crash metrics. They do not rely on historic crash data (which may be poor quality and contain police officer bias), and they do not require waiting on new crashes to occur to obtain new data (which is ethically dubious). Second, using surrogate measures of safety for vulnerable road user safety analysis addresses one of the key challenges of bicycle and pedestrians safety research: under-reporting of biking- and walking-related crashes, especially less severe crashes [22, 18, 20, 23]. Finally, surrogate mea-

sures may be measured more readily and more quickly with the growth of intelligent and connected transportation systems and traffic monitoring via cameras.

The major challenge with surrogate measures of safety, however, is determining which measures of critical events best corresponds to crashes. This is especially difficult for surrogate measures for bicyclists and pedestrians [18, 20]. As per Johnsson et al., the "ideal" metric should be theoretically sound, validated, and reliable [18]. In this context, "validated" metrics are those that consistently relate to crash outcomes, and reliable means that the measure should effectively estimate frequency and/or severity across several contexts. It has been challenging to find surrogate safety measures that are validated and reliable for bicyclists and pedestrians. Many existing surrogate measures of safety do not capture traffic events that most relate to bike and pedestrian crashes, like swerving behavior, bike- or pedestrian-centric metrics (e.g., step changes in pedestrian gate) [18, 20]. There is a clear need for additional metric evaluation, and this report will attempt to contribute to the field in this area.

Summary

Bicyclists and pedestrians are over-represented in crash statistics in the United States and in Georgia, and they are more likely than many other road users to be harmed or killed during a crash. There is a need for data-informed safety counter measures that will most reduce the exposed population to the risk of injury and death. Often these counter measures that are most effective are built environment measures, but decisions about where and how to design the solutions are uninformed without using measures of risk – which require both a measure of an unsafe outcome (e.g., crash or a critical traffic event) and a measure of exposure. The measure of unsafe event must capture validated critical traffic events, like crashes or validated surrogate safety measures, and the exposure measures must capture whether or not a person is at risk on a meaningful scale.

NON-MOTORIZED COUNTING METHODS

The previous section of the literature identified the need for exposure measures in evaluating risk. A reasonable measure for exposure in many cases is traffic volumes. Vehicular volumes (i.e., cars passing a point per time period), vehicular flows (i.e, cars passing a point per hour), or annualized vehicular volumes (i.e., Annual Average Daily Traffic (AADT)) are often used for quantifying risk, exposure, and predicted crashes of people driving.

The same can be said for non-motorized traffic. Biking and walking counts can be collected through transportation agency-based non-motorized count programs, and those counts can be used in various forms (i.e., average annual daily non-motorized traffic, or AADNT) as an exposure metric to quantify risk. In addition to the safety assessments, there are other reasons to quan- tify biking and walking travel. Counts collected from these programs serve a variety of purposes, like justifying infrastructure projects, conducting before and after scenarios, adjusting signal tim- ing, prioritizing projects, informing public policy, and assessing physical activity [5, 24]. The following subsection describes non-motorized traffic programs, which are the primary format for systematically counting bicyclists and pedestrians (and sometimes other forms of micromobility).

Agency-Based Counting

Traditional non-motorized count programs are typically designed and operated at state, regional, or municipal levels of government, often within a department of transportation or a regional transportation government [14]. The structure of these programs is modeled after motorized traffic monitoring programs [25], where activity is measured by the number of bicyclists and pedestrians per hour or per 15-minutes. They can be started with relatively minimal equipment – perhaps several thousand dollars for counters and for database management. They can be expanded to be more comprehensive, however, to contain full-time staff, many high-quality counters at many locations [26], and partnerships with local organizations like school and advocacy groups to obtain more counts.

Collected counts are typically converted into annualized metrics like annual average daily bicycles or pedestrians (AADB/AADP). Generally, a count program consists of two different kinds of monitoring: permanent counters, which continuously collect counts to measure depth of activity; and short-term counters which collect anywhere from a few hours to multiple days of data to measure breadth of activity within the system [24, 27, 25, 28, 29]. With permanent and long-term counters in place (and once there is data for a year of observed activity), all count sites are then grouped by weekly or daily similarities in non-motorized activity, i.e., factor groups [29]. Short-term count locations are grouped by their similarities in activity over time to long-term counters (where possible), resulting in locations that are grouped by similar trends in bike or pedestrian volumes. Examples of factor groups could be "recreational facilities," which have low counts during the week but higher weekend volumes, or "commute facilities," which have regular morning/evening count peaks on weekdays. Multipliers, or "temporal adjustment factors," are derived from the permanent counter(s) in each factor group and applied to the short-term counters in the same factor group to adjust for their short monitoring period; the underlying assumption is that travel trends at one location will be similar to trends at another "similar" location in the network.

The 2013 edition of the Federal Highway Administration's *Travel Monitoring Guide* (TMG) was the first to include guidance for counting non-motorized travelers [29]. It included information on data format and quality requirements for the Travel Monitoring Analysis System Travel Monitoring Analysis System (TMAS), which is a national database used for traffic monitoring nation-wide. Now, the more recent *Travel Monitoring Guide* from 2022 (TMG2022) [25] and the National Cooperative Highway Research Program (NCHRP)'s *Report 797 Guidebook on Pedestrian and Bicycle Volume Data Collection* [28], are the main resources for state-wide count programs [5].

The methods described in these guides indicate multiple years of bicycle and pedestrian data collected at a variety of sites within a network are required to create meaningful estimates. The general structure of a count program is as follows: [24, 27, 25, 28]:

- *Continuous counters*: Central parts of any non-motorized count program are continuous, or

long-term, counters. These counters collect counts 24 hours per day, 365 days per year, usually in 15-minute increments. They are often placed in ”pinch points” of non-motorized traffic, such as river, train track, or bridge crossings. In addition to their value of monitoring counts at those sites, they are also used to create adjustment factors, or multipliers that account for systematic error, like seasonal variations and temporal variations, as well as spatial differences.

- *Short-term counters*: In addition to long-term counters, short-term counters are also key elements of a non-motorized count program. They collect counts for 24 hours to 7 days, where 7 days is the preferred minimum amount of time [5]. Locations of short duration counts should be as contextually varied as possible in order to collect a representative sample of roadway and land use characteristics [30, 31]. There are many ways to decide where to place short term counters, but examples of sampling strata can be socioeconomic factors, road type, proximity to major non-motorized trip generators such as universities, population density, and intersection density. For larger study areas there are two common approaches to count site selection within each strata: by random selection from all road and path segments in each sampling strata or by specific manual selection from sites identified by local experts to have non-motorized activity.

- *Monitoring*: With short term and continuous counters in place, data should be collected from the long-term counters for at least a year prior to creating adjustment factors. To use short-term counts for the purpose of generating AADNT, counts should be collected for 7 days continuously at a single location before rotating that counter to another site. For data management and quality control, this source recommends reviewing the data manually through visual inspection to flag unrealistically high counts, long series of no counts, or sudden changes in count trends at a site.

- *Site Grouping*: Once there is data for at least a year, all count sites can be grouped by similarities in non-motorized activity. The grouping is either done by manual inspection of

the distribution of non-motorized activity or through statistical methods, like clustering or time series analyses. Count sites should be grouped with other sites that have similar travel patterns. Starting with continuous counters, they can be labeled with other counters that have similar weekly trip distributions. Then, short term counters are assigned into each of these groups based on similarities of activity distribution. Often these categories will be based on visual analyses of differences in the weekly temporal distribution, like commute groups (i.e., weekday peaks and weekend troughs in activity) or recreational groups (i.e., weekday troughs and weekend peaks in activity), or by similar groups for urban and rural contexts. These groups can then be used to derive temporal adjustment factors, with the assumption that the temporal distribution of non-motorized travel at one location in the network will be similar to the temporal distribution of travel at a "similar" location in the network.

This approach to counting non-motorized travelers has some strengths. First, the structure of data collection and volume estimation is similar to approaches used for systematic monitoring of vehicular travel, which makes the volume metrics like AADNT more intuitive. It is also good that these counts would presumably be collected and archived in a statewide database so that they can be used for many public, statewide purposes, including planning, project prioritization, and justification for new projects.

Within this structure for non-motorized count programs, there are also several limitations. First, the literature suggest several ways to decide where to place short term counters; examples of sampling strata can be socioeconomic factors, road type, proximity to major non-motorized trip generators such as universities, population density, and intersection density [31]. For larger study areas there are two common approaches to count site selection within each strata: by random selection from all road and path segments in each sampling strata or by specific manual selection from sites identified by local experts to have non-motorized activity. While institutional knowledge and engineering judgement should be a part of all analysis and engineering decisions, we believe there are opportunities for mathematical methods to support this intuition and perhaps guide decision-makers in geographic areas where their intuition lacks.

Second, creating factor groups of "similar" activity patterns lacks specificity. What does it mean to be similar? In practice, there are two main ways for creating groups once counts have been collected. The first is through manual inspection of the daily and/or weekly distribution of non-motorized activity. Practitioners will create categories based on visual analysis of differences in the weekly temporal distributions, like commute groups (i.e., weekday peaks and weekend troughs in activity) or recreational groups (i.e., weekday troughs and weekend peaks in activity), or by similar groups for urban and rural contexts. While this approach for creating factor groups may be effective in cases where there are clear patterns, it may not always be obvious where there are differences in travel patterns, and analyst bias may mislead grouping.

An alternative method for creating groups is through statistical methods, like clustering, or non-supervised machine learning. While clustering is mentioned in the TMG [25], what is not specific in this guide is the method of clustering that should be used to create these categories. Given the nature of clustering time series (e.g., sensitivity to distance metrics, no theoretical basis for determining the number of clusters a priori), the clustering method and approach to selecting the number of clusters can be critical to create effective activity pattern groups.

Other Data Sources and Methods for Non-Motorized Travel Monitoring

Given both the challenges and limitations of formalized count programs, researchers have taken to alternative approaches to estimating biking and walking volumes (as a means of exposure). Increasing digitization and large datasets collected passively from devices give opportunities to harness proxy measures of biking and walking. While each proxy source is inherently incomplete, they can be combined with observed counts in models as a source of additional information, and they may be used to make counting more efficient to conserve resources. The following subsections describe additional sources of count data and methods for predicting activity by mode.

Pedestrian Data and Methods

In general, there is limited research on sources of proxy data and methods for volume estimation for pedestrians in the U.S. context.

One exception to this is the emerging body of work that uses pedestrian actuations and data from traffic signal control logs at intersections as a source of proxy data. In a 24-hour pilot study conducted on an actuated traffic signal in Oregon, researchers showed that pedestrian phases can be used to estimate pedestrian volumes [32]. By harnessing the pedestrian actuations at a traffic signal control log, they showed that in this context – a suburban signal that required actuation for pedestrian crossings – actuations both correlated with observed pedestrian crossings and could be used to create an adjustment factor (i.e., the ratio of observed pedestrian volume to the number of pedestrian phases recorded) to calculate AADP.

Later, a more in-depth series of studies both calibrated and validated models based on traffic signal controller logs for estimating pedestrian crossing volumes [33, 34, 35, 36, 37, 38]. Singleton et al. and Humagin et al. used traffic signals in Salt Lake City, Utah, that were equipped with with Automated Traffic Signal Performance Measures (ATSPM) controllers, which capture all traffic events at a signal at the millisecond level, to predict pedestrian volumes at a variety of intersection types (HAWK signals, high activity signals, signals with long cycles, etc.).

To build these models, they selected locations with a variety of pedestrian activity, where "activity" was defined based on seven different measures of pedestrian activity called Pedestrian Activity Metrics (PAMs), which were calculated from pedestrian phases or actuations. The PAMs were used to cluster signals into different categories of activities. Out of the 521 signals available for analysis (e.g., equipped with ATSPM, with traffic signals, etc.), they selected 90 locations for observations. At least 24-hours worth of pedestrian counts were conducted at each intersection via traffic cameras, where 1 hour was used as the temporal unit of analysis. They then created simple regression models that were designed for a factoring approaching, creating multiplicative factors to estimate pedestrian volumes at intersections.

After iterating with 150 different model possibilities that were combinations of different PAMs,

model forms, and phasing (assessed quantitatively through Root Mean Squared Error (RMSE) metrics), they found the following specifications for each context [34]:

1. *HAWK signals*: Quadratic specifications were selected for HAWK signals, and they performed best using a PAM that is calibrated using captured sequences of actuations. Their models produced high correlation, $R^2 = 0.915$.

2. *Signals with pedestrian recall*: For both high and low levels of pedestrian activity, they found that a quadratic specification performed best. High-level activity locations performed the worst out of all the models in their study, although the model still had a correlation of 0.649. For locations where there were low levels of pedestrian activity, the quadratic model had good correlation of 0.804 and low error metrics.

3. *Signals with **no** recall*: Both high- and low-activity level models were specified with piecewise models. They separately modeled signals with longer cycle lengths and shorter cycle lengths, where "long" lengths were equal to or greater than 90 seconds.

While the models they proposed worked well in a variety of contexts in Utah, further research should explore how these methods apply to other urban contexts, where pedestrian behavior may be different or there are more widely varying levels of pedestrian demand.

Bicycle Data and Methods

While there is limited research on estimating pedestrian volumes with proxy data, there is much more research on using proxy sources and models to estimate bicycling. The most common sources of proxy data are bike share program data [39] and Strava data [40, 31, 41, 42, 43]. Strava [44] is a free smart phone application and social media platform that people use to track human-powered recreational and commute trips, including bike trips. Strava Metro is the research and data division of Strava [45], which provides researchers and public agencies with high-resolution data on non-motorized travel. These data contain segment-level hourly volumes for an entire roadway network – a notable advantage over other third-party (i.e., privately developed) datasets, which often can

only generate area-level estimates of bike activity. A challenge of using Strava data, however, is that it does not capture all bike trips; it is inherently incomplete given that it is self-selected, and the data are biased towards young, white males [46, 47]. Miah et al. note that Strava represents somewhere between 2% and 10% of all bicycle traffic at some locations, but others – likely those that are more recreational facilities – may see much higher shares, sometimes up to 30%. This bias, however, appears to be systematic [40], and Strava data has been shown to produce good-quality estimates of AADB in certain contexts when calibrated with counts collected from a variety of roadway types [42, 31, 43].

There are several common approaches to modeling estimation of bike volumes (usually in the form of AADB), summarized as follows:

- *Data fusion:* [39, 40, 48] Estimating bike volumes from multiple proxy data sources commonly begins with data fusion. Data fusion, in the words of Miah et al., is the "concatenation of data sets with an enormous diversity in terms of information, size, and behavior" [40, p. 04021122-3]. This method varies in complexity depending on the kind of data being fused but often includes temporal alignment of biking data from multiple sources and appending of various categorical variables. Often this data fusion can be geographically weighted, and Proulx found that geographically weighted data outperforms any single dataset's prediction of volumes, provides a precise representation of traffic volumes across the network[39].

- *Segments clustering:* Clustering methods are often used to categorize segments of bicycle data [40, 31]. Since proxy data sources like Strava are biased but systematically so, modeling segments separately (or sometimes partially pooled) allows for estimates to vary by category of bicycle activity.

- *Machine learning:* [40, 43] With fused data sets, machine learning approaches have been used in several ways for estimating bike volumes. Miah et al., for example, estimated the relationship between (clustered) Strava data and daily counts, where the nonlinearity between modeling variables was captured by binary recursive partitioning models (i.e., Classification

and Regression Trees (CART)). CART repeatedly partitions the data into subsets until a state of homogeneity is reached.

- *Land use and network characteristics*: Fused datasets typically make use of many kinds of variables, including land use and roadway network characteristics. A comparison between Strava activity counts and video-based count data in around Miami-Dade County shows a moderate correlation of 0.55 [41]. Factors affecting bicycle in these models included the presence of bicycle facilities; hills; number of jobs; nearby commercial properties; population; origin-destination centrality; transit stops; presence of a median; number of lanes; and number of intersection legs or approaches [5].

Gaps and Challenges in Non-Motorized Counting

There are limitations and challenges to both non-motorized count programs and to estimating volumes with proxy data. First, using the methods described in these state and federal guidelines for non-motorized counting [programs], bicycle and pedestrian analysis requires multiple years of data collected at a variety of sites within a network. This would naturally be labor intensive and require strict maintenance schedules and regular expansion of the program. In their review, Ohlms et al., the found that ”No state had a fully complete picture of non-motorized travel, but a combination of travel survey data, short- and long-term sample-based count data, and travel models provided useful estimates” (p. 76). Given the institutional support and time required to run systematic count programs, there is a need to investigate alternative methods for estimating bicycle and pedestrian volumes.

Also, the provided approach gives limited information about how to categorize locations for the creation of factor groups (or for clustering when using proxy data). Guidance says that visual inspection or clustering of some type should be used, there is little specification to how this might be done. Lack of specificity in this area is particularly challenging, as clustering is an unsupervised machine learning task, which are highly sensitive to algorithm selection, similarity or distance measurement, and the selection of the number of clusters to form [49]. There is need to investigate

clustering of biking and walking activity more rigorously.

Clustering Applied to Non-Motorized Data

There is growing interest in applying temporal clustering methods to count time series, especially non-motorized count data; as noted above, clustering of time series helps predict future volumes or demand on non-motorized modes. Regarding bike data, Miah et al. [40] reduced crowd-sourced bike trip data in Portland, Oregon, from a phone application (i.e., Strava data [44]) to three index features: a ratio of weekend versus weekday cycling traffic (i.e., non-commute measurement), a ratio of morning to mid-day volumes (i.e., morning peak), and a ratio of self-reported utilitarian trips to total activity (i.e., commute measurement), which have been shown to be important measures of predicting activity [50]. The features were then clustered using k-means, and the resulting clusters were used to predict the relationship between Strava data and observed data through non-parametric models (also CART models). Their clustering resulted in three clusters, although it is not clear how they settled on the final number of clusters ($n = 3$). Li et al. took a different clustering approach in time series reduction in their work on clustering bike share system usage data in Chicago, Illinois [51]. They reduced their time series through discrete wavelet transforma- tion before using shape-based time series clustering methods and dynamic barycenter averaging prototype selection, with good results: their clusters ($n = 5$) captured the main characteristics of the time series and performed well based on an adapted Davies-Bouldin index (i.e., a measure of cluster compactness and separation from other clusters).

Recent work with pedestrian data has also made use of clustering methods. Singleton et al. and Humagin et al show that pedestrian push button data (i.e., actuation data) collected from traffic control logs can predict pedestrian volume predictions [33, 38]. They represented these time series (originally recorded as timestamps at the millisecond level) in six different ways, where each of the metrics calculated were a function of actuation (and other events, like the beginning of a new phase). Then they used three different time series clustering approaches – two shaped-based methods and one hierarchical method. Euclidean distance clustering performed best when clustering the

first two metrics (which were people pushing the button and the call registering the traffic control log, respectively), and a structure-based clustering method of temporal correlation on the last two metrics (which measured pedestrian detections and calls, respectively, as shares of the weekly total activity), but there was similar performance across different clustering approaches, and with some overlap between cluster assignment.

Given these applications to both bicycle and pedestrian datasets, we see opportunities for further exploration. First, it is worth understanding which clustering approaches perform best for bicycling and pedestrian data; while each of the studies mentioned here used different approaches successfully, they did not compare the same approaches applied to different kinds of non-motorized count data. Second, while Singleton et al. and Humagain et al.'s work thoroughly investigated three different clustering methods applied to six different representations of the time series, they did not test feature-based clustering, and we see the opportunity to build upon this work and compare shape- and feature-based clustering here. Regarding bike datasets, we also see the opportunity for expanding on Miah et al.'s work [40] by extracting a variety of features to represent the time series.

Summary

Non-motorized counting is a challenging task faced by transportation agencies, but understanding levels of activity within the network is crucial for understanding safety within the network. Non-motorized count programs are the standard way for quantifying biking and walking in a network, but observing biking and walking traffic at enough locations to get a sense of activity within the entire networks, or even large portions of them, is impractical and infeasible for most transportation agencies. Using proxy sources as supplications to observed counts, however, shows promise for improving volume estimates. But there is still much that is unclear about how to leverage proxy data, including how best to use proxy data for determining count site locations and factor groups. The typical process for using proxy data to create models is shown in Figure . There is room for improvement in these methods, however, including improving the granularity of estimates (not just AADNT), rigor around factor group creation, assessment, and defining "similarity" among time

series, and improving the methods used to cluster biking and walking activity captured in the proxy data sources.

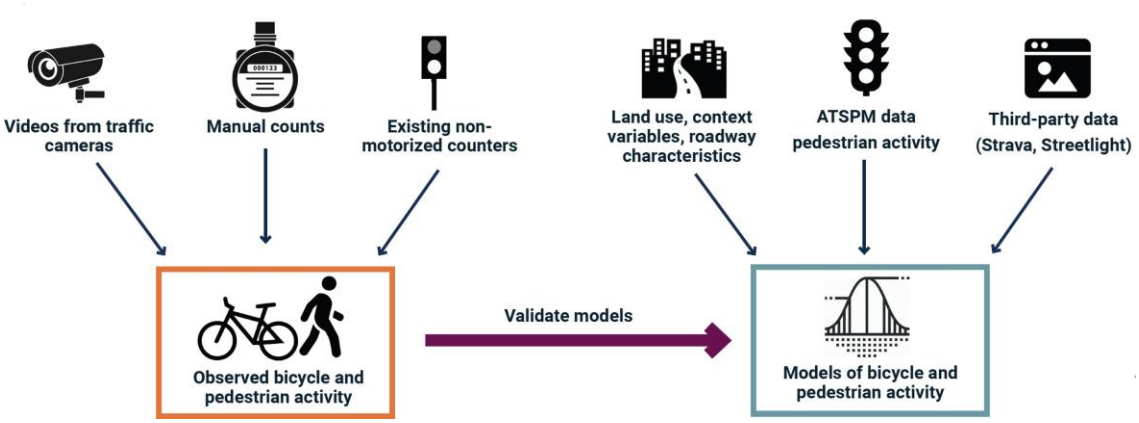


Figure 3: Illustration. General process for estimating non-motorized travel volumes with proxy data sources.

CHAPTER 2: CLUSTERING FOR BIKE AND PEDESTRIAN TYPOLOGIES

This chapter will discuss the different methods that were implemented to accomplish the goal of characterizing bike and pedestrian activity in Georgia through traffic signal data. Based on existing practices and findings from the literature, the methods involved in this study included developing site selection criteria, data-specific acquisition and processing techniques, an applicable clustering approach, and cluster characteristic assessments.

Data Description

Pedestrian Data

To strategically limit the scope of the study while still including a variety of contexts for pedestrian activity, initial site selection criteria were implemented. These criteria were based on four conditions, including (1) the presence of ATSPM signals in the location, (2) the site's proximity to Atlanta and other potential study sites, (3) preliminary equity analyses, and (4) intersection density. The following subsections will discuss each selection criterion as well as how each was used in the site selection process.

The first consideration limiting potential intersections of study was the presence of ATSPM signals in the location. This was because to conduct the proposed study, intersections were required to have signals with ATSPM systems logging relevant pedestrian metrics. While Georgia has implemented ATSPM systems for at a total of 7,034 signals, there are still locations that do not have the software in place and thus may not have the high-resolution data loggers available to collect applicable pedestrian traffic data.

Figure below demonstrates the distribution of signals with ATSPM around the state. From this map, it was observed that signals with ATSPM tended to be located around cities or areas of higher population density, but only sporadically in more rural areas. While rural locations still make the

use of traffic signals at intersections, intersections without the ATSPM data loggers installed could not record the types of data required for this analysis. Additionally, stop-controlled intersections or uncontrolled intersections which can often be found in rural areas with low automobile activity, also could not be considered due to their lack of push buttons and data loggers.

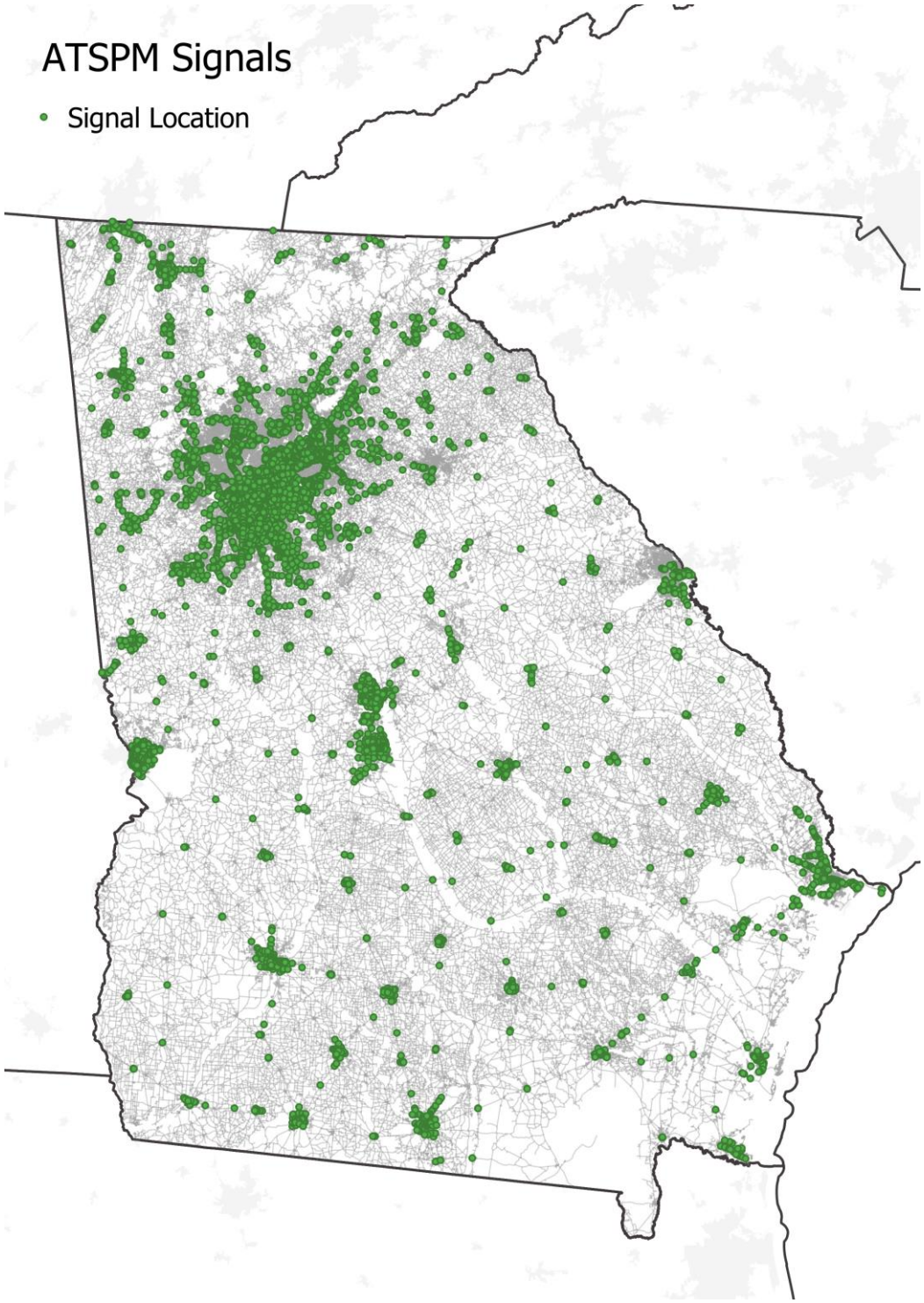


Figure 4: Map. Location of all ATSPM signals throughout the state of Georgia.

Beyond the installation of ATSPM software, a second consideration for a site's inclusion within the study was the location's proximity to Atlanta, as well as other selected study locations. Although the goal of this study was to create a state-wide system for evaluating pedestrian activity, preliminary conditions for site selection restricted locations to be within Interstate 285 if they were selected from Atlanta, or in close proximity to other potential study sites if they were selected outside of Atlanta. For sites in Atlanta, Interstate 285, also known as the Perimeter, was selected as the primary study boundary since it encompasses the greater part of the City of Atlanta and serves as a pre-existing reference for the limits of urban activity.

Other considerations for limiting the study to be within and around Interstate 285 were for the project team's feasibility of travel and understanding of local contexts. Since the project team was based in Midtown, Atlanta, and a secondary goal of this research was to use its results to develop activity estimation models for future studies, it was important that selected sites be within a reasonable driving distance from the City of Atlanta as well as from each other. Thus, in the event that the project team would need to revisit sites after data collection was conducted, it would not be unreasonable to accomplish. Selecting sites near the City of Atlanta also provided the project team with the potential of having personally visited potential study sites in the past, which would give them a prior understanding of the location's context.

In addition, to vary the types of locations selected and thus potentially the types of pedestrian activity observed, preliminary factors that were considered on top of presence of ATSPM signals included the presence of universities in the municipality, urban form (dense versus sprawling), population levels, and general pedestrian facility characteristics.

The third level of criteria included in the site selection process was a preliminary evaluation of equity. Using the Gini coefficient from the American Community Survey (ACS) Gini Index, a high-level evaluation of income inequality, income was used as a proxy for selecting locations with a variety of equity groups. By having multiple levels of income dispersion present in selected sites, different groups of people and types of pedestrian activity would ideally be represented in the site selection and overall modeling process. The ACS Gini Index was ultimately used because data

was easily accessible and provided information at a statewide level, rather than for only specific regions. Figure provides a geographic representation of the Gini Index throughout the state of Georgia by census tract, with a scale of low, medium, and high levels of income inequality.

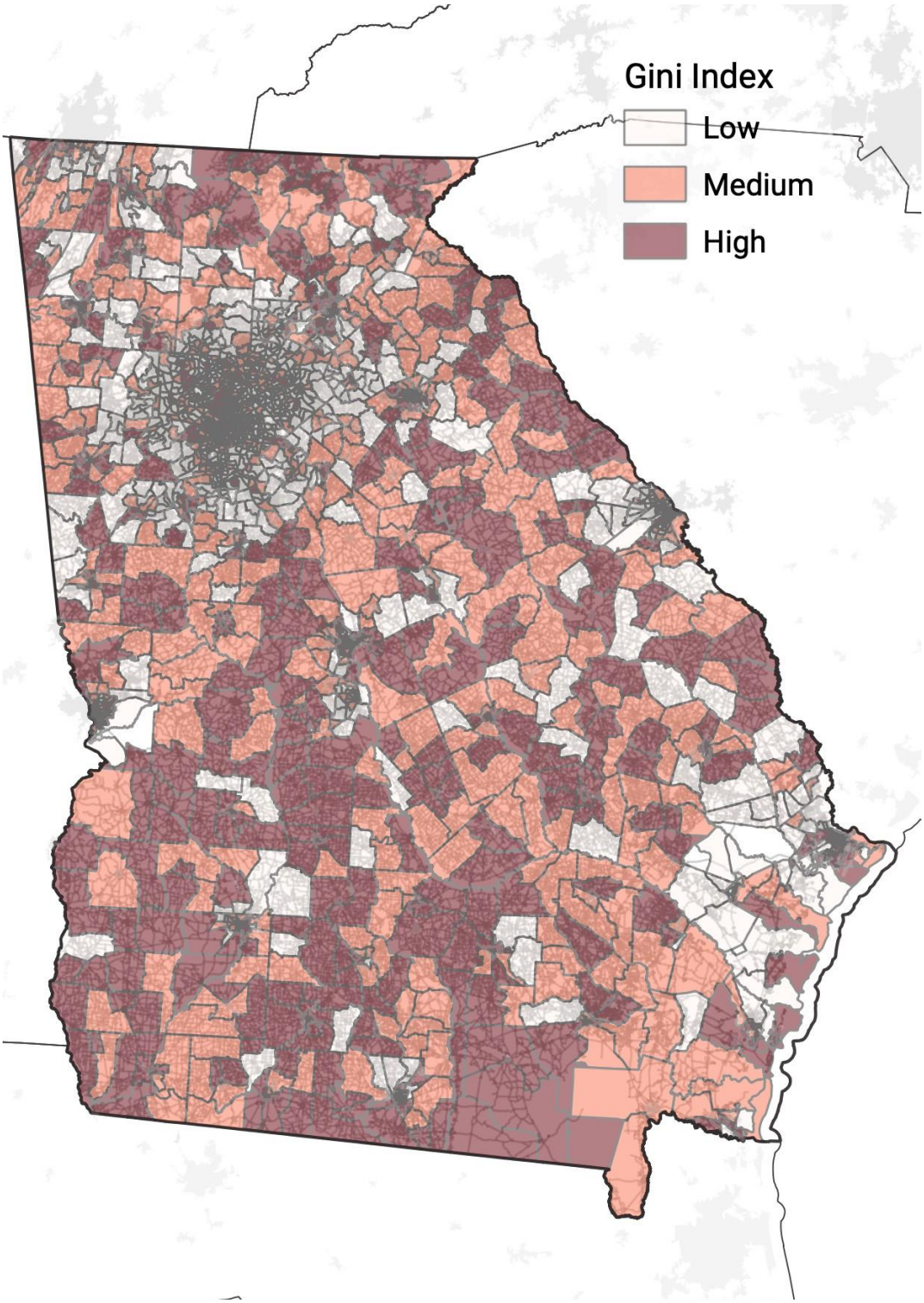


Figure 5: Map. ACS Gini Index throughout the state of Georgia

A final selection criterion for sites to include within the study was level of intersection density. Beyond purely having ATSPM installed within a signal, a certain level of anticipated pedestrian activity was needed to effectively model pedestrian volume, as well as validate funds spent in the future to collect ground truth data. Since previous studies have shown that higher levels of intersection density within an area indicate greater potential pedestrian activity [52], intersection density was deemed to be a suitable indicator of predicted pedestrian activity.

To evaluate intersection density in the Georgia context, the number of intersections per square mile throughout the state was calculated and modeled using Geographic Information System (GIS) through the open-source Quantum Geographic Information System (QGIS) software. High intersection density levels were considered to have between 210-660 signals per square mile, medium density levels between 34–210 signals per square mile, and low density levels 0–34 signals per square mile. These bins of different density levels were determined using Jenks natural breaks optimization, which is a method for categorizing values into different groups by minimizing variance from the group mean but maximizing the variance between other group means). By overlaying this visual with the location of ATSPM signals shown previously in Figure , areas having signals with the highest levels of intersection density could be identified. The distribution of ATSPM signals in Georgia specified by high, medium, and low intersection density is shown in Figure below.

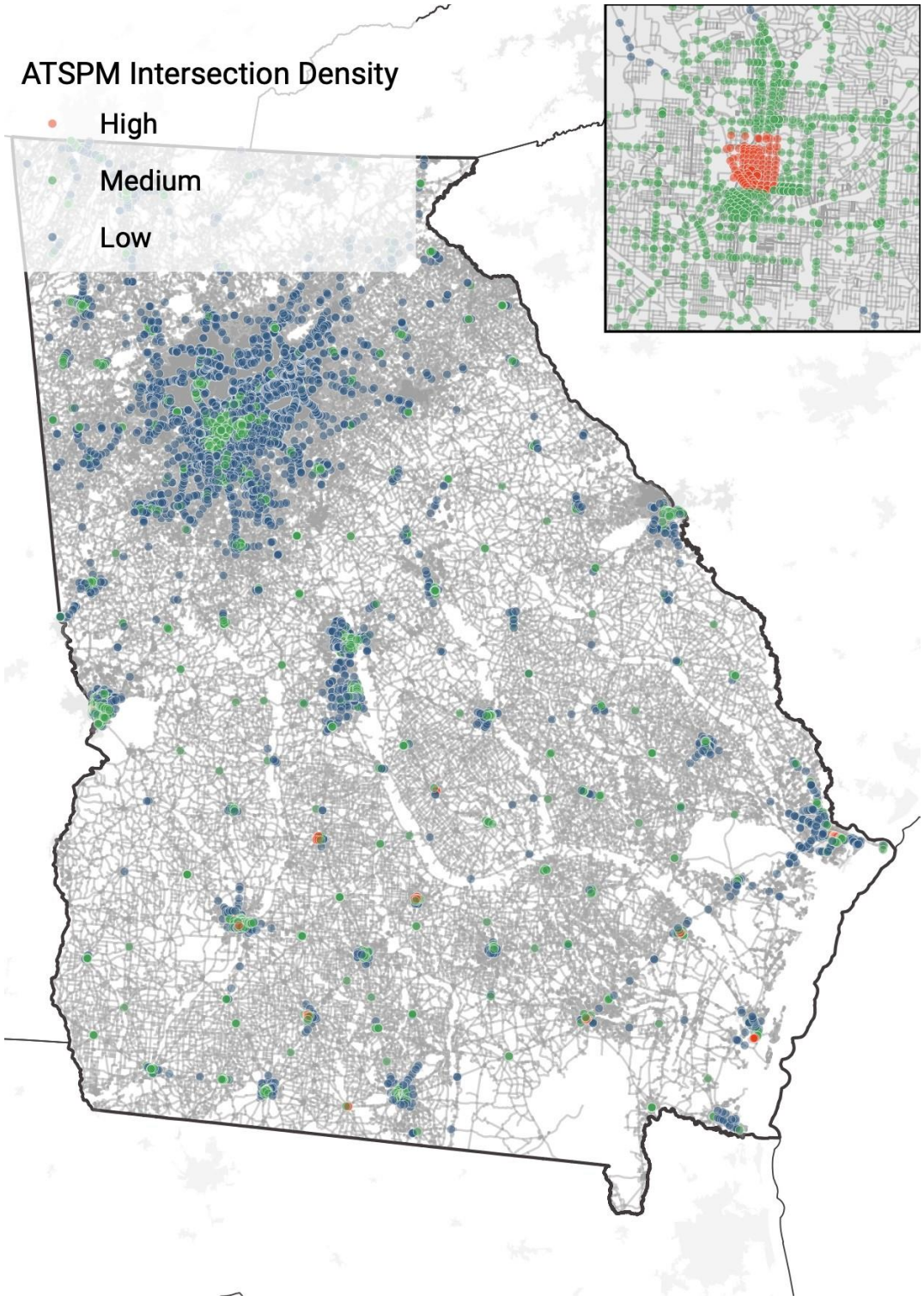


Figure 6: Map. Distribution of ATSPM signals in Georgia specified by high, medium, and low intersection density

Based on these initial selection criteria, a total of 1,799 signals were identified from seven different municipalities, including Atlanta (within the bounds of Interstate 285), Macon, Warner Robins, Milledgeville, Griffin, Rome, and Dalton. Figure shows the geographic location of each municipality, and Figure shows the number of ATSPM signals in each municipality compared to the state-wide ATSPM database.

From these maps, the first two selection criteria, which included the presence of ATSPM-equipped signals at the location and proximity to Atlanta, are noticeable. Generally, two groups of sites outside of Atlanta were developed, with the first including Griffin, Milledgeville, Macon, and Warner Robins, and the second including Dalton and Rome. Both groups at their extents are within a two-hour driving distance (less than 100 miles) from Midtown, Atlanta, and are approx- imately within a one-hour driving distance (about 50 miles) from each other. Overall, the various municipalities selected provided a variety of different pedestrian activity contexts, including some areas that were university-based, had more dense urban form, more sprawling urban form, smaller population levels and larger population levels.

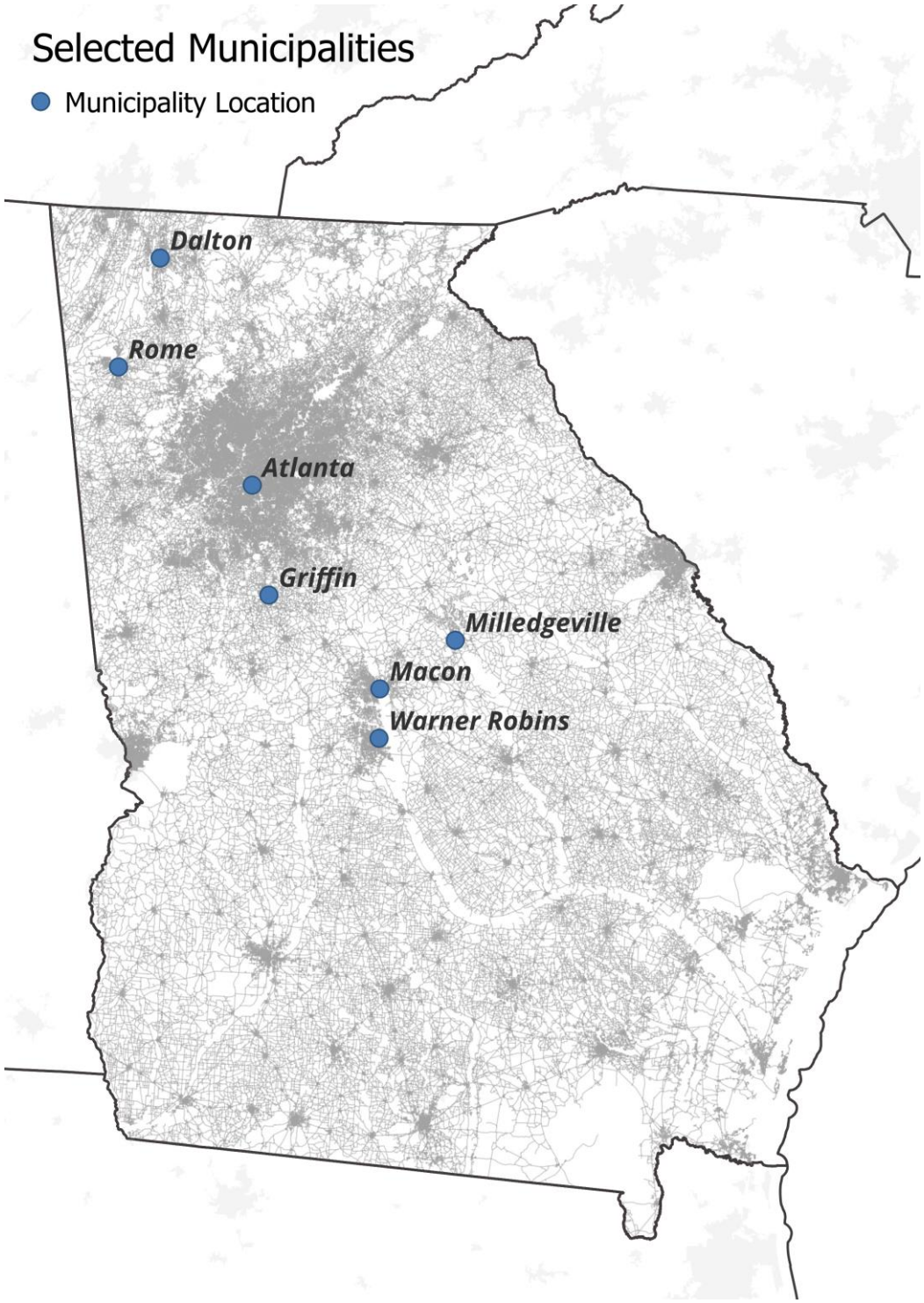


Figure 7: Map. Location of selected municipalities throughout Georgia.

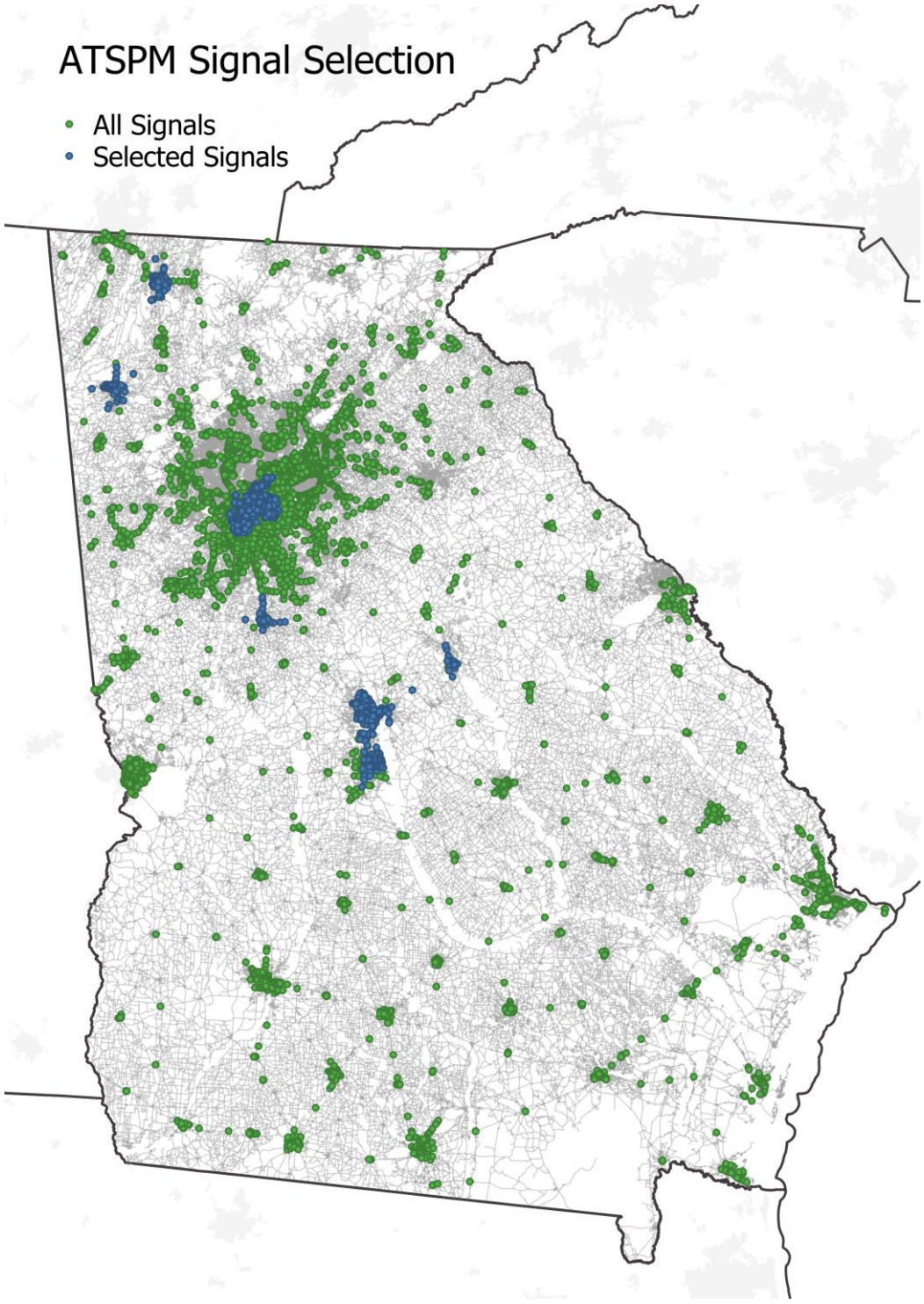


Figure 8: Map. Visual representation of selected ATSPM signals versus all ATSPM signals in Georgia.

Table outlines each municipality, the number of ATSPM signals from each, and the number of

signals as a percentage of the total. It should be noted that while the level of intersection density was a selection factor considered for all municipalities, all signals were kept regardless of whether they were located in intersection-dense areas or not for all municipalities except Atlanta. Only signals located in medium to high intersection density segments were selected within Interstate 285 to help limit the number of signals included in the initial site selection process.

Overall, the greatest share of signals was located within Atlanta, having 65% of the total number of signals. This was followed by Macon, which had a share of 14%. The remaining municipalities all had between 2% to 6% of the signal share, which was likely correlated to their smaller occupied area when compared to Macon and Atlanta.

Table 1: Initial selected study locations and associated number of ATSPM signals

Municipality	Number of ATSPM Signals	Percentage of Total (%)
Atlanta	1,177	65
Macon	249	14
Warner Robins	84	5
Milledgeville	42	2
Griffin	61	3
Rome	77	4
Dalton	109	6
Total	1,799	100

Prior to beginning the clustering analysis, several steps were required to acquire and process the approach time series data. This section will describe the data types utilized in this study, the preliminary evaluation of the acquired dataset, final compilation steps, and the pedestrian activity metric calculation process.

Six months of time series traffic signal data, including all events from all 7,034 ATSPM-equipped signals, were provided by GDOT through an external hard drive. Date ranges for these data began January 1, 2022, and went through June 30, 2022. The dataset was originally stored in the Apache Parquet file format, which uses column-oriented storage, but was converted to Comma-Separated Values (CSV) files to facilitate analysis. The data itself included every registered event code by each signal during the six-month time frame, the timestamp for each event, as well as the

event's associated phase parameter.

Although all signal event codes were included in the original dataset, only pedestrian-related events were relevant to this analysis. Table below highlights common pedestrian event codes, their labels, and descriptions. Specifically, event 45, "Pedestrian Call Registered", which is the event logged after the push button is initially pressed during a phase and event 90, "PedDetector On", which is the logging of the actual push button press are of particular interest when it comes to evaluating pedestrian activity. However, it should be noted that for a signal in recall, which is a special case when the push button does not need to be pressed for a pedestrian call to be registered, event 45 will not occur.

Table 2: Common pedestrian signal event codes and descriptions (Sturdevant et al., 2012)

Event Code	Signal Event	Event Description
0	Phase On	Event occurs at activation of phase, either at the beginning of a green phase or walk interval.
21	Pedestrian Begin Walk	Event occurs at the activation of the "Walk" indicator.
22	Pedestrian Begin Clearance	Event occurs when the "Don't Walk" indicator begins flashing.
23	Pedestrian Begin Solid "Don't Walk"	Event occurs when the "Don't Walk" indicator turns solid, either at the end of a pedestrian clearance interval or after a pedestrian dark interval (when pedestrian indicator is off).
45	Pedestrian Call Registered	Event occurs when a call to service is registered during a phase by pedestrian demand. This event will not occur during a phase set to recall.
89/90	PedDetector Off/PedDetector On	Event occurs when the signal from a pedestrian push button is deactivated or activated, or after any detector delay or extension processing. Pedestrian detector events may occur mulitple times for a single pedestrian call.

To evaluate the nature of the Georgia DOT ATSPM data set, a preliminary week of time series data from April 24, 2022, to April 30, 2022 was assessed for the selected set of 1,799 signals. These dates were chosen under the assumption that they would represent periods of relatively regular

pedestrian activity at the selected sites, since the month of April in Georgia is typically temperate with respect to weather and lies within the general university academic calendar. Initially, only data for event codes 0, 45, and 90 were extracted for each signal. Signals without any event 45s or event 90s during the specific week were eliminated from the selection process. From this one week of data, anomalies in the time series were assessed, and further cuts were made to the number of signals considered in the study based on quartiles of activity.

Data Anomalies

After assessing the time series data for just one week at each selected signal, several data anomalies either led to the exclusion of certain signals, or the need to clean signal time series. With respect to signal exclusion, signals that were clearly malfunctioning were removed from consideration in the study. This included signals with a number of actuations that were considered to not be possible within the span of a week, such as signals with more actuations than there were seconds in a week, as well as signals with long strings of event 45s unaccompanied by event 90s with no detectable pattern. A total of 28 signals were removed during this process.

A number of signals in the dataset, however, were found to have certain recurring patterns of event 45s unaccompanied by event 90s. After meeting with Georgia DOT traffic engineers, it was confirmed that patterns of event 45s could be triggered by event 150, "Coordinated Cycle State Change", and occasionally by event 184, "Power Restored". Event 150 occurs when a signal goes from freely operating into coordination with other signals, and event 184 occurs when a signal restarts from a flash phase, typically as a result of a power outage. The number of extra event 45s that were triggered by these two event codes varied between one to four, and a script was written to remove these additional events.

Table demonstrates an example of event 45 patterns occurring as a result of an event 150 for signal 2131 located at the intersection of North Columbia Street and Hammock Road Northwest in Milledgeville, Georgia. As shown in the table, the timestamps for each event code are exactly the same between event 0, 45 and 150. With respect to the sequence of events, at exactly 7:30 AM, the signal phase changes and the signal switches from free to coordinated, throwing an additional

two event 45s unaccompanied by event 90s.

Table 3: Example of time series data that demonstrates a case of event 45 patterns being triggered by an event 150.

Signal ID	Date	Time (in milliseconds)	Time (in UTC)	Event Code	Event Parameter
2131	4/24/2022	27020500	4/24/2022 7:30	0	2
2131	4/24/2022	27020500	4/24/2022 7:30	0	6
2131	4/24/2022	27020500	4/24/2022 7:30	45	2
2131	4/24/2022	27020500	4/24/2022 7:30	45	6
2131	4/24/2022	27020500	4/24/2022 7:30	150	1

Quartile Testing

After removing malfunctioning signals and correcting for data anomalies, additional cuts to the number of considered signals needed to be made to stay within the limits of available computational power. Thus, signals were selected based on minimum thresholds of activity. However, rather than arbitrarily selecting a minimum number of event 45s or 90s as a point of reference, data was assessed using quartiles. Table and Table outline the minimum and maximum weekly totals of event 45s and 90s, respectively, for each quartile.

Table 4: Quartile thresholds for weekly event 45 totals between April 24, 2022 and April 30, 2022.

Location	Maximum Quartile Values (#45s)				
	0	1	2	3	4
Milledgeville	2.0	33.8	164	932	3,930
Griffin	2.0	32.4	62.5	138	379
Warner Robins	1.0	32.5	58.5	97.5	174
Macon	1.0	12.8	25.0	59.0	7,100
Atlanta	1.0	41.8	148	476	118,000
Rome	1.0	32.8	96.5	179	7,650
Dalton	1.0	15.0	54.0	185	7,920

Table 5: Quartile thresholds for weekly event 90 totals between April 24, 2022 and April 30, 2022.

Location	Maximum Quartile Values (#90s)				
	0	1	2	3	4
Milledgeville	2.0	89.0	429	2,810	14,200
Griffin	7.0	89.0	211	764	3,230
Warner Robins	1.0	77.0	219	435	1,780
Macon	1.0	24.0	63	197	24,300
Atlanta	1.0	213	719	2,270	268,000
Rome	1.0	157	418	871	23,500
Dalton	1.0	49.5	212	736	17,100

Ultimately, it was decided that all higher activity signals falling within quartiles three and four would be kept within the dataset, while only 25% of signals in quartiles one and two would be kept through random selection. This percentage of lower activity signals in quartiles one and two was selected to help reduce the set signals that had extremely low levels of pedestrian activity, but still keep some as low activity signals are still meaningful to the analysis. Additionally, fewer low activity signals would ideally prevent data collection endeavors of future studies from collecting little to no pedestrian counts during the collection period.

An additional consideration for this process was whether quartile values for event 45 or event 90 should be used to make signal cuts. Since signals with pedestrian recall do not log event 45s, it was determined that quartile thresholds for event 90 should be used to prevent recall signals from being eliminated from the analysis. Thus, after retaining all signals falling in the upper quartiles and randomly selecting 25% of signals falling under the first two quartiles of event 90s, 738 signals remained.

Data Compilation

To perform the future clustering process, it was prudent to use as much data as possible to assess different pedestrian activity trends over time. However, due to constraints from available computing power and the large number of signals still remaining within the study, it was not feasible to use all six months' worth of data. Therefore, three months of data were pulled from the original six-month dataset with dates ranging from April 1, 2022, to June 30, 2022. By using

the second quarter of the year rather than the first, it was assumed that the data would be more representative of higher pedestrian activity levels than the colder, winter months at the beginning of the year.

Time series data was pulled from the external hard drive in batches to avoid exceeding computer memory. To accomplish this, only half of the 738 signals were extracted at a time, and each month of data was pulled individually resulting in six time series files. After combining each half of each month once loaded into R Studio, each time series file was then combined chronologically.

By calculating PAMs 1 and 2, initial magnitudes of the most relevant pedestrian actuation events, including event 90 and event 45, could be assessed for a specific unit of time. In the case of this study, PAMs were assessed for all phases, rather than a single phase, and the selected unit of time was hourly. PAM 3 and PAM 4 normalized the previous PAMs by dividing the initial magnitudes of pedestrian detections (event 90) and pedestrian calls (event 45) by the number of phase starts (event 0) within a time unit *t*. This accounts for different signals having different cycle lengths, which would contribute to the number of cycles that are able to occur during a specified period and thus the number of pedestrian event codes. Specifically, event code 45 is influenced by cycle length, as only one event code 45 can be registered per phase.

Each of the four PAMs was calculated for every hour of the three-month time series for each signal. They are summarized in Table . Additionally, each PAM was also calculated as a three-month weekly average for each signal to assess the differences in clustering results. It should be noted that signals with pedestrian recall did not have PAM 2 or 4 due to their lack of coded 45 events.

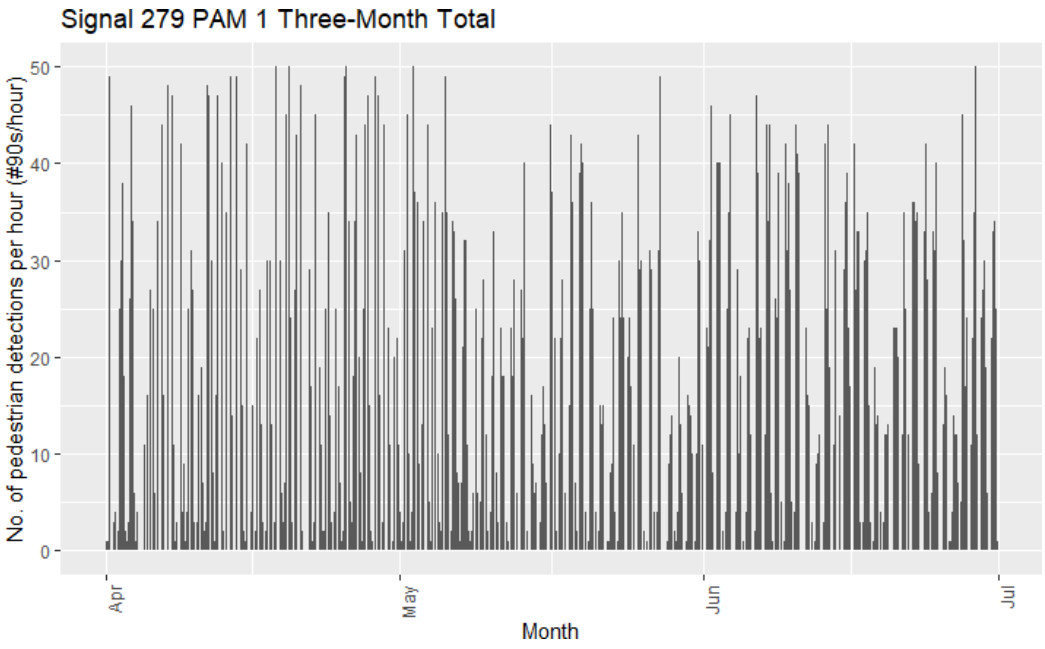


Figure 9: Graph. Example distribution of PAM 1 for full three-month period.

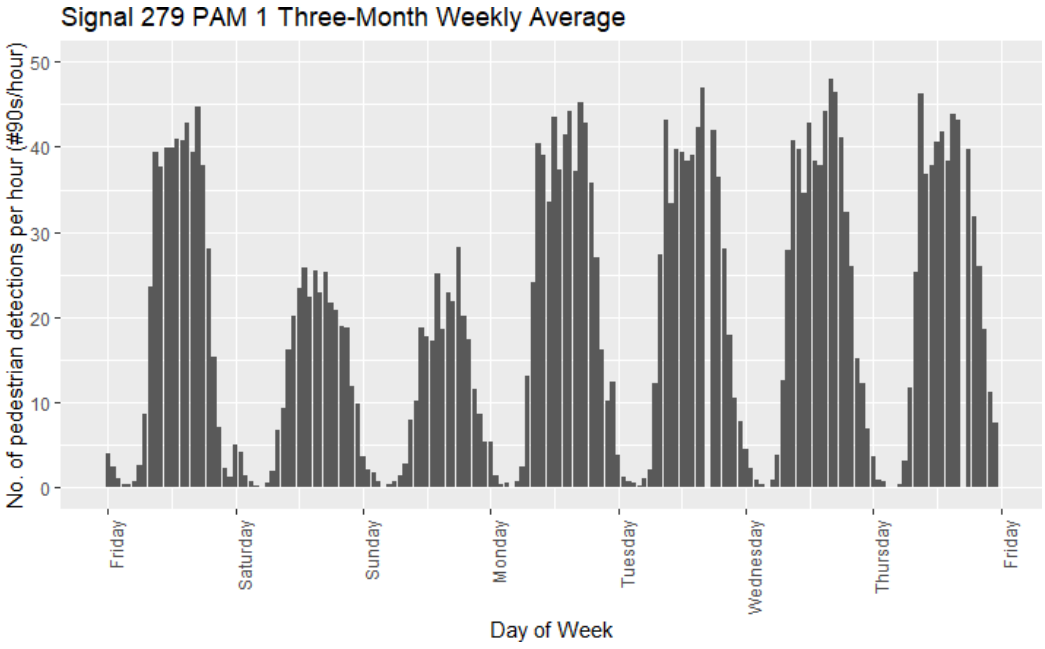


Figure 10: Graph. Example distribution of PAM 1 for the three-month weekly average.

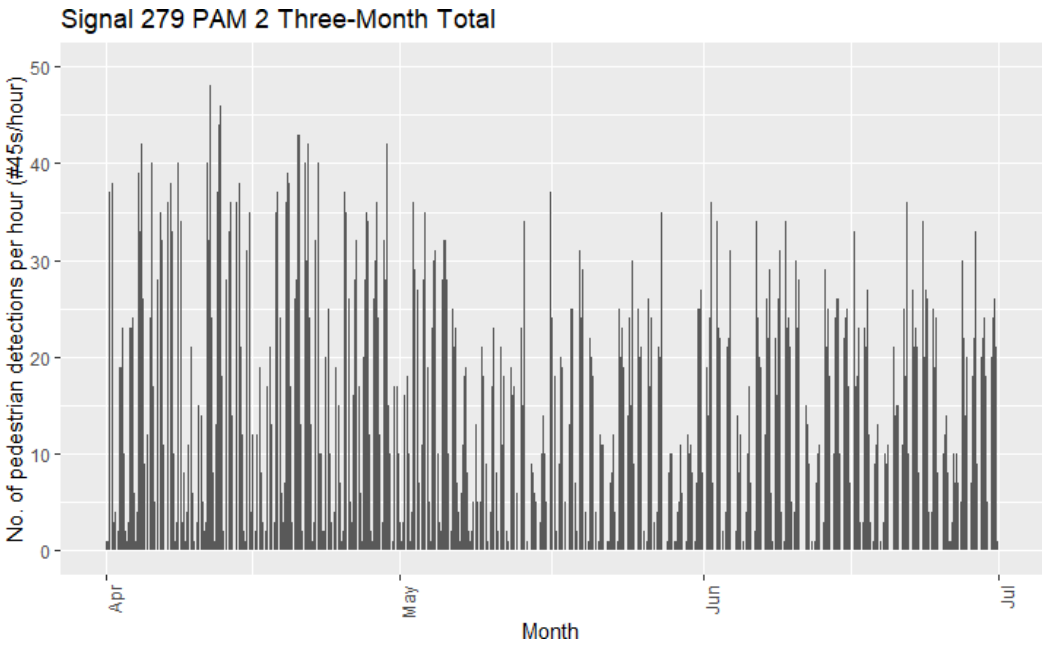


Figure 11: Graph. Example distribution of PAM 2 for full three-month period.

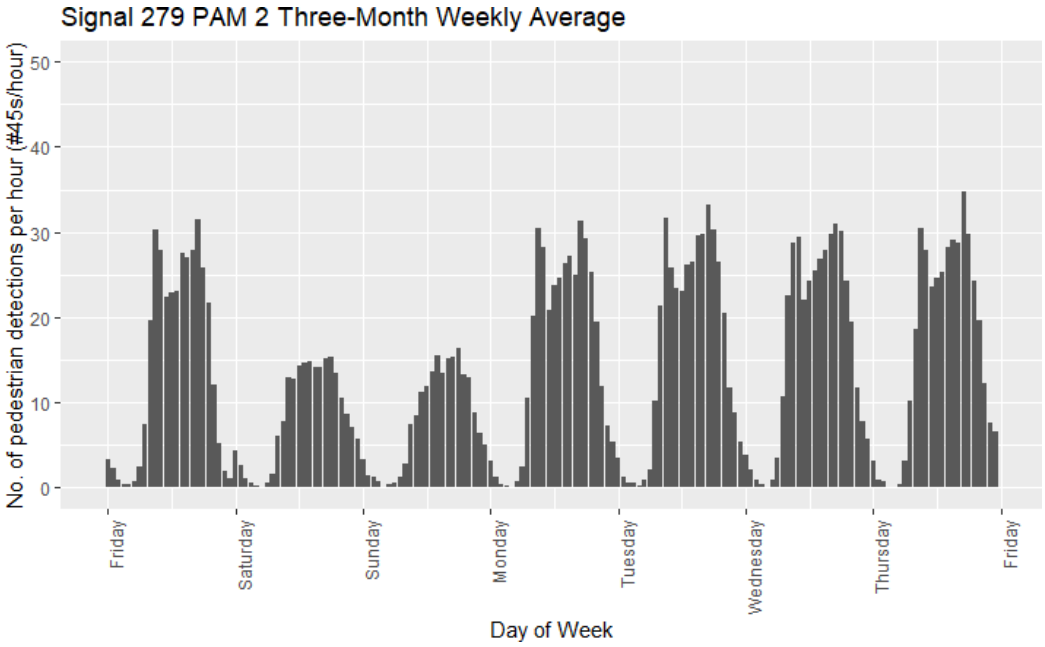


Figure 12: Graph. Example distribution of PAM 2 for the three-month weekly average.

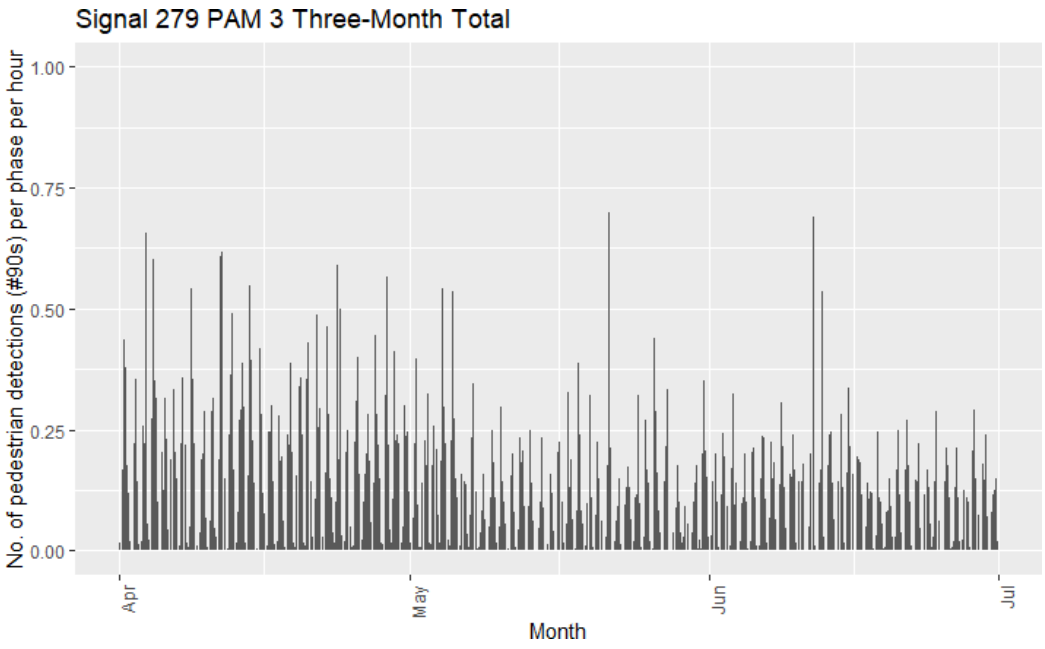


Figure 13: Graph. Example distribution of PAM 3 for full three-month period.

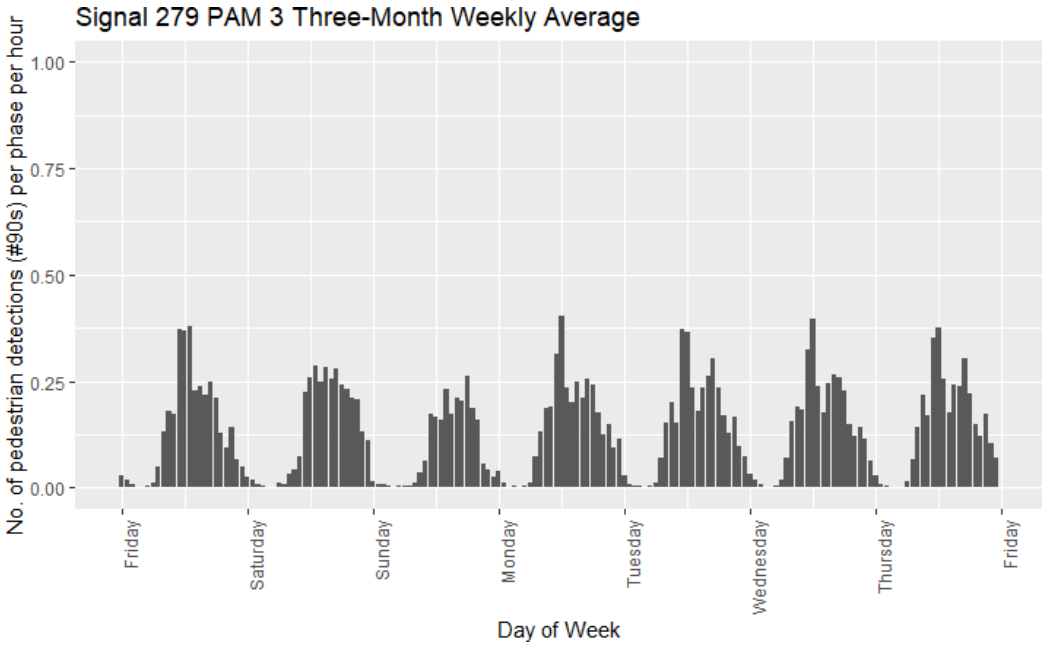


Figure 14: Graph. Example distribution of PAM 3 for the three-month weekly average.

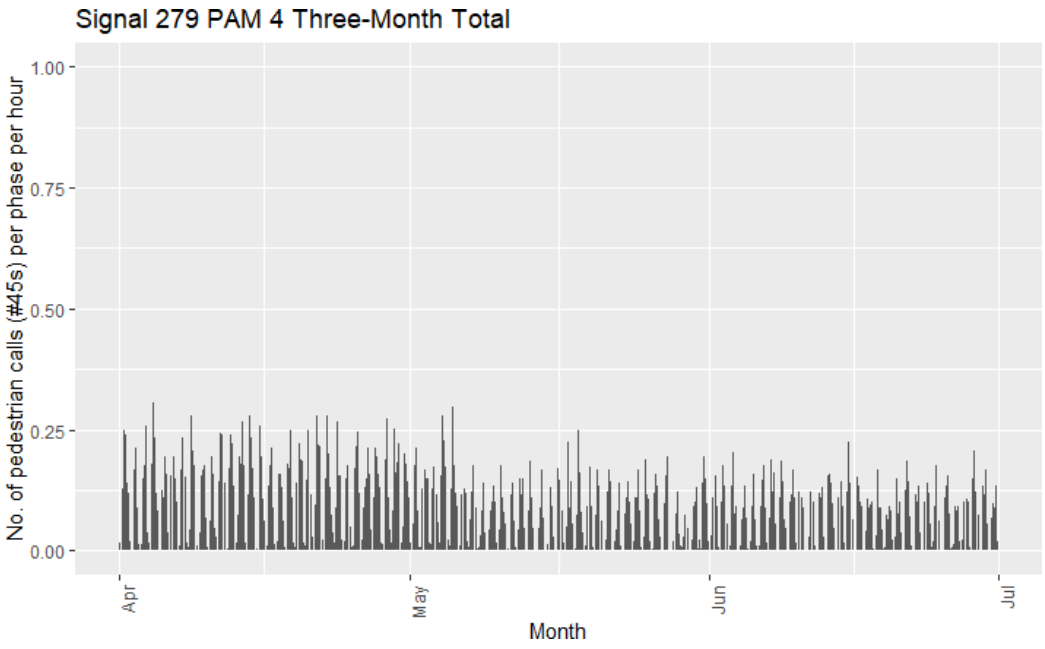


Figure 15: Graph. Example distribution of PAM 4 for full three-month period.

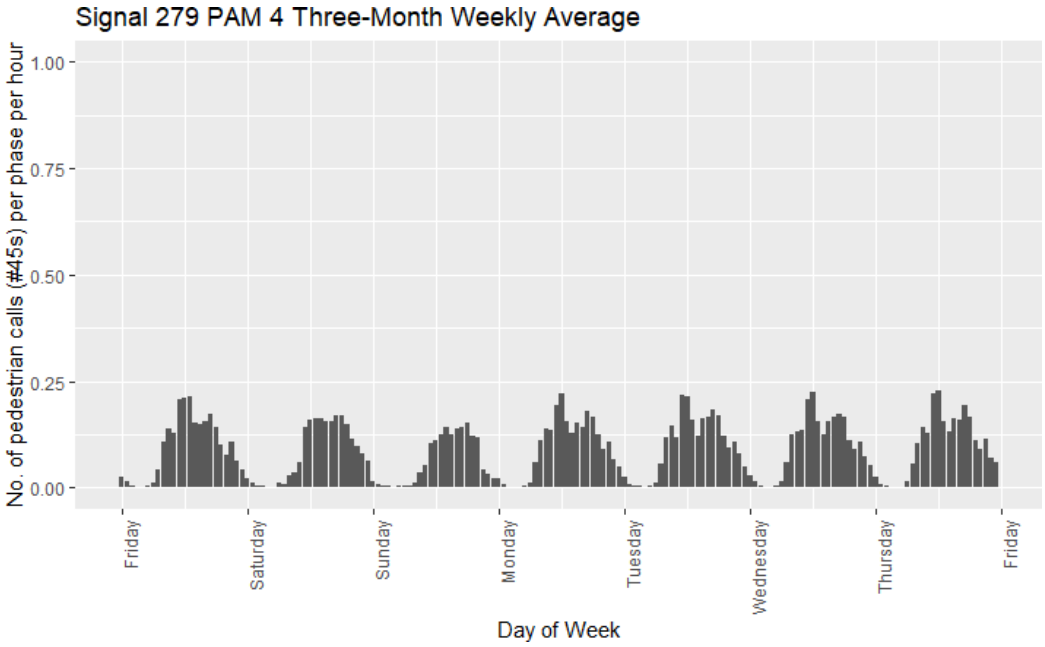


Figure 16: Graph. Example distribution of PAM 4 for the three-month weekly average.

For further analysis, it was decided to use the more high resolution data – the full three-month period – for comparing clusterings across different methods.

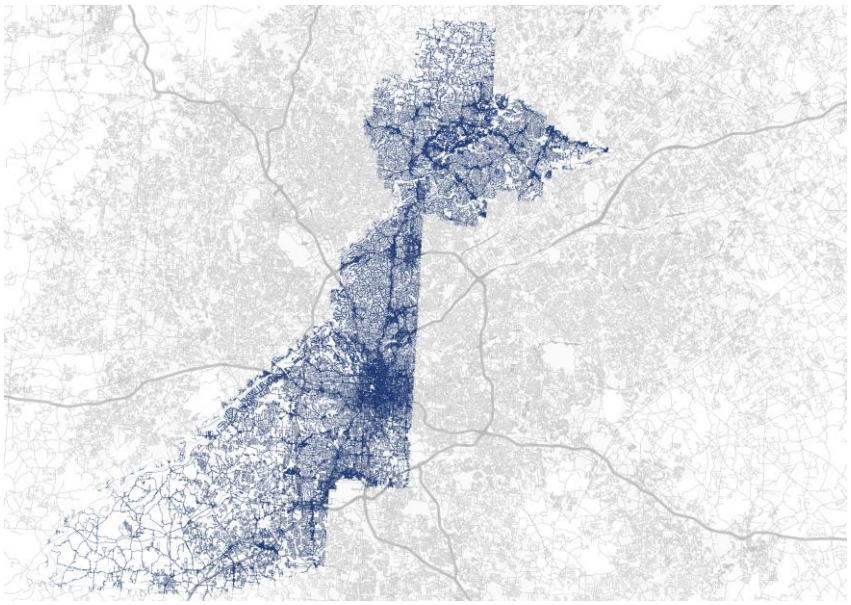
Bicycle Data

In addition to the pedestrian data described in the previous section, clustering method performance on bicycle data were also assessed. Strava data is used as the proxy data source. Strava [44] is a smart phone application and social media platform that people use to track human-powered recreational and commute trips, including bike trips. Strava Metro [45] is the research and data division of Strava, which provides researchers and public agencies with high-resolution data on AT. These data contain segment-level hourly volumes for an entire roadway network – a notable advantage over other third-party and sparsely collected datasets, which often can only generate area-level estimates of bike activity.

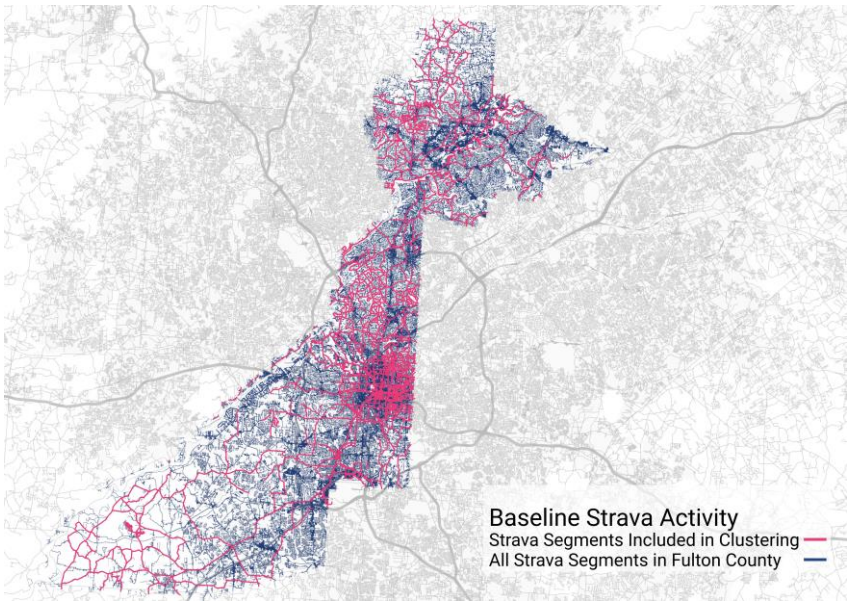
There are two main challenges of using Strava data: First, it does not capture all bike trips; it is inherently incomplete given that it is self-selected, and the data are biased towards young, white males [46, 47]. Miah et al. note that Strava represents somewhere between 2% and 10% of all bicycle traffic at some locations, but others – likely those that are more recreational facilities – may see much higher shares, sometimes up to 30%. This bias, however, appears to be systematic [40], and Strava data has been shown to produce good-quality estimates of average biking rates in certain contexts [42, 31, 43]. Second, Strava data is rounded to the nearest five riders for privacy purposes, which eliminates some of the information in the already self-selected data.

For this work, only Strava segments in Fulton County (Georgia) are considered for analysis, excluding segments that result in dead ends (e.g., segments that lead to parking lots or single family homes). This results in over 41,000 segments. A moderate level of activity is needed to perform clustering, so we considered two iterations of excluding data across two units of time. The four Bike Activity Metrics (BAMs) we use in this work are a combination of these two factors. The first metric is the number of bicyclists reported per hour for each segment (cutting about 50% of the data) in the dataset (BAM 1), and the second is the number of bicyclists reported per day for each segment (BAM 2). We considered these same temporal metrics to a trimmed dataset as well, where we only included segments that had 10 rides per week recorded across all three months (BAM 3 and BAM 4). This limited the dataset to 1,359 segments, only 3% of the data. We also

considered two additional metrics (one at the hourly level and one at the daily level) that reported bike activity as a ratio of commute-flagged trips to all other trips, but ultimately there were too few commute trips for in the dataset (relative to the number of segments and other trips) to be a meaningful measure, so we ultimately did not include these metrics in this work.



(a) All segments in Fulton County



(b) Segments included (red) that included at least one ride over three months

Figure 17: Maps. Strava segments in BAM 1 and BAM 2 are shown in red, where all segments in Fulton County are shown in blue. Only segments in midtown were included in BAM 3 and BAM 4.

Activity Metrics Summary

All activity metrics used in this analysis are shown in Table .

Table 6: Metrics of pedestrian and bike activity, where pedestrian metrics are labeled with *P* and bike metrics are labeled *B*

Metric	Equation	Definition
PAM 1	$\#90_t$	Number of pedestrian detections per hour, *t*
PAM 2	$\#45_t$	Number of pedestrian calls registered per hour, *t*
PAM 3	$\#90_t \div \#0_t$	Number of pedestrian detections per hour, *t*, per phase
PAM 4	$\#45_t \div \#0_t$	Number of pedestrian calls registered per hour, *t*, per phase
BAM 1	$\#B_{t1}$	Number of self-reported bicyclists reported per hour
BAM 2	$\#B_{t24}$	Number of self-reported bicyclists reported per 24 hours
BAM 3	$\#B_{t1,Cut300}$	Number of self-reported bicyclists reported per hour on data trimmed to have at least 300 recorded rides per three months
BAM 4	$\#B_{t1,Cut300}$	Number of self-reported bicyclists reported per 24 hours on data trimmed to have at least 10 recorded rides per week for three months

CLUSTERING ANALYSIS

Given these two sets of data and four metrics for each set, we then assessed the performance of three different methods of time series clustering: two shape-based clustering methods, (1) Euclidean Distance (ED) and (2) Dynamic Time Warping (DTW) (3) one feature-based clustering method, listed in Table . Given one distribution of activity $X = (x_0 \ldots, x_n)$ and another distribution of activity $Y = (y_0 \ldots, y_n)$ over time period T which contains n observations during the time period, where $d(x_i, y_i)^2 = (x_i - y_i)^2$, Table summarizes the similarity metrics for each clustering method used here. Also, for the DTW approach, π is a list of indexed pairs (i_k, j_k), with $0 <= i_k < n$ and $0 <= j_k < m$, and these indexed pairs begin with (i_0, j_0) and end with (i_{n-1}, j_{m-1}). We chose a four-hour warping window [53, 54]. All data were z-standardized prior to conducting the clustering analysis, which enables the clustering algorithms to better compare the different time series by giving them similar magnitudes [49]. Two different clustering algorithms were used; for the shape-based approaches, we used the "partitioning around medoids algorithm,"

which begins by randomly selecting initial representative objects and then iteratively replacing representative objects by other objects until the quality of the clustering cannot be improved [55].

Table 7: Summary of clustering methods applied to each dataset

Method	Approach	Similarity	Equation	Representation	Prototype	Algorithm
1	Shape	Euclidean distance	$ED(X,Y) = \sqrt{\sum_{(i=1)}^{T} d(x_i, y_i)^2}$	Z-standardized	Medoid	Partitioning around medoids
2	Shape	Dynamic time warping, 4-hour window	$DTW(X,Y) = min_{\pi} \sqrt{\sum_{(i,j)\in\pi}^{T} d(x_i, y_i)^2}$	Z-standardized	Medoid	Partitioning around medoids
3	Feature	Euclidean distance	$ED(X,Y) = \sqrt{\sum_{(i=1)}^{T} d(x_i, y_i)^2}$	Z-standardized	Centroid	K-means

As noted in the literature review, feature-based clustering reduces the time series to a set of metrics that describe them, and simpler clustering mechanisms are applied to this reduced data. There are hundreds of possible features that can be extracted from a time series. Ultimately, the features that were chosen to represent the time series in this work were those that are common in traffic trend assessment. Seasonal and Trend decomposition using Loess features [56], like trend strength, seasonality, and autocorrelation factors, were also explored, but the features did not lead to high-quality clusterings for these data, likely because the features are designed for continuous data, not count data. The features selected for analysis are shown in Table . In the table, W is the number of weeks in the time series; (t_i, v_i) represents the timestamp t_i and observed value v_i for the i-th data point; commute trips are v_{ct} and n is the number of non-zero observations in the dataset. Also, note that the "weekly commute share" feature was only calculated for the bicycle proxy data [44], as this information was not available in the pedestrian proxy dataset.

Cluster Evaluation

The final step of applying clustering approaches to count data is evaluating the quality and number of clusters. Without having a labeled dataset as a reference, evaluating the validity of clustering results is difficult; external measures of validity cannot be used because this data is unlabeled. To assess the quality of the clusters created, to determine the optimal number of clusters, and to

determine which methods create the most cohesive clusters for AT monitoring, several cluster Cluster Validity Indices (CVIs) were measured for each clustering.

Table 8: Features extracted from each time series for clustering.

Feature	Calculation
Average of Weekly Maximum Values	$\frac{1}{W}\sum_{i=1}^{W} \max(v_{n_{i-1}+1}, v_{n_{i-1}+2}, \ldots, v_{n_i})$
Average of Weekly Average Values	$\frac{1}{W}\sum_{i=1}^{W}\frac{1}{7}\sum_{j=1}^{7} v_{i-1} \cdot 7 + j$
Weekday Commute Share	$\frac{\sum_{w=1}^{W}\sum_{t\in \text{Weekdays}_w} v_{ct}}{\sum_{w=1}^{W} \lvert\text{Weekdays}_w\rvert}$
Weekend Daily Max	$\max_{w=1}^{W} \max_{t\in \text{Weekdays}_w} v_i$
Weekday Daily Max	$\max_{w=1}^{W} \max_{t\in \text{Weekends}_w} v_i$
Squared Covariance of Non-Zero Values	$\in \sqrt{} = 0, \frac{1}{n}\sum_{i=1}^{n}(v_i - \bar{v})^2$

We used four internal CVIs (shown in Table): the Silhouette Index (SI), the Sum of Squared Error (SSE), the Dunn Index (DI), and the Davies-Bouldin index (DBI). SI measures cluster cohesion, or how similar time series objects are to their own cluster compared to other clusters [57, 49], based on average distance a_i from an object in the same cluster and the average distance b_i from objects in a different cluster, minimized over cluster separation $max(a_i, b_i)$. SI was calculated for each time series object and then averaged over all SI values in the data. SI ranges from -1 to 1, where negative values indicate incorrect cluster assignment, 0 indicates closeness to other clusters (i.e., poor clustering separation), and 1 indicates the objects fit within their own cluster well and poorly within another cluster (i.e, high-quality clustering through separated clusters).

SSE is a measure of clustering compactness based on within-cluster variance. Compact clusters are those that are well defined and separate from neighboring clusters. SSE values for each cluster were calculated by squaring each time series object's distance to its respective cluster's prototype, and then summing the squared distances. It is a general measurement of error within the clusters, where error is the deviation of each object from its cluster's prototype [49, 33]. We present SSE as a percent change from each clustering iteration.

While SSE measures individual clusters' compactness, DI is a global measure of cluster com-

pactness as a measure of clustering quality; it considers the entire set of clusters in its calculation. For DI (equation shown in Table), C_i, C_j...C_k are clusters within a clustering analysis, where the

k is the total number of clusters. The compactness of the clusters is measured on the distances between two clusters $dist(C_i, C_j)$, where the minimum of $dist(C_i, C_j)$ is the distance between the closest pair of clusters, and $dist(C_k)$ represents the maximum distance within the largest cluster in the clustering. DI is a distance ratio, so it is unitless and boundless.

Like DI, DBI is also a measure of cluster compactness, but it measures compactness and separation in a slightly different way; where DI measures compactness based on a ratio of distances, DBI is a measure of the average similarity between clusters, and it is calculated by comparing the similarity of each respective cluster to the cluster that is most similar to it. The similarity between two clusters $sim(C_i, C_j)$ that are the most similar to each other (i.e., $max_{j=i}$) are then normalized by the distance between them. In this analysis, the measure of similarity varies by the similarity metric used in each analysis (e.g., Euclidean distance). Together, we used a combination of each of these metrics (Table) to assess the clusterings' compactness and separation from other clusters, as well as the cohesion of the data points within the cluster.

Table 9: CVIs used to determine clustering quality.

CVI	Calculation
Silhouette index (SI)	$SI = \frac{b_i - a_i}{max(a_i, b_i)}$
Dunn index (DI)	$DI = \frac{\min_{1 \le i=j \le m} dist(C_i, C_j)}{\max_k dist(C_k)}$
Davies-Bouldin Index (DBI)	$DBI = \frac{1}{k} \sum_{i=1}^{K} max_{j \; i} \frac{sim(C_i, C_i) + sim(C_j, C_j)}{dist(C_i, C_j)}$
Sum of Squared Errors (SSE)	$SSE = \sum_{i=1}^{n} \sum_{j=1}^{k} dist(x_i, c_j)^2$

In addition to these CVIs, clusterings with well-performing metrics were assessed visually. To visualize the clusterings, the data were reduced to two dimensions by extracting the first two principal components of the data, where these two dimensions are those that capture the most variation in the data. Each segment of the bike data/intersection of the pedestrian data are plotted along each of these dimensions and the color coded by cluster assignment. This helps to visualize the clusters, which is important for interpreting the meaning of the CVIs.

RESULTS

Each metric's clusterings are shown in Figure and Figure , where the pedestrian metrics are P1-P4, and the bike metrics are B1-B4. We present CVIs for each clustering approach (ED, DTW, and feature-based) for each metric, with one exception: because of the scale of the bicycle data used in this analysis, we could not produce DTW metrics for the B1 metric without exceeding computer memory, so DTW results are not included in B1. In addition to the CVIs, we also present the percent change in the sum of squared error between clusterings, as this is often used to inform cluster selection.

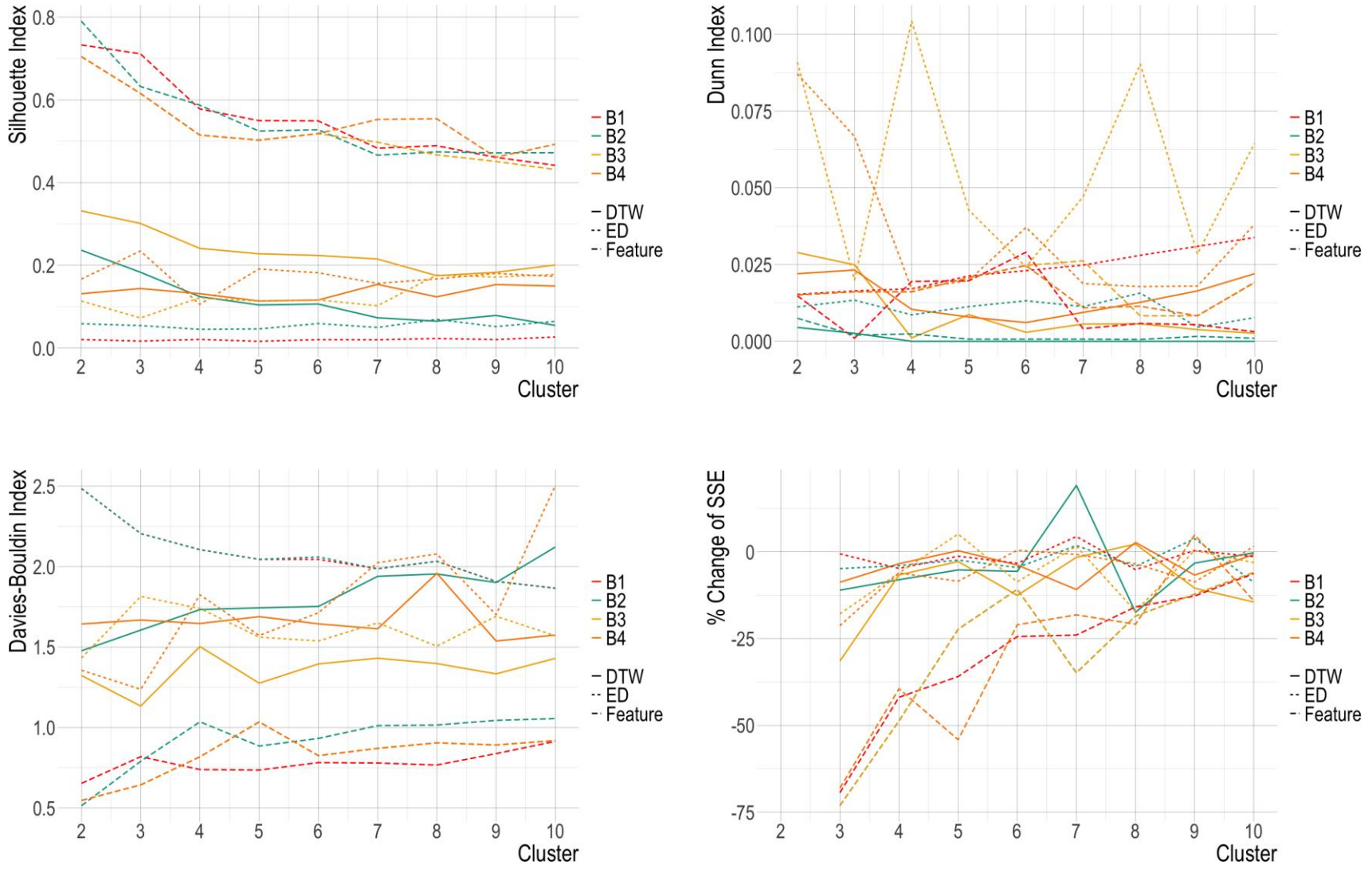


Figure 18: Graphs. Bike proxy data clustering results by CVI

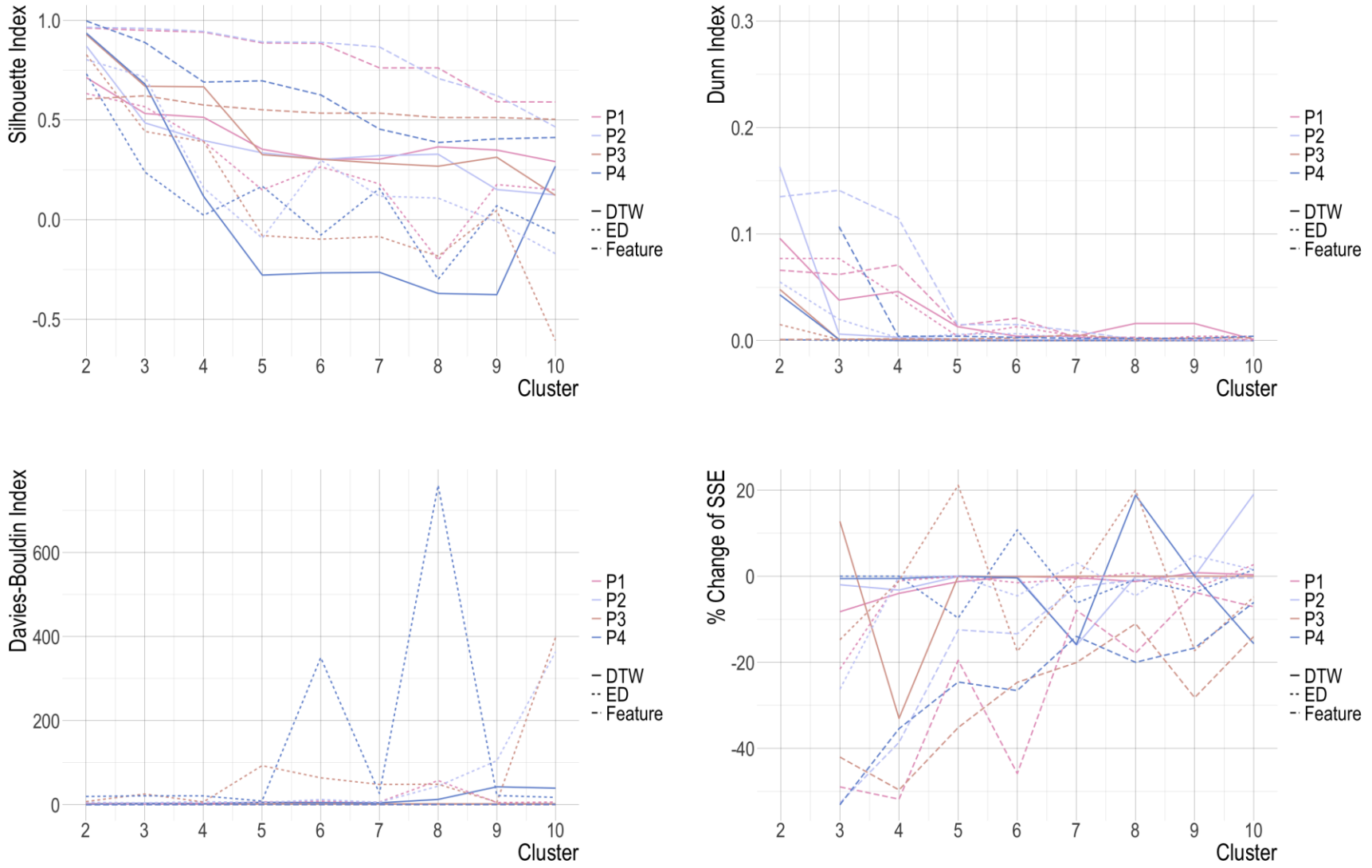


Figure 19: Graphs. Pedestrian proxy data clustering results by CVI. Note that the 2 clustering solution had a DBI of 21.5, which is not shown due to graph limits.

Table 10: Summary of clustering results (**bold** = best performing metrics)

Metric	Measure	Average SI	Average DI	Average DB	Median SI	Median DI	Median DB	Max SI	Max DI	Min DB	Indicated Cluster	Best Method
P1	ED	0.256	0.025	11.281	0.179	0.004	6.368	0.632	0.077	2.808	2-4 clusters	Feature/ DTW
	DTW	**0.414**	**0.026**	**2.020**	**0.353**	**0.016**	**2.134**	**0.712**	**0.096**	**1.195**	**2 clusters**	
	Feature	**0.814**	**0.027**	**0.555**	**0.884**	**0.014**	**0.547**	**0.961**	**0.071**	**0.405**	**4 clusters**	
P2	ED	0.214	0.010	61.584	0.118	0.002	7.618	0.802	0.055	4.537	Conflicting results	Feature/ DTW
	DTW	**0.368**	**0.019**	**2.202**	**0.328**	**0.001**	**2.365**	**0.870**	**0.163**	**1.171**	**2 clusters**	
	Feature	**0.812**	**0.048**	**0.662**	**0.889**	**0.015**	**0.738**	**0.965**	**0.141**	**0.393**	**2-3 clusters**	
P3	ED	0.073	0.002	77.580	-0.080	0.000	48.359	0.827	0.015	6.465	2 clusters	Conflicting Results
	DTW	0.431	0.006	2.521	0.313	0.000	2.408	0.928	0.048	2.225	2 clusters	
	Feature	0.550	0.002	0.500	0.534	0.002	0.505	0.621	0.006	0.449	3 clusters	
P4	ED	0.104	0.000	138.560	0.070	0.000	21.267	0.731	0.001	8.879	Conflicting results	Feature
	DTW	0.049	0.005	13.185	-0.264	0.000	5.314	0.935	0.043	1.485	2 clusters	
	Feature	**0.617**	**2.407**	**0.479**	**0.625**	**0.004**	**0.524**	**0.996**	**21.534**	**0.002**	**2 clusters**	
B1	ED	**0.021**	**0.023**	**2.075**	**0.021**	**0.023**	**2.045**	**0.027**	**0.034**	**1.866**	**10 clusters**	ED
	DTW	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
	Feature	0.555	0.011	0.781	0.549	0.006	0.780	0.733	0.029	0.653	2 clusters	
B2	ED	0.056	0.011	2.077	0.055	0.011	2.045	0.070	0.016	1.866	8-10 clusters	Feature
	DTW	0.114	0.001	1.803	0.104	0.000	1.752	0.237	0.005	1.477	2 clusters	
	Feature	**0.550**	**0.002**	**0.921**	**0.525**	**0.001**	**1.012**	**0.791**	**0.007**	**0.515**	**2 clusters**	
B3	ED	**0.130**	**0.057**	**1.612**	**0.116**	**0.047**	**1.570**	**0.178**	**0.104**	**1.432**	**8 clusters**	ED/ DTW
	DTW	**0.233**	**0.009**	**1.358**	**0.224**	**0.006**	**1.395**	**0.332**	**0.029**	**1.133**	**3 clusters**	
	Feature	0.523	0.017	0.828	0.503	0.016	0.871	0.705	0.026	0.547	Conflicting results	
B4	ED	**0.173**	**0.036**	**1.779**	**0.173**	**0.020**	**1.713**	**0.235**	**0.087**	**1.237**	**3 clusters**	ED
	DTW	0.135	0.014	1.664	0.132	0.013	1.644	0.154	0.023	1.537	Conflicting results	
	Feature	0.546	0.016	0.828	0.518	0.016	0.871	0.705	0.025	0.547	Conflicting results	

For the pedestrian results, Figure , which shows results from all clusterings, and Table , which summarizes the CVIs, suggest that 2-4 clusters are optimal. With few exceptions, feature clustering results returned the highest SI and DI and the lowest DB, indicating that regardless of the number of clusters, the feature-based clustering returned the most cohesive, compact, and separate clusters. P1's feature-based results perform best (by the authors' definition), which indicates 2 or 4 clusters. For P2, feature-based clustering generally performs the best, although DTW performs slightly better in terms of cluster separability (i.e., Dunn index). CVIs from P3 are less consistent; ED clusterings did not perform well across any CVI, but DTW clusterings performed better on some CVIs, and feature clusterings performed better on other CVIs. While P4's feature clusterings were the most compact, separated, review of the clusterings plotted into two dimensions via Principal Components Analysis (PCA) of the data indicated poor clustering, as only one signal clustered

separately from the rest. After reviewing the clusterings and CVIs from of the pedestrian metrics' clusterings, P1 and P2 with 2 clusters appear to be the highest quality clusterings, shown in Figure .

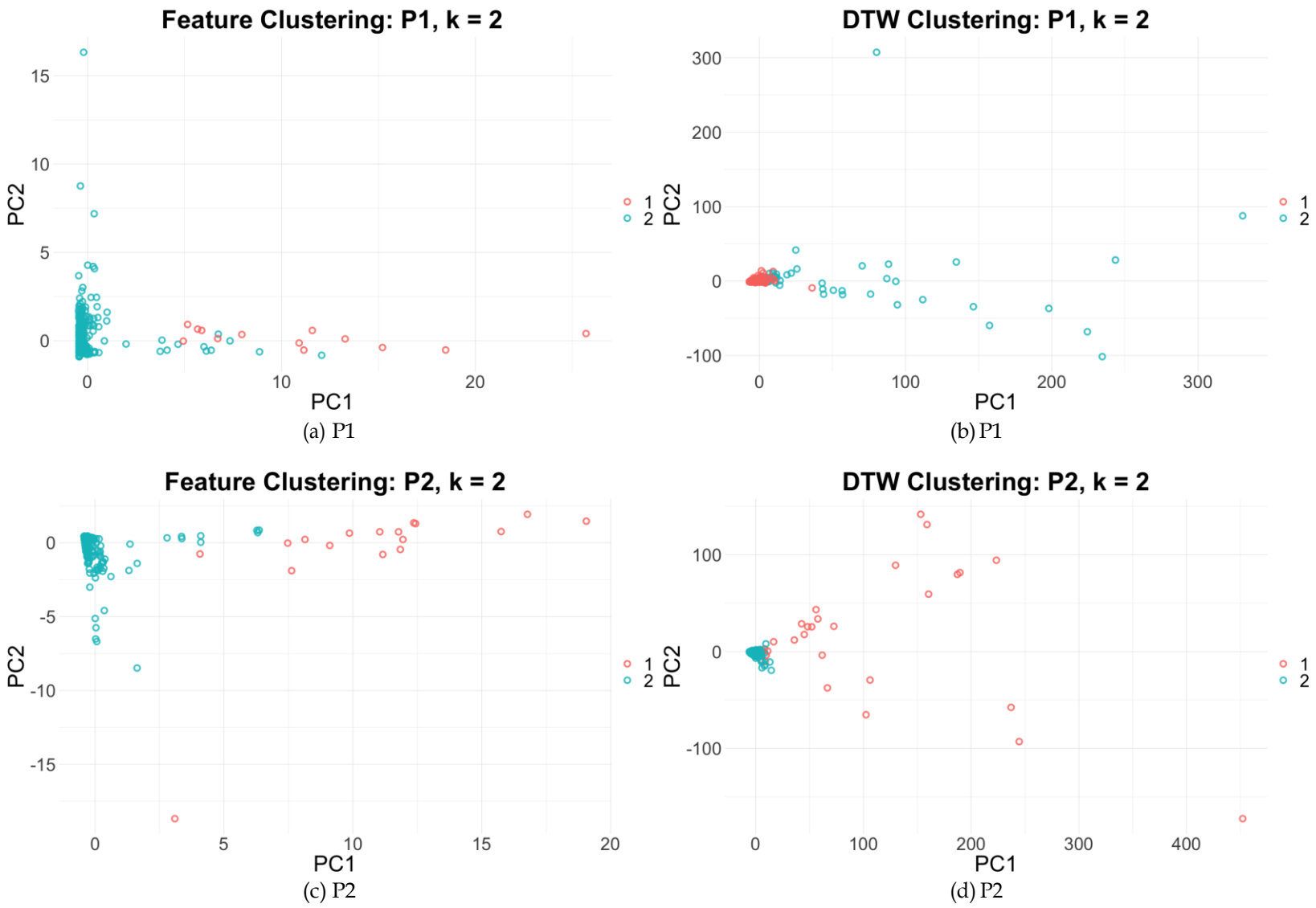
In general, the results of the bike metrics CVIs are less clear and in some cases contradictory. ED results for both B1 and B2 pointed to 8-10 clusters. However, the ED results for B3 and B4,

which was the only method that pointed towards a consistent result in both of these metrics, pointed towards 8 clusters in B3 but 3 clusters in B4. The ED results had comparatively poor performance across the clusterings' CVIs, except for the Dunn index. Feature-based clustering on all but one metric (B2) point towards conflicting results. It appears that the clusterings could not create both separated and distinct clusters, as the SI, DI, and DB metrics were contradictory; the silhouette index points towards more clusters (8-10 clusters), but the DB index strongly points to 2 clusters. DTW results were generally poor in all applications except B3, where they clearly pointed to three clusters. Overall the ED similarity measure clustering performed well generally (except for B3, which did not have CVIs that pointed to consistent clusterings across the clusterings), and they pointed to a range of possible clusters from 8-10 clusters, but it should be noted that in general these are poorer clustering results. After reviewing the CVIs and visualizations of the clusterings, we present the best clustering from each metric in Figure .

The spatial distribution of select cluster results are shown in Figure and Figure . In the pedestrian results, the first cluster is concentrated in the Midtown and Downtown Atlanta areas. Specifically, the signals appeared to be located around areas such as Metropolitan Atlanta Rapid Transit Authority (MARTA) stations, Centennial Olympic Park, Piedmont Park, the Georgia Institute of Technology campus, major recreational pedestrian areas like the Mercedez-Benz Stadium, as well as significant arterials such as North Avenue Northeast, Ponce Del Leon Avenue Northeast, and 10th Street Northwest. These are signals that are highly seasonal week-to-week and near major pedestrian trip attractors. Most of the second cluster are located in rural areas, suburban areas, and in less-trafficked urban places; they have less consistent weekly trend cycles, stronger monthly seasonality (but less clear weekly seasonality), and lower overall activity levels.

In Figure , the mapped and categorized segments of the best performing B4 solution indicates three patterns – highly recreational segments in cluster one, segments that connect to the highly recreational segments in cluster two, and (relatively) low activity segments. We also reviewed the map of the B3 DTW three-cluster solution, and we found nearly identical solutions, where only the ends of several connected segments were sparingly assigned different clusterings. This

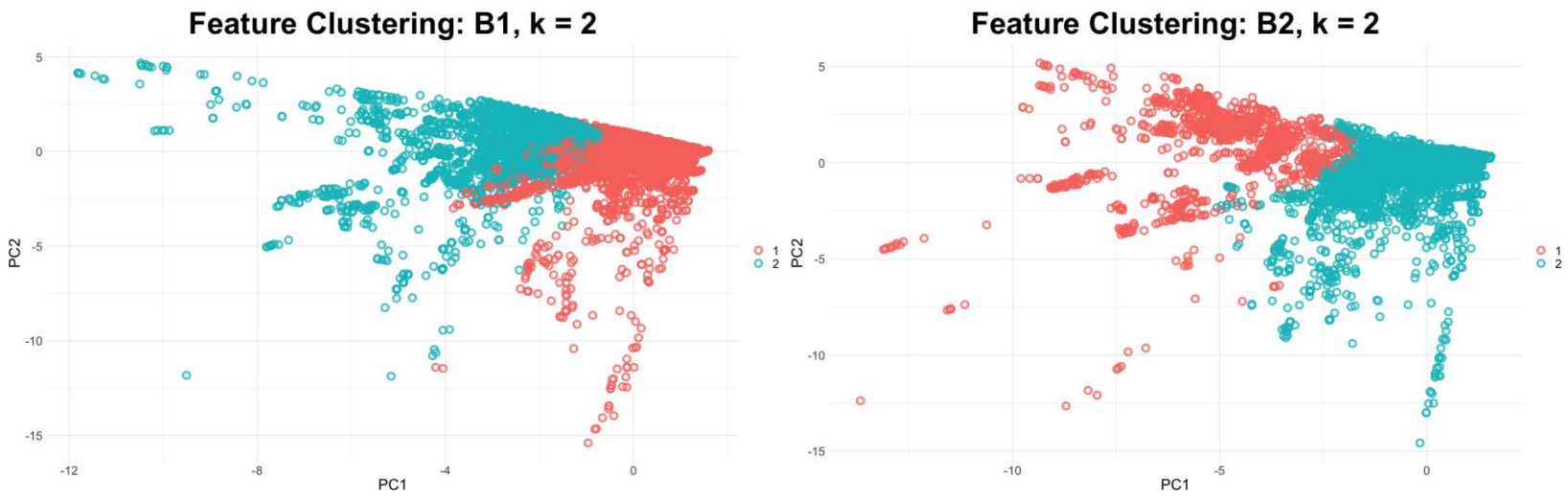
consistency across methods shows stable results, and indicates that Strava data that is filtered for the most active segments can produce stable clusterings across clustering approaches.



(a) P1 (b) P1

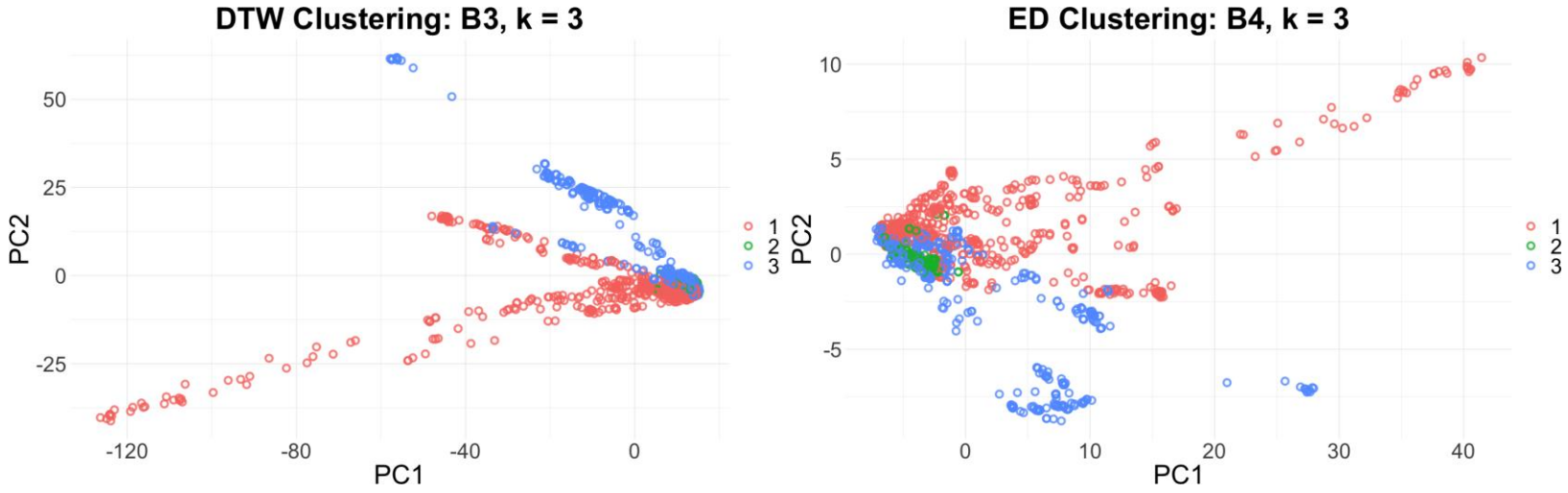
(c) P2 (d) P2

Figure 20: Graphs. Best performing clusterings for P1 and P2, where PC1 and PC2 are the first principal components extracted for 2-D plotting of clustering results



(a) B1

(b) B2



(c) B3

(d) B4

Figure 21: Graphs. Best performing clusterings for B1-B4, where PC1 and PC2 are the first principal components extracted for 2-D plotting of clustering results

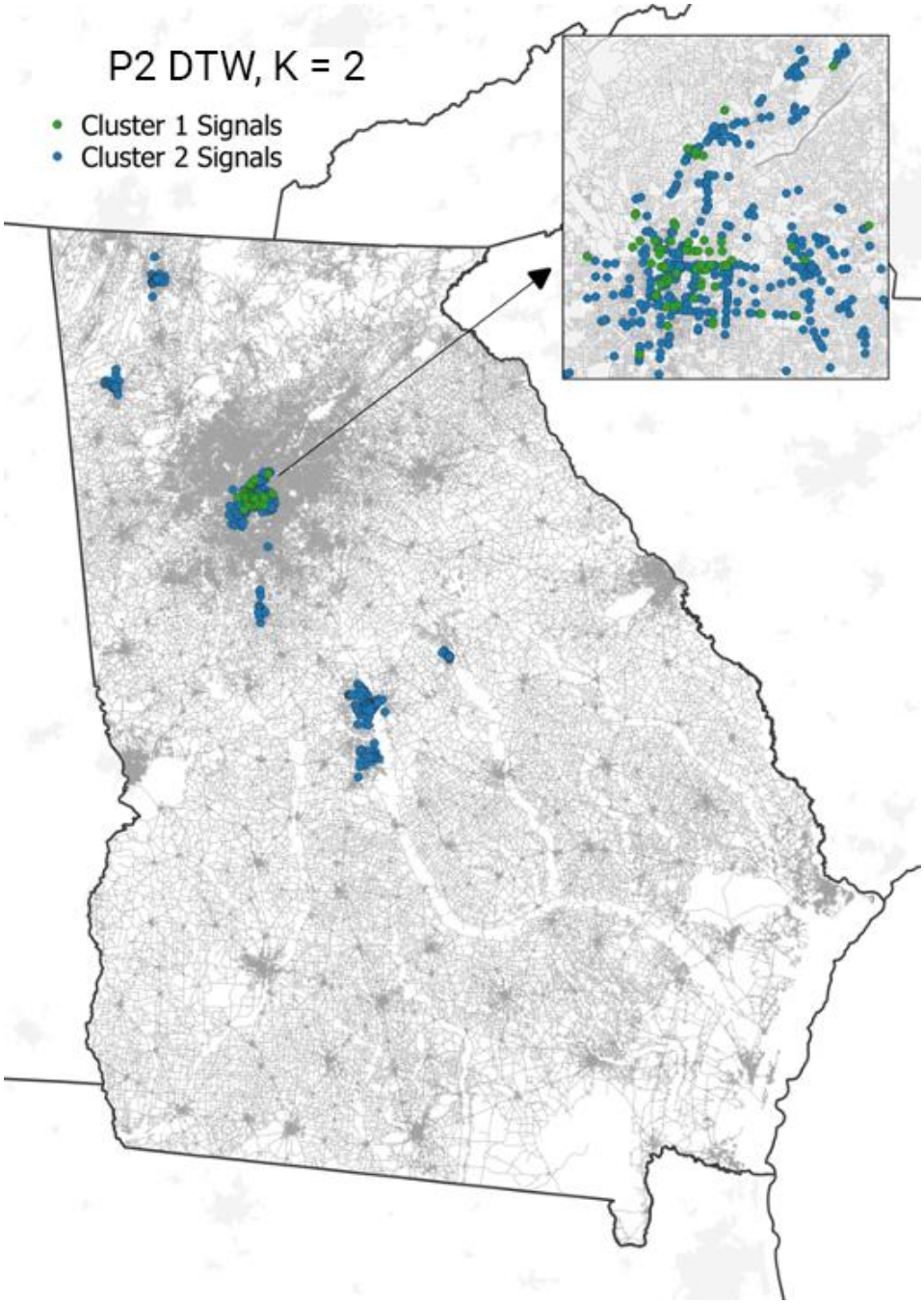


Figure 22: Map. Spatial distribution of the DTW two-cluster solution for metric P2

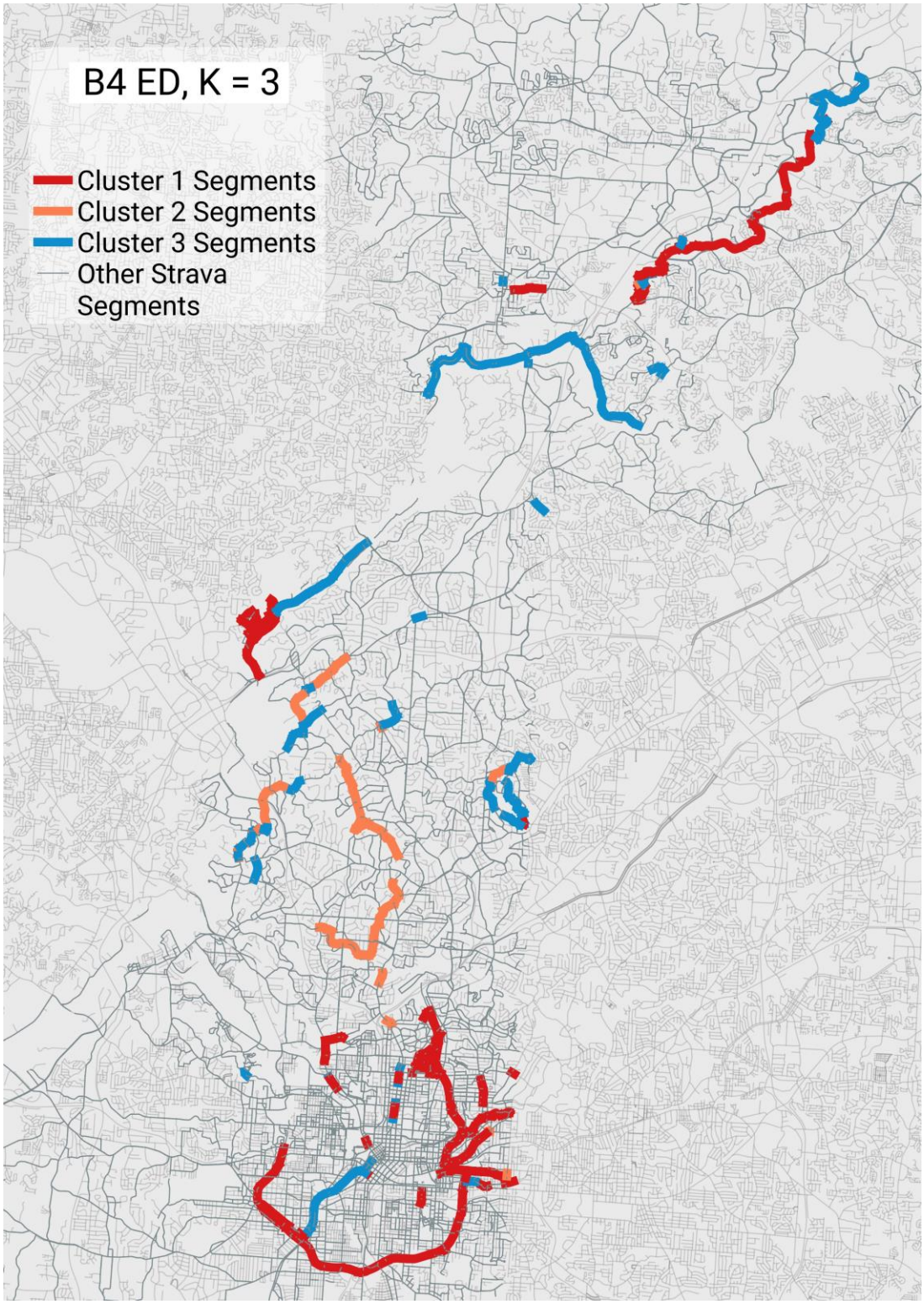


Figure 23: Map. Spatial distribution of the ED three-cluster solution for metric B4

INTERPRETATION OF RESULTS

This work assessed three clustering methods' performance applied to two datasets that are proxies for biking and walking. High-quality clusterings were defined as those that result in compact, separated, and cohesive clusters, aligning with best practice in count programs and with classification problems in unlabeled machine learning. Because the data used in this analysis were proxies of data instead of direct counts of biking and walking, we assessed clusterings across several different metrics of activity that are functions of the proxy data to see which metric(s) may result in the best clusters. The clusters were assessed based on four well-established internal validity indices that capture separation, compactness, and cohesion.

Discussion of Findings

The results indicated that the pedestrian proxy data – thousands of pedestrian actuations at hundreds of Georgia traffic signals – can be used to form cohesive and separated clusters using two of the clustering methods tested here. Across all metrics P1-P4, each clustering method resulted in clusterings with CVIs that consistently pointed to a small range of reasonable clusterings (2-4 clusters). Because existing research has indicated that ATSPM data does correlate to pedestrian counts across several contexts [33, 34, 35, 36, 37, 38], it is likely these findings point to the capacity for clustered ATSPM data to be used as a starting point to begin to collect pedestrian count data for factor group development.

Across all pedestrian clustering approaches, the CVIs indicate that DTW and feature-based clustering results form the most cohesive and separate clusters. To the authors' knowledge, this work was the first to assess the performance of feature-based clustering on data across several metrics. An important finding for application of this work to practice is that feature-based clustering in general performs as well or better than DTW on the pedestrian data. As noted previously, DTW approaches are computationally complex; DTW analyses took between 12 and 14 hours to run for each pedestrian metric (and 125 hours to run on the bike metric B2). Feature-based clustering,

which only take a few seconds, appears to be a satisfactory alternative to DTW across metrics P1-P4 – a time-saving finding for future work in this area.

P1 and P2 performed best of of the pedestrian metrics when assessed holistically, including both the CVIs and and the PCA-plotted clusterings. It made sense that a metric that represents button pushing would have relatively clear patterns over time, and it made sense that it may be more cohesive than P3, (which is button pushes normalized by the number of phase starts within an hour) because there may be multiple button pushes per pedestrian event at a location (i.e., P3), but there will only be one pedestrian call registered to the system (i.e., P4). We recommend further research in quantifying the relationship between P1 and pedestrian volumes at intersections.

The clustering results derived from the self-reported bicycling data (Strava) produced lower quality results overall compared to the ATSPM data clusterings. While the best performing methods across all metrics indicated 2-3 clusters in the data, no method showed a clear advantage across the metrics. In some cases, feature-based clustering performed well (B1 and B2), and in other cases, this approach performed poorly (B3 and B4). In general, the shape-based approaches created better clusterings when there were fewer (more information rich) segments in the data, and feature-based clustering performed well when all non-zero segments were included in the analysis. We consider several possibilities that could have led to this result. First, it is possible that bike volumes are highly variable and the self-reported data is unevenly reported in the population; it could be that the clusterings that result in the most cohesive clusters (i.e., clusters with the most similar locations) need to be larger to capture all the differences in variation, but the differences are not so large as to result in separated clusters. If this is the case, then the cluster assignment should be selected based on the specific use-case; if the analyst wants to capture variation in activity even if those variations are small, perhaps more clusters would be better, but if the analyst wanted clusters that were the most distinct from each other, then fewer clusters would be better.

Another reason for these Strava clustering results could be related to the zero-inflated nature of this self-reported data; such a large share of zero values (which is the case with most of Strava data) could preclude formation of compact clusters. While we attempted to address this concern

in part with metrics B3 and B4, likely zero inflation still notably impacted the results. Much of the true information is missing from Strava data, so perhaps it contains too little of the complete distribution of activity to form distinct clusters. Methods of time series clustering that apply well to zero-inflated series should be explored, such as state-space modeling or model-based clustering. Also future iterations of this work should assess clustering results when a stricter limit for inclusion in the analysis is used; these analyses used a threshold of at least one recorded bicycle trip for three months to include the segment in analysis for B1 and B2, and 10 rides per week for B3 and B4, but perhaps a higher threshold, like 20 or more rides per day, would have resulted in better clusterings. In summary, it does not seem likely that Strava data can be used in Georgia to determine distinct AT typologies, but more research is needed to explore how Strava data can be used to supplement observed count data for typology development, or how this method could be used in other locations with more biking.

Limitations

There are a few limitations to this work which point to opportunities for future research. First, given the nature of factor group development in general and this clustering task specifically, it was not possible to use external CVIs to validate clustering selection or to further assess the effectiveness of different clustering methods. It was not known which segments/intersections in each dataset should be assigned to which cluster a priori, so external CVIs were not applicable. This is, unfortunately, the reality of many clustering problems, however, and not a limitation unique to this work.

Another limitation of this work relates to equity concerns. For both proxy datasets used in this research, (Strava) segments and (ATSPM) intersections with too little data were not considered for further analysis. While removing low-activity locations was a necessity for several reasons, including justification of the use of Georgia Department of Transportation Georgia Department of Transportation (GDOT) resources to collect the counts and ensuring that clustering was possible, it inherently removed locations that could lead to inequitable decision-making with the clustering results. For example, Strava segments that remain in the model for further consideration are likely

those that have whiter and more male ridership given the inherent bias in Strava data [46]. Regarding the pedestrian data, intersections that remain in the models will be those that are less rural and nearer major pedestrian destinations. Also, during the intersection selection process, it became clear that the infrastructure quality was lower in majority minority areas, often resulting in missing ATSPM-compatible equipment and likely resulting in different pedestrian behaviors (e.g., crossing at different locations). These are unfortunate but acknowledged limitations of the proposed approach that should be addressed with future research.

An additional limitation was that these clustering analyses were completed on proxy data, not on *observed counts*, which may demonstrate different temporal patterns. This especially seems likely with bike data used to form clustering, as Strava data is shown in the literature (and perhaps also in this work) to capture only a small portion of the complete distribution of biking activity at any given location [31, 40]. The choice to cluster on proxy data was deliberate: the authors wanted to test how the different clustering approaches performed when given larger datasets from a variety of different locations in Georgia, and to assess the feasibility for practitioners to use proxy data to inform where to begin counting if they have yet to start a count formal count program. But, this choice was based on an important assumption that push-button data and Strava data are proportionately representative of the true walking and biking data, respectively. Although the validity of this assumption was not specifically tested in this study, findings from the literature have previously justified this relationship between pedestrian traffic signal data and actual pedestrian activity [26, 33, 34, 35, 36, 37, 38, 31, 40, 32].

CHAPTER 3: PREDICTIONS OF BICYCLE AND PEDESTRIAN VOLUMES

BAYESIAN FRAMEWORK FOR VOLUME ESTIMATION

Given the importance of non-motorized volumes to quantifying risk and justifying safety projects, it is imperative that transportation agencies have estimates of biking and walking at many places in the system. Biking and walking volumes are essential for selecting and prioritizing safety projects that can reduce the burden of injury on people biking and walking. The reality of many agencies, unfortunately, is that they do not systematically monitor biking and walking, so these volumes and the resulting risk calculations that could prioritize safety interventions are not available to guide their decisions. Creating a count program requires a long-term investment of resources and staff, and many agencies are unable to fund the amount of monitoring needed to meaningfully understand non-motorized travel in their system or to produce risk-based estimates. There is a need to make use of alternative sources of data to help states and communities who do not have monitoring programs, to expand the reach of communities that do have them, and to improve estimation accuracy and precision where possible. This chapter addresses two needs:

- First, non-motorized count programs or agencies without these programs should harness existing data as a source of proxy (or indirect) data. There are some readily-available data sources, like Strava data [44] and ATSPM data, that have high-resolution coverage of activity, but it is not clear from existing literature how to best use this kind of data for creating equally high-resolution estimates of biking and walking. There is a need to understand how to best use these alternative sources of data to expand agencies' understanding of non-motorized travel, especially when these are the only sources of data that many agencies have.
- Second, non-motorized data suffer from many of the challenges that limit the use of frequentist models to predict high-resolution estimates of biking and walking. Due to their highly

variable trends and the zero inflation of proxy data, more traditionally used methods in this area like Poisson regression are often unable to deal with such complex system dynamics with few observations and biased, noisy proxy data. There is a need to introduce more advanced modeling techniques to non-motorized volume estimation to allow for predictions despite limited observations and incomplete proxy data.

To improve on existing approaches for monitoring non-motorized travel, this report demonstrates the strength of multilevel predictive Bayesian models that use high-resolution proxy data for estimating equally high-resolution biking and walking volumes. It proposes a framework for leveraging four variables: relatively limited observations of biking and walking, temporal categorizations of biking and walking activity (described in Chapter 3), proxies of biking or walking, and a temporal spline variable. Two proxies were used to make predictions: crowd-sourced counts of bicyclists (i.e, Strava Metro data [45]) and archived pedestrian actuations from traffic signals (i.e., ATSPM). Both of these datasets were described more completely in the previous chapter. Different iterations and combinations of these variables were assessed for the ability to make predictions of hourly biking and walking volumes, even when relatively little observed data was available. The framework was demonstrated across various urban and rural town contexts in Georgia.

A simplified overview of the elements of Bayesian models is shown in Figure . At a high level, there are three components to a Bayesian model: prior beliefs, evidence, and posterior beliefs. Prior beliefs are "expert guesses" about the how the system of interest to the modeler behaves, and these assumptions are made in the form of distributions. Evidence in a Bayesian model is in the form of collected data about the system. The evidence is then used to update the prior beliefs about the system, resulting in posterior beliefs (also in the form of a distribution) about how the system behaves. This allows for the modeler's opinion to be directly infused into the model and the results, which can benefit posterior information when data are sparse (but can also result in poor or incorrect results if the assumptions are poor or incorrect).

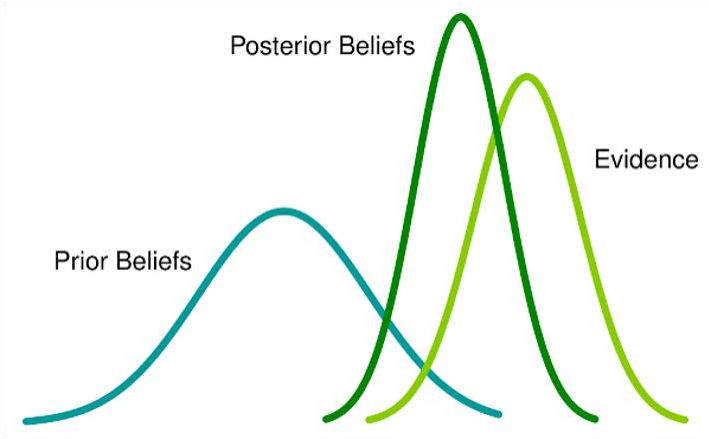


Figure 24: Illustration. Simplified schematic of Bayesian model

While there are three main elements to Bayesian models generally, in application a Bayesian model has several additional components [58, 59, 60]:

1. *Prior distributions:* First, Bayesian models incorporate assumptions in the form of parameter distributions called "prior distributions," and they represent the modelers' background knowledge of each parameter as an influential part of the model. If models are considered to be simplified descriptions of what is actually happening in the system, then the prior distributions describe the range of possible ways that the parameters can influence the system; they are intended to be specified as to best capture the generative process that underlies the outcome(s) of interest. The prior distributions selected for the model are inherently subjective; they implement what the modelers assume about variables in the model, incorporating both the uncertainty of those assumptions and the expected constraints of the processes being modeled. Priors can also be hyperparameterized to create multilevel models, where variables' parameters are also given distributions (e.g., given a variable α, we could assume $\alpha_i \sim Normal(\mu, \sigma)$, where μ and σ are not point estimates but instead assigned their own prior distributions). While more complex, this approach can create more efficient sampling and better fitting models. Priors can be determined through a variety of approaches, but regardless of the approach used to select the priors, the assumptions built in should be checked with simulated data to ensure that they produce reasonable estimates.

2. *Likelihood function and observed data:* The likelihood function is the next component of a Bayesian model. It is the joint probability (or density) of the observed data as a function of the parameters of a statistical model. If the prior distribution is $P(\theta)$, then the likelihood function is $P(y|\theta)$. This portion of the modeling framework "updates" the Bayesian model, and the function used in the model should capture the way that the variables included in the model are expected to interact in the system of interest.

3. *Posterior probability distribution, posterior predictive distribution, and predictions:* If the first three "ingredients" of a Bayesian model are the prior distributions, the observations, and the likelihood, then posterior distributions (and predictions that are made from the posterior distribution) would be the fourth "ingredient." Mathematically, the prior distribution $P(\theta)$ multiplied by the likelihood $P(y|\theta)$ is proportionate to the posterior $P(\theta|y)$:

$$P(\theta|y) \propto P(\theta) * P(y|\theta)$$

The posterior probability distribution describes the conditional distribution of model parameters, given the data observed, and it also describes the uncertainty of the parameter estimates. Then, Markov Chain Monte Carlo (MCMC) methods are used to sample from the posterior distribution of model parameters. A combination of sampling distributions from all possible parameter values is used to create the posterior predictive distribution, where each sample is weighted by its posterior probability [60]. The estimates are *distributions*, not point estimates, and they reflect the uncertainty expressed in the prior distribution and the observations. However, point predictions can still be made from these distributions; for example, the mean of a given posterior parameter's distribution can be calculated to create a single point estimate as needed, but generating a distribution of estimates is more flexible and realistically conveys uncertainty.

There were several reasons for investigating a multilevel Bayesian framework for estimating non-motorized travel volumes in this report. First, unlike the frequentist approaches used in tradi-

tional methods for non-motorized monitoring, Bayesian approaches are generally well-suited for sparse, highly variable, and/or noisy data. Directly measured non-motorized travel is often limited, as noted in the literature review, and it is often noisy for many reasons – indicating that there could be benefits to using a set of Bayesian models instead of the traditional approaches. Proxy data in general (and the proxy data used in this report specifically) are noisy, as they inherently capture a portion of the activity in the system with both systematic and random error.

Second, Bayesian modeling methods also provide a quantitative estimate of both the likelihood *and* uncertainty of the model's predictions, instead of a single point estimate [60]. Point esti- mates may convey a sense of certainty that is not truly present (even when confidence intervals are used). In Bayesian models, results are presented as a distribution with a credibility interval given the data observed (sparse or otherwise); small amounts of observed data lead to broad posterior distributions and/or large credibility intervals, which appropriately capture uncertainty.

Third, Bayesian models are able to contain more parameters than observations, which is useful in the case of sparse observations of non-motorized count data within a heterogeneous network. Finally, much data that does not directly measure non-motorized travel volumes (i.e., proxy data) are not included in many traditional approaches to estimating non-motorized travel because of their bias and/or incompleteness, as discussed in the literature review. But, Bayesian modeling can make use of this kind of data, appropriately including the certainty of that proxy measure into prior distributions.

Some of the models developed in this chapter also used a "pooled approach," or multilevel approach. Using a multilevel approach to the Bayesian model also has several benefits that come from pooling information across cases in the model. First, multilevel models are less prone to over-fitting and are more efficient compared to entirely unpooled models and thus should be used when there is theoretical basis to do so [60]. Because it is logical that information about number of people biking or walking in one location may be useful for inference about the number of people biking/walking in another similar location, a multilevel approach makes sense for these models.

General Research Design

The following research design is used to make predictions of biking and walking volumes (illustrated in Figure): First, proxy data were used to create categories of biking/walking activity (see the previous chapter). Second, a sample of locations in each cluster were selected for counts. Observations of biking and walking activity were conducted at 40 locations throughout Georgia for three days: Thursday, Friday, and Saturday. The three-day observations were collected over several months in November and December 2023 and February 2024. Third, these counts were used to update Bayesian models – separate models for biking and for walking – that model the relationship between the proxy data, the observed data, and temporal correlations (as well as other model iterations that used categories of biking and walking activity). Several model specifications were tested. The posterior predictive distributions of the model were then estimated, and MCMC sampling was used to create a sample from the posterior predictive distribution, the mean of which became hourly predictions of biking and hourly predictions of walking at each location where count data were collected. The performance of these predictions was then assessed from several perspectives, including predictive accuracy, information criteria, and absolute error.

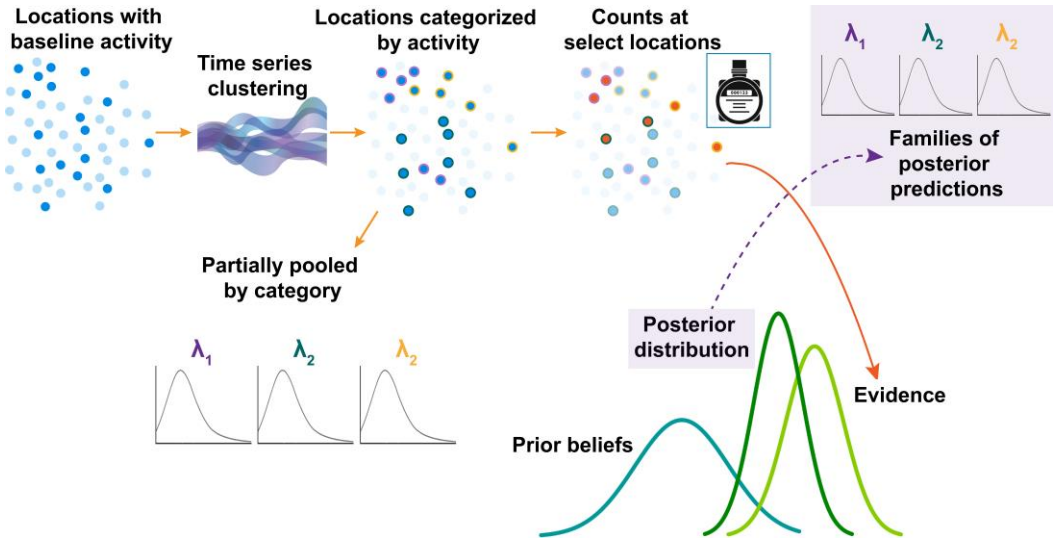


Figure 25: Illustration. Schematic representation of research design

The models used to make hourly predictions of biking and walking activity were studied in several ways. First, a simplified specification was assessed through a training-testing regime, meaning that the model was calibrated on a subsample of the data (the training data), and then predictions were compared with the remaining set of data (i.e., the testing data). Then, iterations of the model that added complexity were developed, and the performance of these models was tested through cross-validation techniques, where sub-samples of data were iteratively considered as training and testing samples.

The rest of this chapter is organized as follows: First the model assumptions, including directed acyclic graphs, the criteria used to select sites for observation, and the methods used to assess model performance are described. Then, results are shown, including the count volumes and corresponding proxy values at each location, the pedestrian volume estimations, and the bicy- cle volume estimations. The chapter concludes with a discussion of the results and directions for future research.

MODEL ASSUMPTIONS AND DEVELOPMENT

Observation Samples

Following the simplified approach to Bayesian models shown in Figure , to support the prior beliefs, evidence in the form of non-motorized counts was collected at the hourly level, and these counts were used to update the prior beliefs – assumed distributions of model parameters – to become posterior beliefs. Mathematically, the predictive models of biking and walking were regressed on observed counts of biking and walking (respectively). Thus, selecting locations for these counts was critical to making predictions. These locations were selected through several considerations:

1. *Scope of proxy data:* The two proxy datasets – ATSPM data and Strava data, described in the previous chapter – had varying scopes that set boundaries for model application in this report. ATSPM-equipped traffic signals were located at 7,034 intersections across Georgia.

Accessible Strava data, however, is limited to only Fulton County in Georgia. Within Fulton County, there are more than 49,000 segments of Strava data.

2. *Clustering and computational limitations:* Sites selected for observation must be sites that can be assigned a category of activity from the time series clustering described in the previous chapter. The full process for selecting ATSPM-equipped signal locations is discussed by Shorey [61], but the site selection included measures of intersection density, equity, quartile evaluations, data errors, and computational limitations of time series clustering methods. Ultimately, 716 signals were considered for further analysis, the results of which were summarized in Chapter 3. The process for limiting Strava data inputs were also described in the previous chapter, but approximately 41,000 Strava segments were initially considered for count locations.

3. *Urban-rural context and population considerations:* To assess the robustness of the predictions in this model, it was important to select sites that represent a variety of contexts throughout the state of Georgia. In existing literature, often predictions of biking and walking activity focus exclusively on urban settings due to the greater quantity of non-motorized activity in these locations, but this work attempted to consider rural town locations and suburban locations in the models.

4. *Budget and state contracts:* Finally, the budget that funded this work allowed for data collection at 40 locations across the state. Due to existing subcontractor negotiations, sites were further constrained by GDOT district; counts could be conducted in districts 2, 3, and 7. In districts 2 and 3, three small-to-medium-size towns in predominantly rural area were selected for further focus: Milledgeville, Macon, and Warner Robins. Budget constraints also limited the total number of days and hours included in the sample. Counts were collected for 15 hours a day (6:00 AM ET to 9:00 PM ET) for three days (Thursday, Friday, and Saturday) in 15-minute increments during December 2023 and January 2024.

5. *Cluster assignment:* Clustering results were also used as a means of selecting sites. Clus-

tering results from the feature-based time series clustering (see the previous chapter) were assigned to each Strava link and ATSPM intersection, and sites were selected to ensure that there were multiple cases of each cluster present in the sample.

In summary, count locations were selected to maximize the variety of activity, population, and infrastructure contexts within budget limitations. These counts were collected to include three different days that typically have distinct non-motorized travel patterns. These samples are then matched temporally with the proxy data (ATSPM and Strava) recorded during the same time. The location of the 40 counts are shown in Figure .

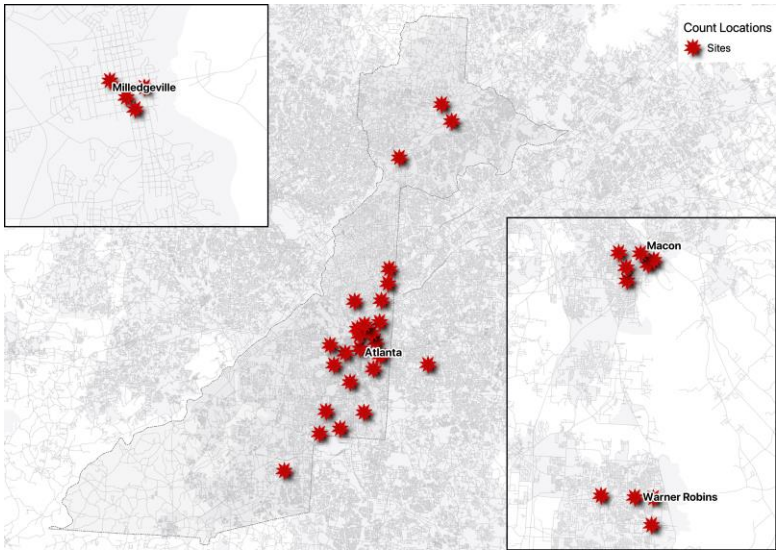


Figure 26: Map. Locations of 40 count sites around Georgia

In addition to these counts, archived counts from 2022 (locations 40-47) were also used in the model. The archived counts were accessed through an Atlanta-based non-profit, Midtown Alliance, who serves as a coalition of business leaders in Midtown. Midtown Alliance maintains several long-term counters, each of which collect hourly data of people biking and walking 24 hours a day. Four of the locations counted pedestrians, and three of the locations counted bicyclists. Specifically, these counters were located at or near the following intersections:

- Spring Street and 14th Street (pedestrian count)

- West Peach Tree and 15th Street (pedestrian count)
- Peachtree Street and 17th Street (pedestrian count)
- Peachtree Street NE and Ponce de Leon Avenue (pedestrian count)
- 5th Street near Williams Street NW (bicycle count)
- 10th Street NE near Monroe Drive (bicycle count)
- 10th Street NE near Myrtle Street NE (bicycle count)

Arrival Rate Assumptions

The proposed specification of the models in this work assumed that the counts follow a gamma-Poisson (GP) process (which is a special case of the negative binomial process). While most models in the literature assume that biking and walking counts follow Poisson processes, the models in this work employed the gamma-Poisson distribution for several reasons. It is reasonable to assume that the counts of biking and walking arise from a mixture of different processes that vary by location, which means that there would be more variation in rates of activity than is amenable to a pure Poisson process and that the data would be overdispersed [60]. A gamma-Poisson model assumes that the rates vary by observation along a gamma distribution; said another way, the gamma-Poisson model inherently assumes heterogeneity in the rates. While pure Poisson processes can be adapted to have variable rates, this model approach is perhaps more straightforward and intuitive.

Formally, when observations $X_1, ..., X_n$ are independently distributed $Poisson(\lambda)$ (where λ is the expected number of events over the time interval of interest that varies across observations), then $\lambda \sim Gamma(k, \beta)$, where k is the shape parameter and β is a rate parameter. Setting the prior distribution of λ as $Gamma(\alpha, \phi)$ is setting a conjugate prior. The gamma-Poisson process has two parameters [60]:

$$y_i \sim GP(\lambda, \phi)$$

where λ is the expected number of events over the time interval of interest (like Poisson models), and ϕ is a strictly positive variance term. Applied to this work, the number of bicyclists/pedestrians per unit of time for multiple segments/intersections can be modeled as a random variable Y, or the number of travelers per *ith* segment/intersection. An example probability mass function is shown in Figure .

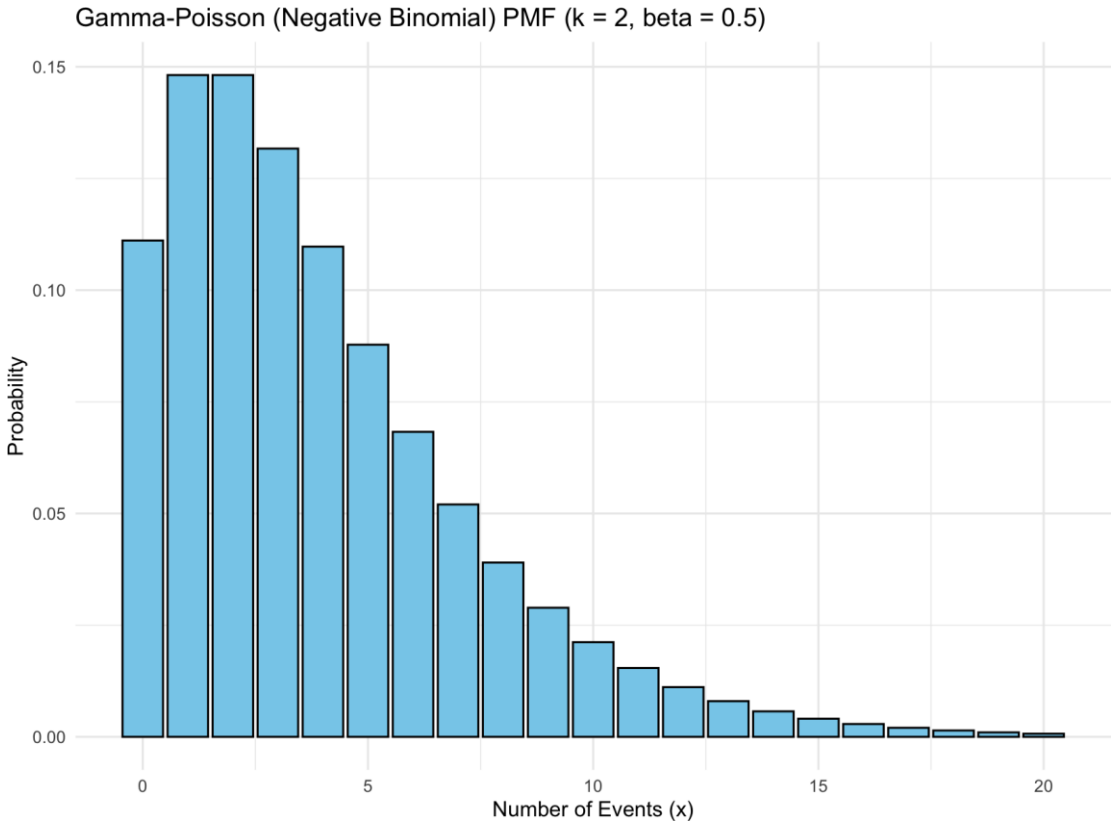


Figure 27: Graph. Example probability mass distribution of of the gamma-Poisson process, where the shape parameter *k* is assumed to be 2, and the rate parameter β is assumed to be 0.5.

To estimate arrivals of the number of bicyclists/pedestrians per segment/intersection (respectively) on a sample of segments/intersections (some of which will have no observations), the following "base specification" is used, where other model iterations include these variables along with others to assess improvements in predictive performance. Given the following definitions:

- $N_i =$ number of observed bicyclists/pedestrians per segment/intersection per hour;
- $P_i =$ count of bicyclists/pedestrians from proxy data;

- $\alpha =$ mean hyper parameter for category intercept;
- $f\,(H)_L =$ a spline fit to the hourly data, where the spline is indexed by location L and captures the temporal autocorrelation for each observed hour H;

the first iteration of the models to predict bicyclists and pedestrian volumes (separately) is:

$$N_i \sim GP\,(\lambda, \phi)$$

$$\lambda = \alpha + \beta_P P * L + f(H)_L$$

$$\alpha \sim Normal$$

$$\beta_P \sim Normal$$

$$L \sim Normal$$

$$f(H)_L \sim Student's\ T$$

$$\phi \sim Exponential$$

It is helpful to describe how each element of the predictive model works. First, the intercept term α captures the average activity. In this specification, it is not indexed by location (as other variables in the model capture variation by location). Then, the proxy data multiplier β_P and the vector of proxy data P are interacted by location L (as a categorical variable) so that the effect of the proxy data and the effects specific to the location are modeled jointly. Finally, the spline term $f(H)_L =$ is indexed by location so that the spline captures hourly variation for each location in the model. Collectively, these variables capture average activity and how that average activity relates to the proxy data over time.

Prior Predictive Simulation

Deciding on the specification for model priors is a key step in Bayesian models, especially when the sample size is small (which is the case for most of the observations in the models). This research set priors through simulations of the data and predictions based on the simulations; predictions are made based on the prior distribution applied to simulated data (see example of simulated data in Figure , which helps elucidate the implications of the prior selection [60]. An example of possible prior values predicted from the simulated data is shown in Figure . The following priors were selected for the base model specification:

$$\alpha \sim Normal(0, 1)$$

$$\beta_{P\ bike} \sim Normal(2,$$

$$1)\ \beta_{P\ ped} \sim Normal(3,$$

$$2)$$

$$L \sim Normal(0, 1)$$

$$f(H)_L \sim T(3, 0, 2.5)$$

$$\phi \sim Exponential(1)$$

Model Performance Assessment

Four different methods of assessing the models' predictive performance were tested in this work:

- *Training data prediction comparisons:* Data were separated into training versus testing sets by day of the week; all days in the dataset except for Friday were used to train the model, and Fridays were separated out as testing data to be used in the next steps of performance assessment. Training predictions and data were compared visually, where predictions were overlaid with the observed data to visually indicate in-sample performance.

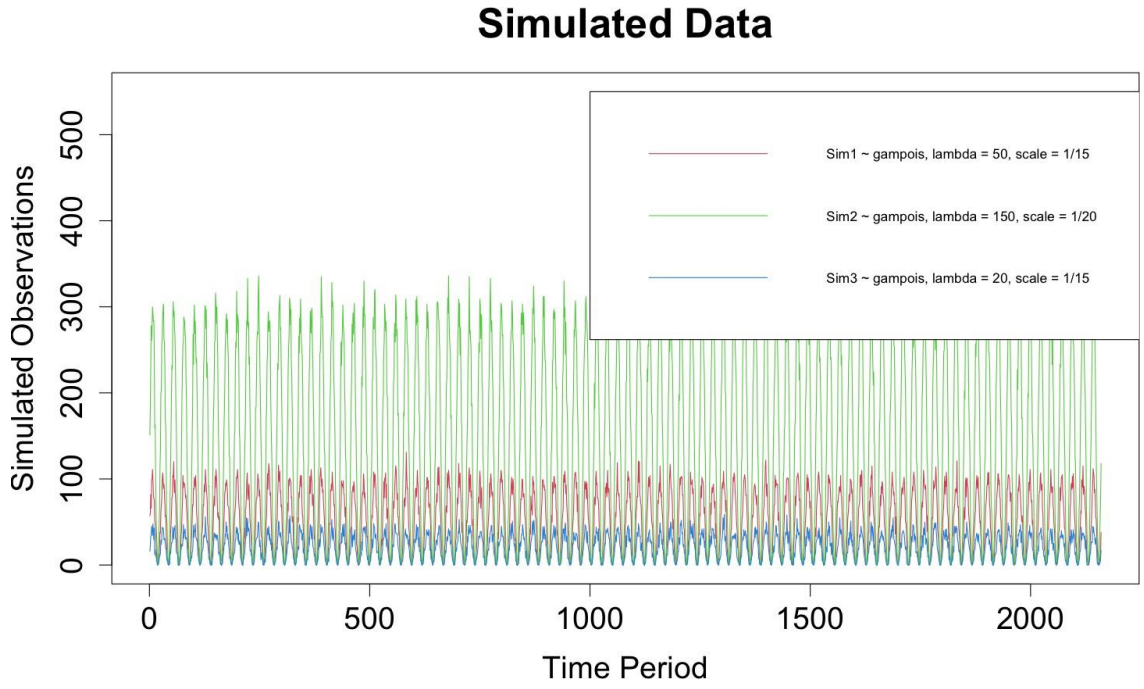


Figure 28: Graph. Simulation of observed data used to create prior predictive simulation

- *Difference between mean observed hourly value and predicted observed hourly value*: For each of the models developed in this work, differences among the predictions and observations were calculated. This was done visually and numerically; predictions were plotted against observed averages, and the differences were captured numerically by calculating the percent differences among the hourly average values of biking and walking observations at each location and the hourly average values of the models' predictions at each location.

- *Predictions' deviance from observed values for training and testing data*: Deviance is a measure related to LLog Point-Wise Predictive Density (LPPD):

$$\text{LPPD}(y|\Theta) = \sum_{i}^{n} \log \frac{1}{S} \sum_{S} p(y_i|\Theta_S)$$

where S is the number of samples and Θ_S is the s-th set of sampled parameter values in the posterior distribution [60]. Deviance is $-2 * \text{LPPD}$, so that larger deviance values are indicative of poorer model predictions, which is (arguably) more intuitive. Deviance is calculated

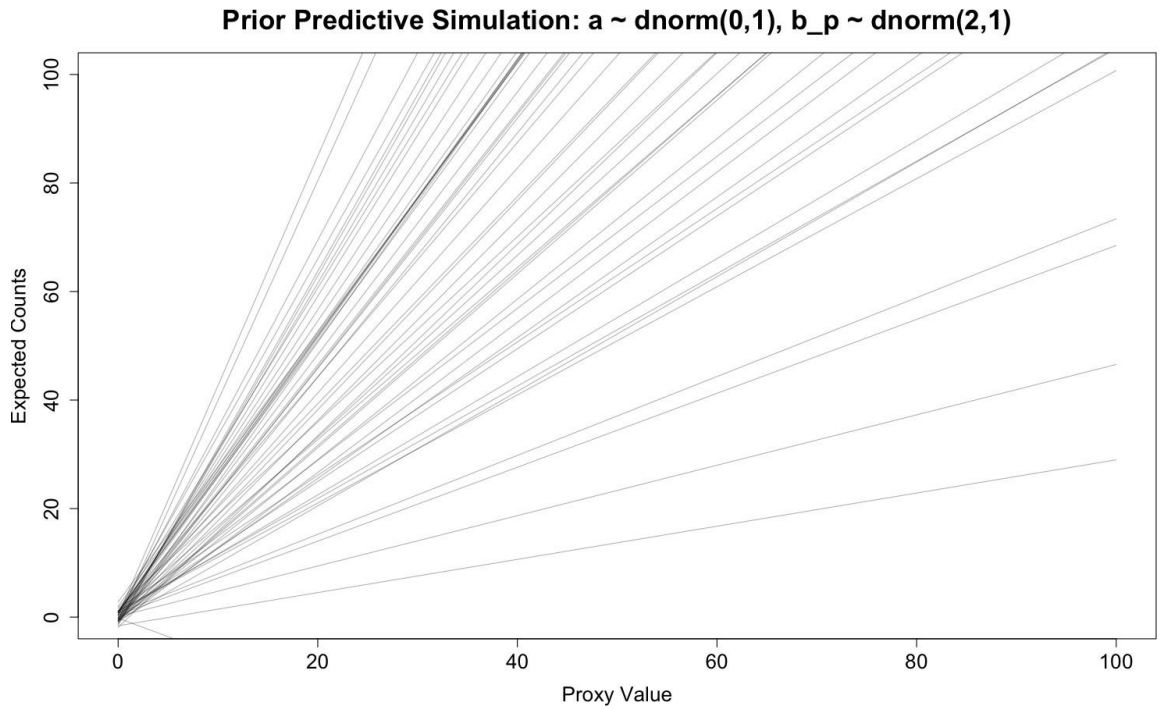


Figure 29: Graph. Prior predictive simulation, where projected lines are distributed possible values of proxy variable given the priors

for each locations' predictions in two ways: in-sample deviance (i.e., deviance from the predicted and observed data that is used for model development), and out-of-sample deviance (i.e., deviance from the predicted values developed from the model from observed data that was not used to train the model). Data are separated into training and testing data temporally. Also, the training RMSE and Mean Absolute Error (MAE) were reported for each location as a measure of absolute error. If y_i are observed values and $\hat{y}_i$ are predicted values:

$$RMSE = \sqrt{\frac{\sum_{i=1} (\hat{y_i} - y)^2}{n}}$$

$$MAE = \frac{1}{n} \sum_{i=1} |\hat{y}_i - y_i|$$

- *Cross-validation:* Finally, four model iterations (that add variables to or hyperparameterize the base specification) are trained on all the data where an approximation of Log Point-Wise

Predictive Density Leave-One-Out Cross Validation (LPPD) and the Watanabe-Akaike Information Criterion (WAIC) are used to measure model performance [62]. LPPD-LOOCV approximation is developed through , which is a technique for Bayesian model performance assessment accurately estimates the each model's LPPD. Whereas the previously mentioned use of LPPD (and deviance) reports the differences between the predicted and observed values of each model prediction by location, this measure of each model's full LPPD, approximated by PSIS. Higher LOOCV-LPPD and lower deviance values indicate predictive accuracy. PSIS is a more efficient than directly calculating LPPD, as it uses importance weights (where importance is measured as an observation's ability to change the posterior distribution) [60]. It assumes that the importance weights follow a Pareto distribution. The largest weights are used to estimate a Pareto distribution, which is then smoothed using the same Pareto distribution. The WAIC criterion is also derived from LPPD, but it includes a penalty for model complexity to offset additional predictive power that sometimes accrues to more complex models. Specifically:

$$WAIC = -2 * (LPPD - pWAIC)$$

where *pWAIC* is the effective number of parameters. Therefore, WAIC considers both model fit and model complexity. Lower WAIC values indicate better model performance while balancing fit and parameters. Also, location-specific RMSE and MAE were also reported for the best performing specification.

RESULTS

Non-Motorized Data

As described in the previous sections, counts were collected from two sources: 3-day counts 40 locations across Fulton County and Georgia that were collected in November 2023, December 2023, and February 2024; and 90 days of hourly count data from each of the permanent counters

in Midtown Atlanta. Details about each location are shown in Table .

The table is color coded to show which locations were ultimately included in the bike models and pedestrian models. Locations that are shown in red text were those that were not included in the pedestrian models due to missing ATSPM data during the time of data collection. Locations in black text are those that are used in the pedestrian models. Locations in green text are those where data was available for both bike and pedestrians models. Note that the bike models were not able to use many of the locations where data were collected. While Strava data was available at all of these locations generally, many of the locations where only three days of data were collected did not have any recorded rides during that time, so they were not useful for modeling the relationship between Strava and observed data.

Table 11: Count Locations in Georgia (red = locations that *were not* included in pedestrian models; green = locations that *were* included in the bicycle and pedestrian models)

Count ID	Major Street	Minor Street	Segment/ Intersection	Count Data	Area	Count Duration
1	Ferst Dr NW	Atlantic Drive	Intersection	Both	Fulton County	15 hours x 3 days
2	Academy Street	(Near)	Segment	Bike Only	Fulton	15 hours x 3 days
3	Azalea Drive	(Near) Haynes Bridge Road	Segment	Bike Only	Fulton County	15 hours x 3 days
4	SR 22/ E Handcock Street	South Atlanta Street Eiberi Street	Intersection	Both	Milledgeville	15 hours x
5	Monroe Drive NE	HAWK	Segment	Both	Fulton	15 hours x 3 days
6	Candler Road SE	HAWK	Segment	Both	Atlanta	15 hours x
7	14th Street NW	HAWK	Segment	Both	Fulton Metro	15 hours x 3 days
8	US-29	John Wesley Ave	Intersection	Both	Fulton County	15 hours x 3 days
9	Dekalb Avenue NE	Krog Street NE	Intersection	Both	Fulton County	15 hours x 3 days
10	Marietta Street NW	Baker Street NW	Intersection	Both	Fulton	15 hours x
11	North Point Parkway	Kimball Bridge Rd	Intersection	Both	Fulton County	15 hours x 3 days
12	Northside Drive NW	Woodward Way NW	Intersection	Both	Fulton County	15 hours x 3 days
13	Dunwoody Road NE	Winall Down Road NE	Intersection	Both	Fulton County	15 hours x
14	Cherokee Avenue SE	Georgia Avenue	Intersection	Both	Fulton County	15 hours x 3 days
15	Cleveland Avenue SE	Browre Mill Road SE	Intersection	Both	Fulton	15 hours x
16	Bluffington Road	Sable Run Raod	Intersection	Both	Fulton County County	15 hours x 3 days
17	MLK Drive NW	Joseph E Lowery Boulevard NW	Intersection	Both	Fulton County	15 hours x 3 days
18	Lenox Road NE	E Paces Ferry Road NE	Intersection	Both	Fulton County	15 hours x 3 days

19	Eisenhower Parkway	C Street	Intersection	Both	Macon	15 hours x 3 days
20	MLK Jr Boulevard	Oak Street	Intersection	Both	Macon	15 hours x 3 days
21	Greene Street	Wilkinson Street	Intersection	Both	Milledgville	15 hours x 3 days
22	North Avenue	Parkway Drive NE	Intersection	Both	Fulton County	15 hours x 3 days
23	5th Street	Cherry Street	Intersection	Both	Macon	15 hours x 3 days
24	SR 22/ W Montgomery Street	SR 243/ Clarke Street	Intersection	Both	Milledgville	15 hours x 3 days
25	Virginia Avenue	Rainey Avenue	Intersection	Both	Fulton County	15 hours x 3 days
26	Watson Boulevard	Holly Street/ Watson Boulevard	Intersection	Both	Warner Robins	15 hours x 3 days
27	SR 3/ Metropolitan Parkway	Dill Avenue	Intersection	Both	Fulton County	15 hours x 3 days
28	Russell Parkway	Davis Drive	Intersection	Both	Warner Robins	15 hours x 3 days
29	SR 243	Franklin Street	Intersection	Both	Milledgville	15 hours x 3 days
30	10th Street NE	Myrtle Street NE	Intersection	Both	Fulton County	15 hours x 3 days
31	College Street	Hardeman Avenue	Intersection	Both	Macon	15 hours x 3 days
32	SR 247/ Watson Boulevard	Carl Vinson Parkway	Intersection	Both	Warner Robins	15 hours x 3 days
33	Pio Nono Avenue	Hillcrest Avenue	Intersection	Both	Macon	15 hours x 3 days
34	SR 237/ Piedmont Road	Morosgo Dr	Intersection	Both	Fulton County	15 hours x 3 days
35	SR 139/ RDA Boulevard SW	Cascade Avenue	Intersection	Both	Fulton	15 hours x 3 days
36	Joseph E Boone Boulevard NW	West Lake Avenue NW	Intersection	Both	County Fulton	15 hours x 3 days
37	17th Street	I-85/I-75 Ramp	Intersection	Both	County Fulton County	15 hours x 3 days
38	Mercer University Drive	Canton Street	Intersection	Both	Macon	15 hours x 3 days
39	SR 247	Davis Street	Intersection	Both	Warner Robins	15 hours x 3 days
40	GA 14/ East Point Street	W Cleveland Avenue	Intersection	Both	Fulton County	15 hours x 3 days
41	Spring Street	14th Street	Intersection	Ped Only	Fulton County	90 days
42	West Peach Tree	15th Street	Intersection	Ped Only	Fulton County	90 days
43	Peachtree	17th Street	Intersection	Ped Only	Fulton County	90 days
44	Peachtree Street NE	Ponce	Intersection	Ped Only	Fulton County	90 days

45	5th Street	(Near) Williams Street NW	Segment	Bike Only	Fulton County	90 days
46	10th Street NE	Monroe	Intersection	Bike Only	Fulton County	90 days
47	10th Street NE	Myrtle Street NE	Intersection	Bike Only	Fulton County	90 days

Observed Pedestrian Counts and ATSPM Distributions

A summary of the total observed pedestrian counts and total ATSPM proxy counts (where P1, the number of event #90 per hour is used as the proxy metric) is shown in Table . Note that the last four locations, 41 - 44, are places where 90 days of archived data were available, and the rest of the locations are places where only three days of data were recorded.

The distribution of the observed data at locations where 90 days of data were available are shown in Figure . In this Figure, all days' observations are plotted on top of each other, where the x-axis shows 24 hours from 0 (12:00AM) to 23 (11:00PM), to show the hourly trends vary- ing over time. For ease of visualization, 3-day observations were separated into two multi-part figures: Figure which shows the recorded data at high-volume locations separated by the month of observation, and Figure which shows the 3-day observation results at low-volume locations by observation month.

Table 12: Total pedestrian observations at each location, with corresponding P1 totals

Location ID	ATSPM ID	Observed Pedestrians	P1
4	2121	492	519
6	11511	180	5
7	278	2124	851
8	3299	4772	0
9	8242	11139	1655
10	8	22858	601
11	39017	142	114
12	7147	4492	416
13	477	457	47
14	8077	4255	63
15	8080	848	9
17	8141	11062	885

18	7060	8250	993
19	3021231	121	24
20	3021159	329	4
22	246	4412	446
23	3021146	3054	48
25	49002	2135	146
26	3273	1217	182
27	7170	985	202
28	3153044	113	89
29	2264	533	537
30	472	10372	820
32	3322	471	534
33	3021087	587	116
34	7116	9101	6417
35	8614	2136	953
36	316	1320	52
37	171	5104	115
38	3021217	2673	1630
39	3299	707	191
40	8565	1604	253
41	195	48416	47069
42	193	126568	29936
43	263	50152	5162
44	261	123056	12792

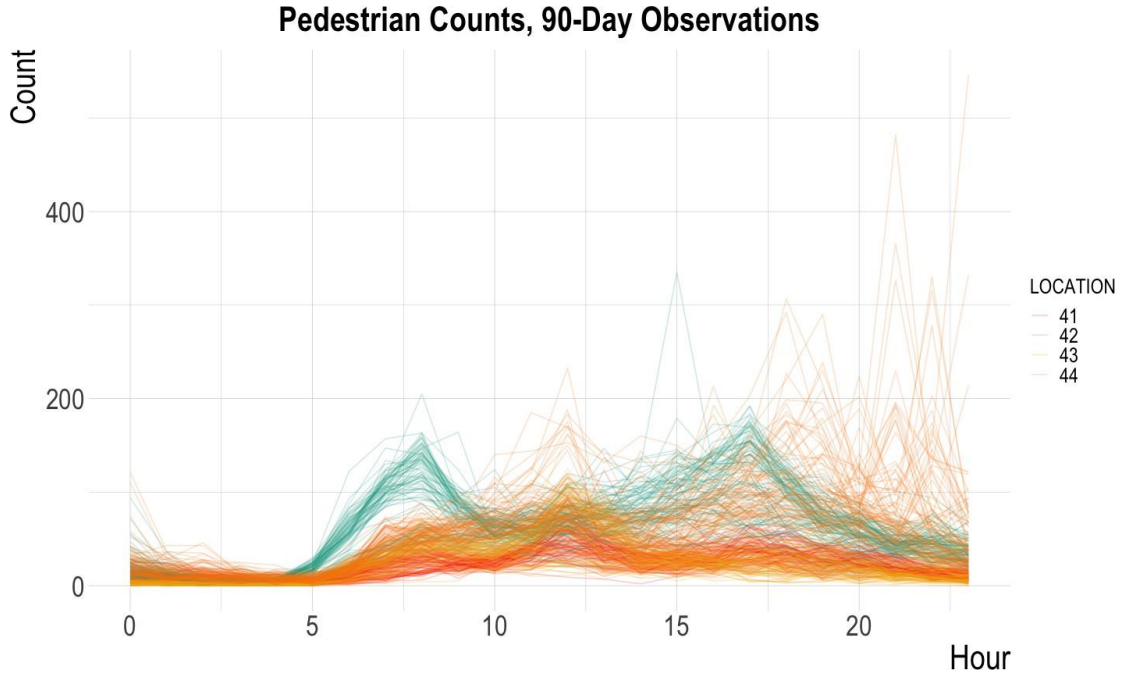


Figure 30: Graph. Variety of pedestrian activity trends at four locations where 90 days of data were collected

The relationship between the observed data and the ATSPM varied across locations. Figure and Figure show two of the 90-day locations (14th Street and Ponce de Leon Avenue, respectively), where the cooler hued colors represent the ATSPM data and the warmer hued colors represent the observed data. Comparing Figure and Figure shows that the relationship between the ATSPM data and the observed data are not the same at each location. At the Ponce de Leon Avenue location, the proxy data did not follow the same trends as the observed data. It also showed lower variability than the observed data and was only representative of a relatively small proportion of the activity that actually occurred at this location. The 14th Street location, however, showed a much wider variety of ATSPM event trends, both in variance and magnitude, despite a quite consistent trend of pedestrian activity.

Similarly, Figure through Figure show the relationship between the ATSPM events and the observed data at several locations where only 45 hours of data were recorded. At the 14th Street HAWK location in Fulton County (Figure), actuations somewhat match the trend of observed ac-

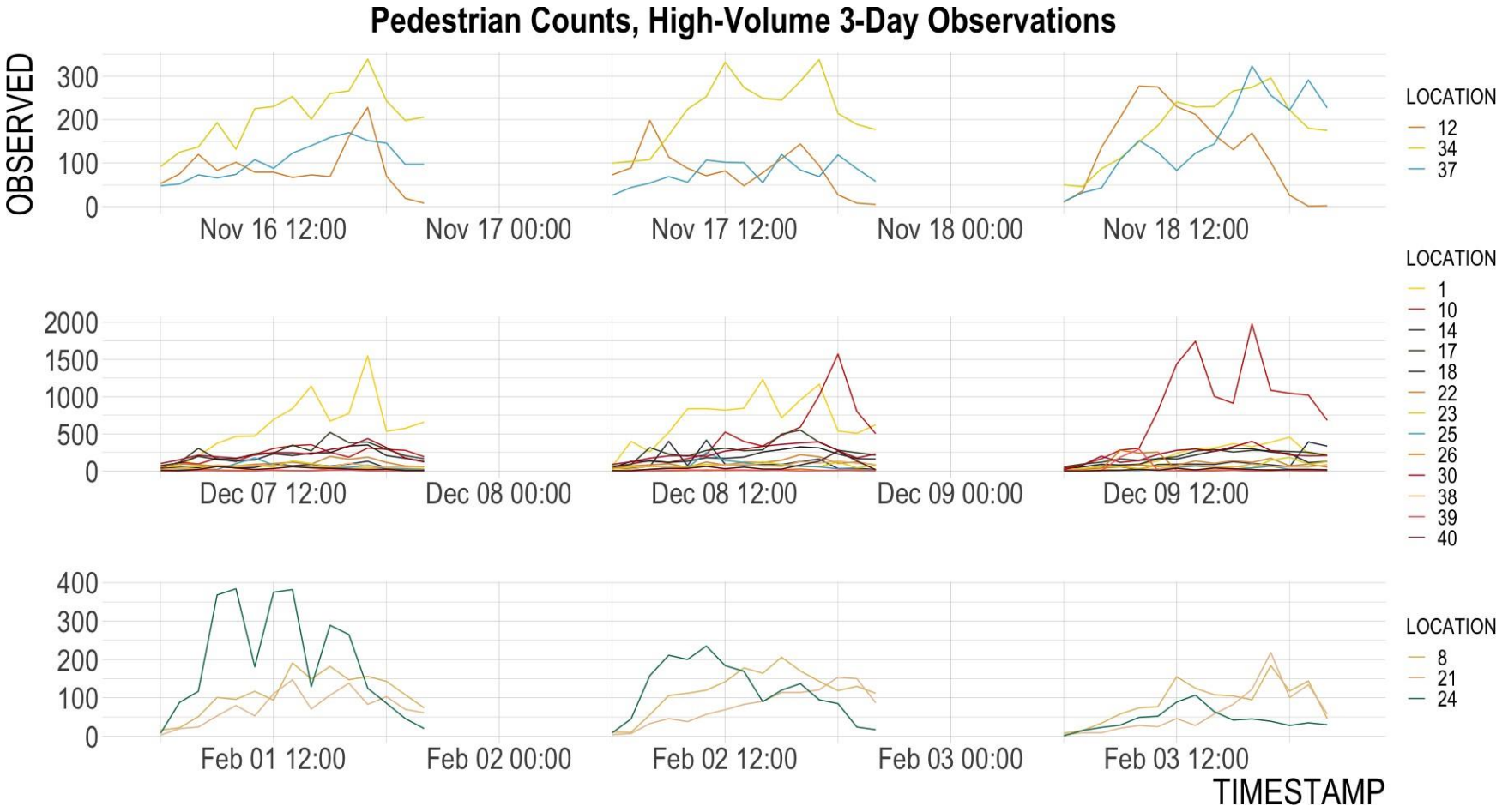


Figure 31: Graphs. Pedestrian counts at high-volume locations where three days of data were collected

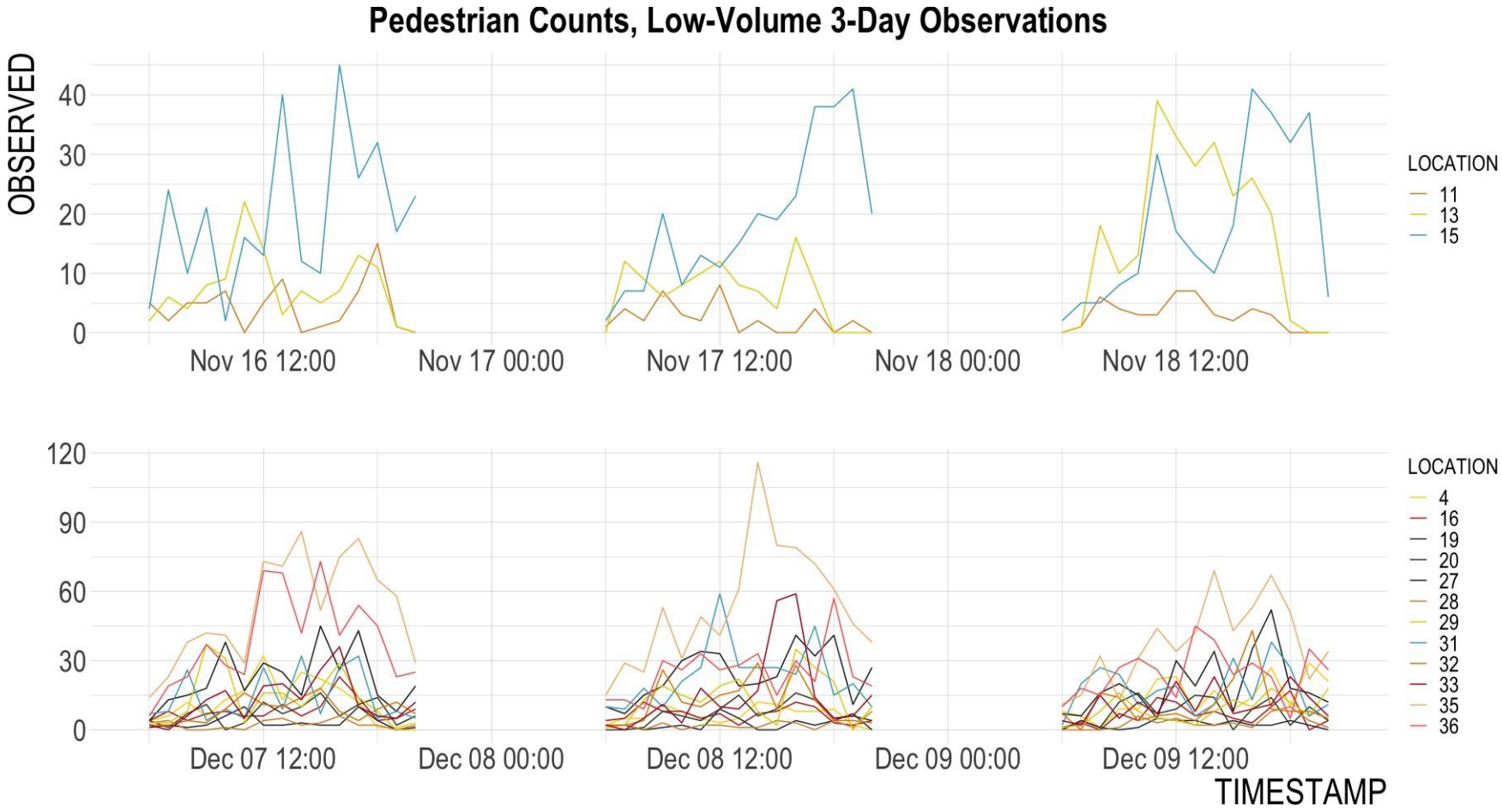


Figure 32: Graphs. Pedestrian counts at low-volume locations where three days of data were collected

tivity, especially in the early morning hours. The Krog Street location (which is a pedestrian scramble) (Figure) and the Northside Drive location (Figure), both of which are in Fulton County, both exhibit a drop in ATSPM activity in evening hours, and in both locations the ATSPM events only capture a small sample of the actual activity there. The Franklin Street location in Milledgeville (Figure) had similar levels of actuations and observations, but there is a clear outlier of observed data at 5PM, and the Watson Boulevard in Warner Robins location (Figure) exhibited a similar outlier volume of pedestrians at 11AM. Finally, the Oak Street location in Macon (Figure) had sporadic and limited pedestrian activity, and almost no ATSPM activity, except for two actuations at 12:00 on two different days.

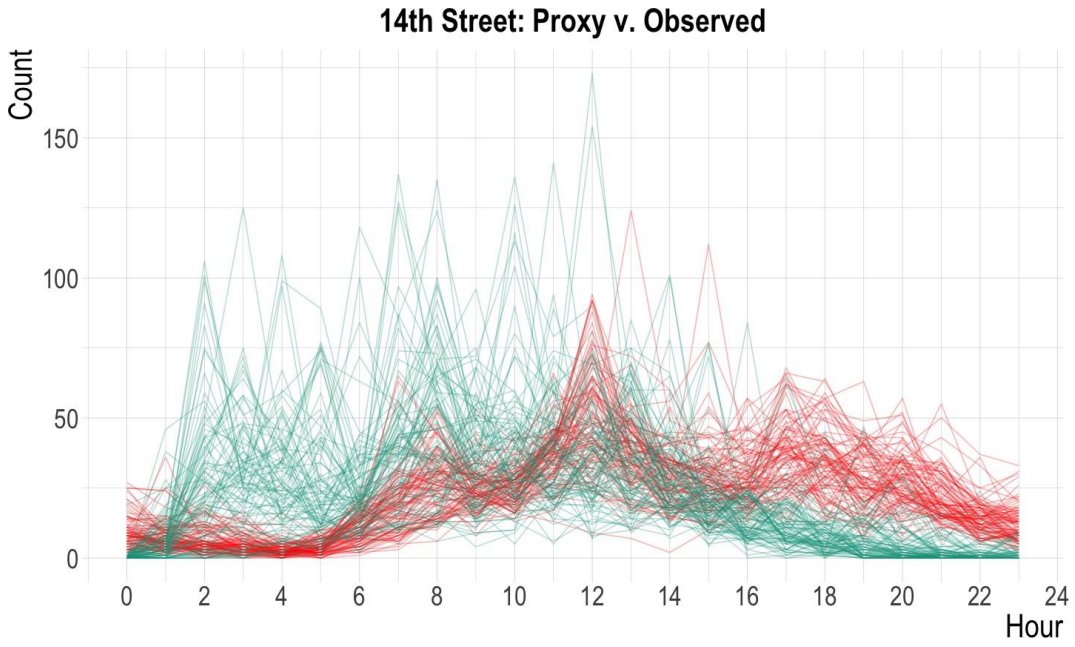


Figure 33: Graph. 90 days of observed pedestrian counts (red) and ATSPM data (light green) at 14th Street Location in Fulton County

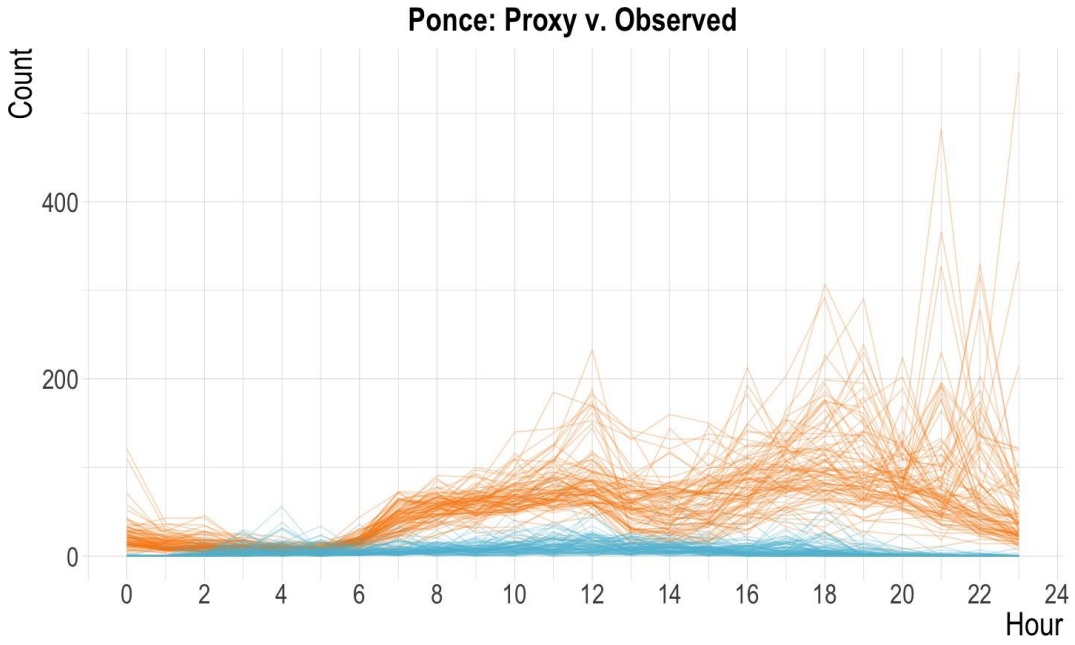


Figure 34: Graph. 90 days of observed pedestrian counts (orange) and ATSPM data (blue) at Ponce de Leon Avenue and Peachtree Street NE intersection in Fulton County

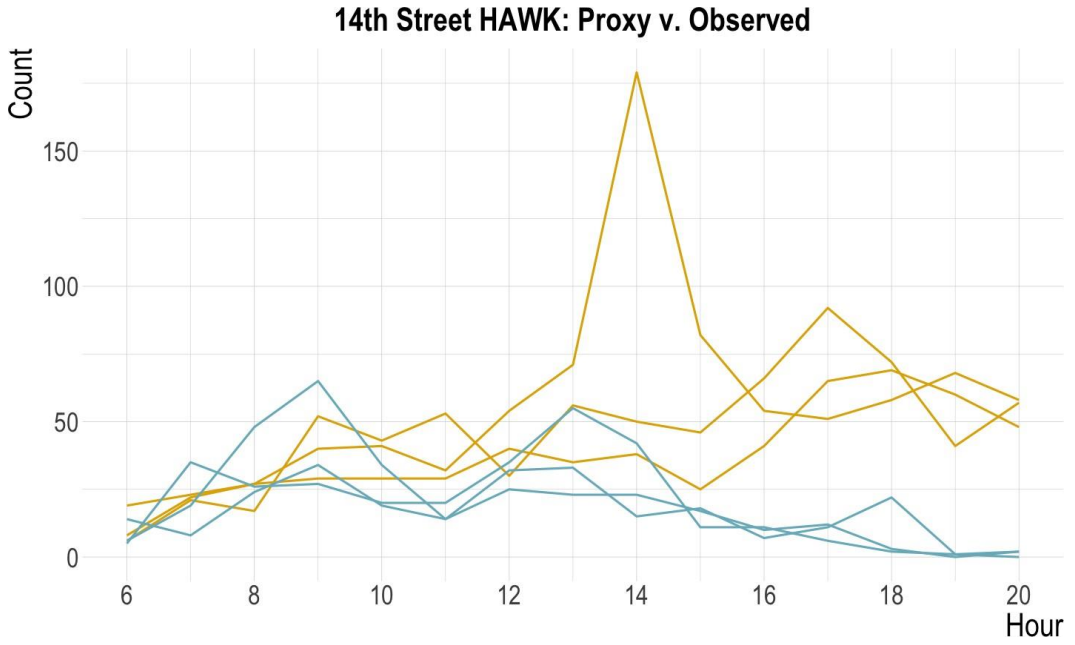


Figure 35: Graph. Three days of observed pedestrian counts (orange) and ATSPM data (blue) at 14th Street HAWK crossing in Fulton County

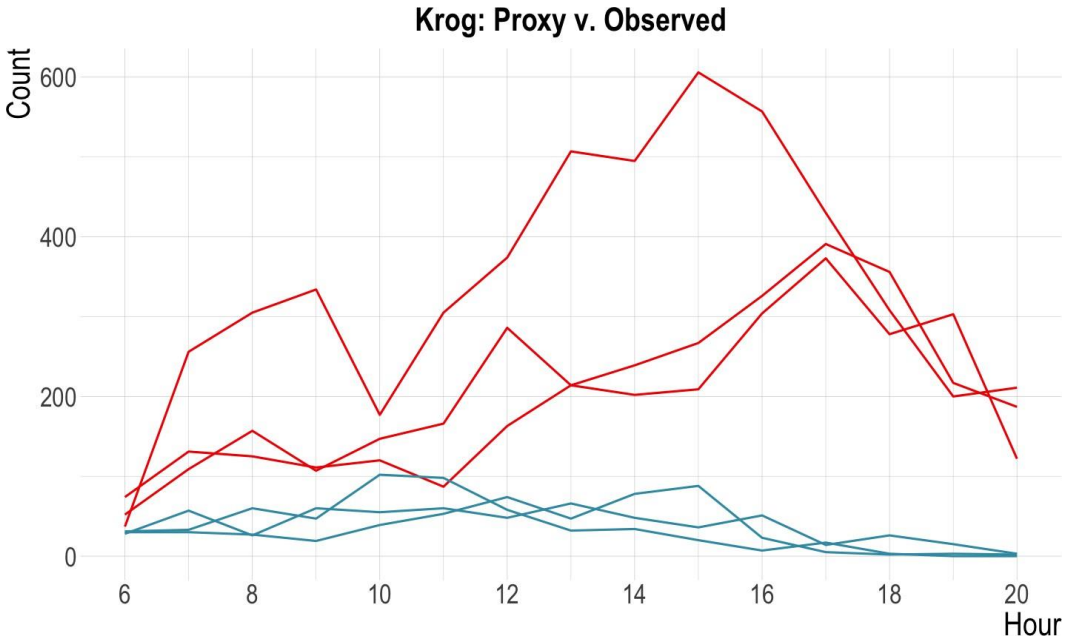


Figure 36: Graph. Three days of observed pedestrian counts (red) and ATSPM data (blue) at Krog Street Tunnel Scramble in Fulton County

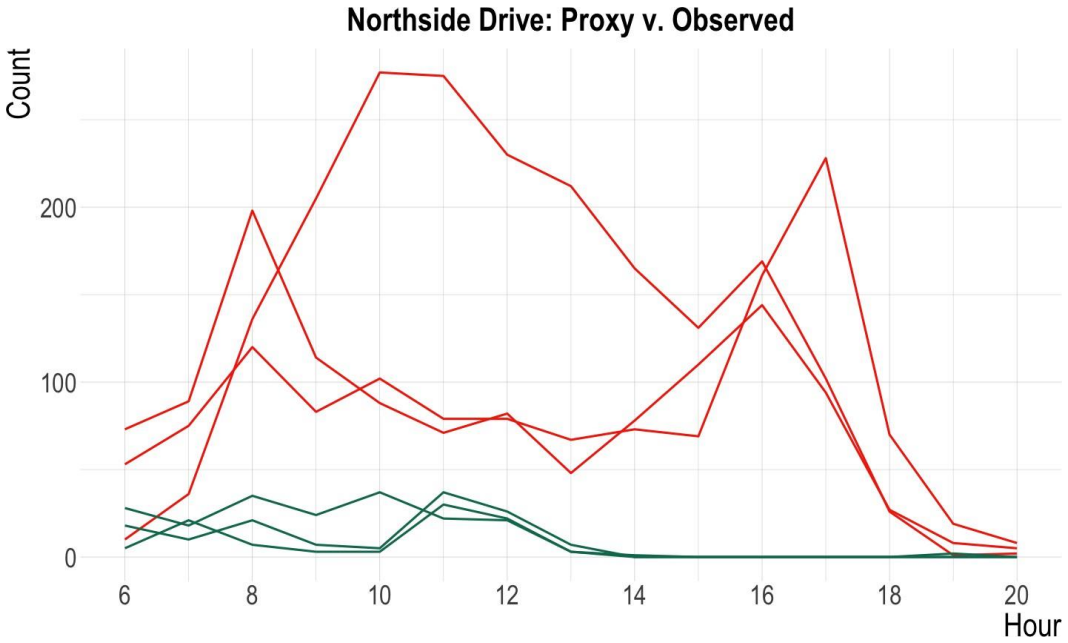


Figure 37: Graph. Three days of observed pedestrian counts (red) and ATSPM data (light green) at Northside Drive NW and Woodward Way NW intersection in Fulton County

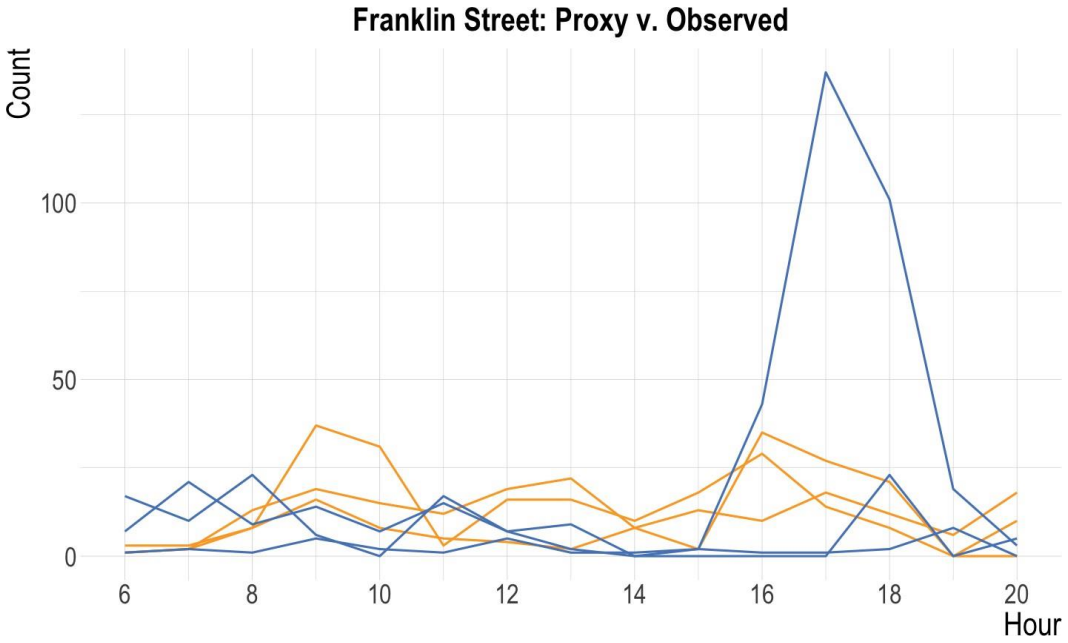


Figure 38: Graph. Three days of observed pedestrian counts (orange) and ATSPM data (blue) at Franklin Street and SR 243 in Milledgeville

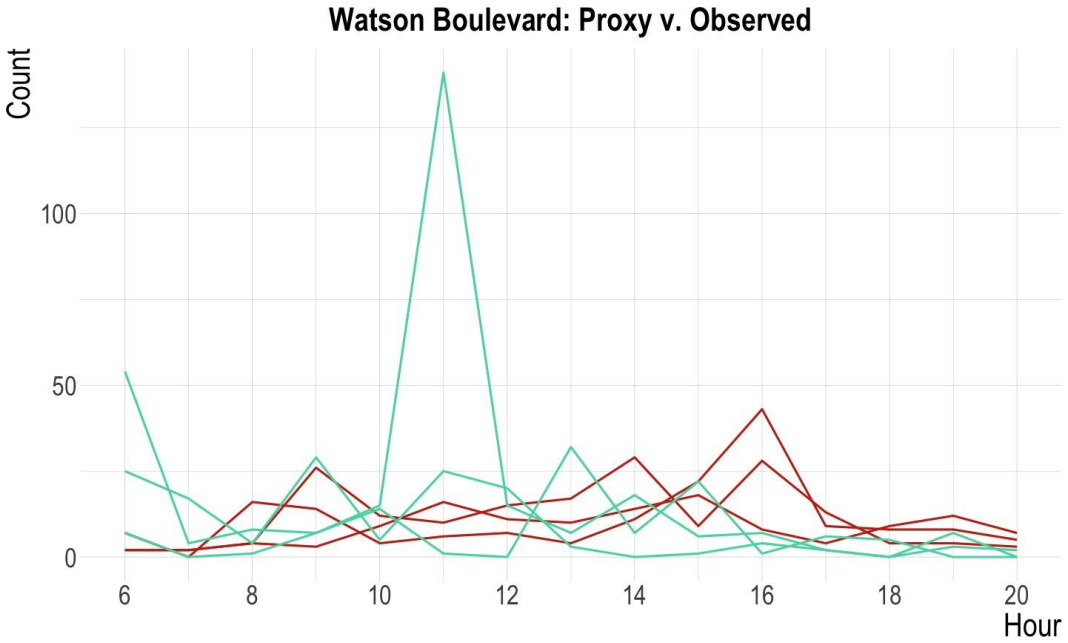


Figure 39: Graph. Three days of observed pedestrian counts (red) and ATSPM data (light green) at Watson Boulevard and Carl Vinson Parkway Warner Robins

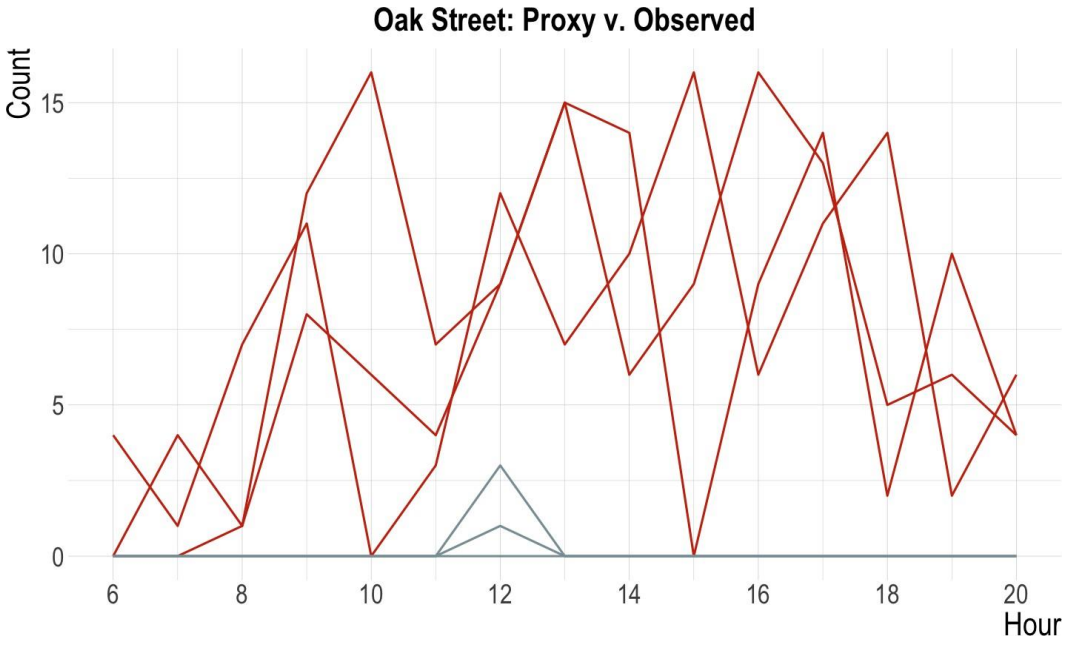


Figure 40: Graph. Three days of observed pedestrian counts (red) and ATSPM data (grey) at Oak Street and Martin Luther King Jr Boulevard in Macon

Observed Bicycle Counts and Strava Distributions

The previous section summarized both the observed pedestrian counts and the relationship between the observed counts and the ATSPM proxy data. The relationship between proxy and observed data at locations where bike data were collected were quite different than the pedestrian observations. A summary of the data collected at the 3-day observation locations is shown in Table . It is important to note that the unit of analysis for the bike data is different than the pedestrian data. All observation data were collected at the intersection level, where counts from all four legs of the intersection and sidewalks were gathered at disaggregate levels. While the pedestrian data is summarized to the intersection level, the bike data is parsed to the segment level to match Strava's unit of analysis. Data were collected at each intersection, which was then disaggregated to each roadway segment. To avoid spatial autocorrelation in the model, only the north/south or east/west direction (i.e., two parallel segments at an intersection) were used in each model. The directions that jointly had the most self-reported rides and the most observations were selected for analysis. The last column

in Table shows which model direction was selected for each location. The data collected at the 90-day locations were collected as segments, so selecting a direction to use for the models was not necessary. A summary of the 90-day data is shown in Table .

Table 13: Summary of the data from locations where three days of observations were collected

Location	North Segment Observed	South Segment Observed	East Segment Observed	West Segment Observed	North Segment Strava	South Segment Strava	East Segment Proxy	West Segment Proxy	Model Direction
9	800	149	647	31	10	135	0	150	N/S
10	61	60	9	13	45	60	15	0	N/S
12	46	4	73	76	0	5	5	5	E/W
13	32	52	0	5	5	5	0	0	N/S
14	80	97	63	28	5	0	0	5	N/S
37	1	0	132	172	0	0	5	0	E/W
40	4	14	0	0	5	5	0	0	N/S

Table 14: Summary of the data from locations where 90 days of observations were collected

Location ID	Strava	Observed Bicyclists
45	280	20647
46	3295	73421
47	385	30733

The distributions of the observed bicycle data are illustrated in Figure and Figure , where each day of data are plotted on top of each other to show the hourly trends. Figure shows different trends for locations 45-47. Location 46 has a great magnitude of activity than the other locations, and the trend is quite muddled during mid-day hours. Locations 45 and 47 have less overall activity and clearer trends during each day.

Figure captures the distribution of activity at locations where only three days of data were collected and where there was enough activity for the location to be included in teh the model. There appears to be a trend: at locations where there are higher volumes of bicycling, biking volumes increase hourly from 6:00AM until 4:00PM - 5:00PM, after which biking volumes decrease notably in evening hours of 6:00PM - 8:00PM. At the lower-volume locations, however, biking volumes were more sporadic and follow a less clear trend, pointing to potential difficulty in predicting levels of activity in these locations.

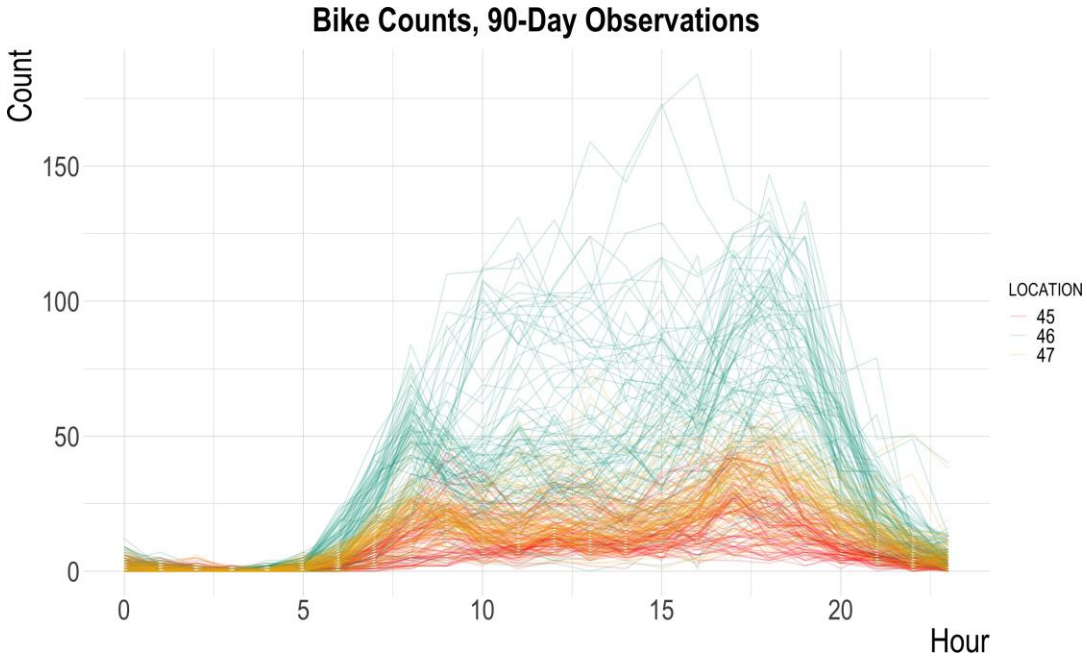


Figure 41: Graph. Observed bicycle activity trends at three locations where 90 days of data were collected

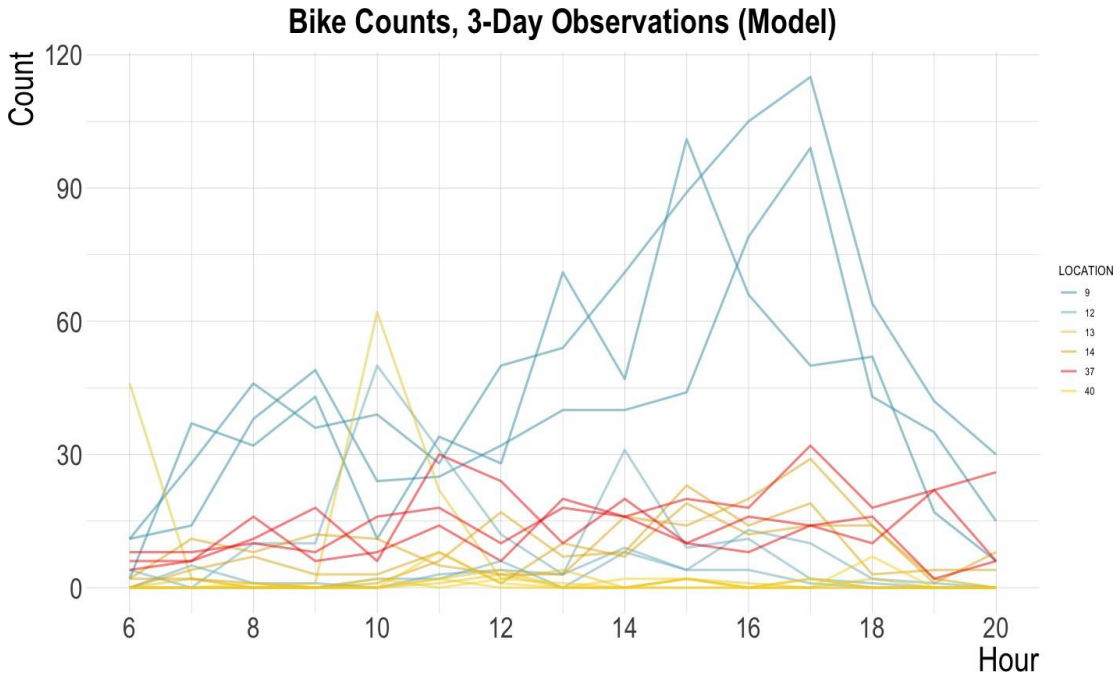


Figure 42: Graph. Observed activity trends at locations that were included in models where three days of data were collected

The next figures, Figure through Figure , show the distributions of Strava data during the same time periods as the observed data collection. In all of these figures, warm hued colors are observed data, and cool hued colors are Strava data during the same time and at corresponding Strava links. Figure and Figure were 90-day observation locations. Unlike the ATSPM data in the locations where the pedestrian volumes were created, the Strava data does not seem to follow the observations at these locations in terms of magnitude nor temporal trend. This seemed even more pronounced at locations where only three days of data are collected, as shown in Figure and Figure . While trends may be present in the observed data – there appeared to be a bimodal distribution of activity at Krog Street, with a small morning peak and larger evening peak, and an evening peak at the Cherokee Avenue location – Strava data did not represent many of the observed rides at these locations. This was especially true for the Cherokee Avenue location, where there were only 1-5 (Strava Metro reports rounded data for privacy purposes) rides recorded during one day of the observation period.

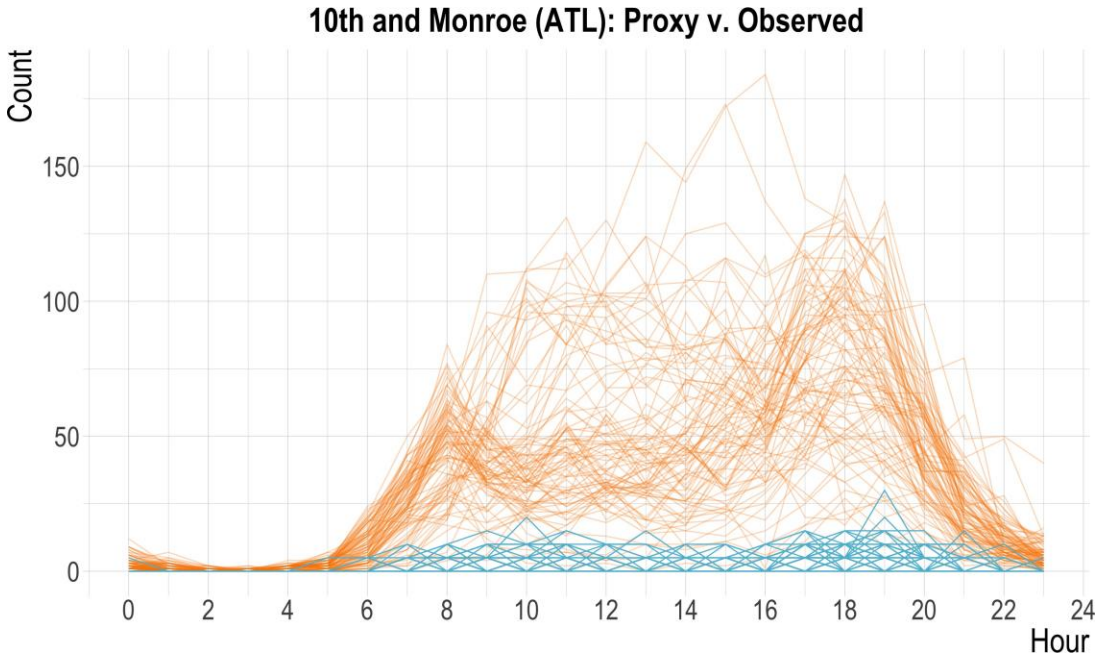


Figure 43: Graph. 90 days of observed bicycle counts (orange) and Strava data (blue) at 10th Street and Monroe Drive in Fulton County

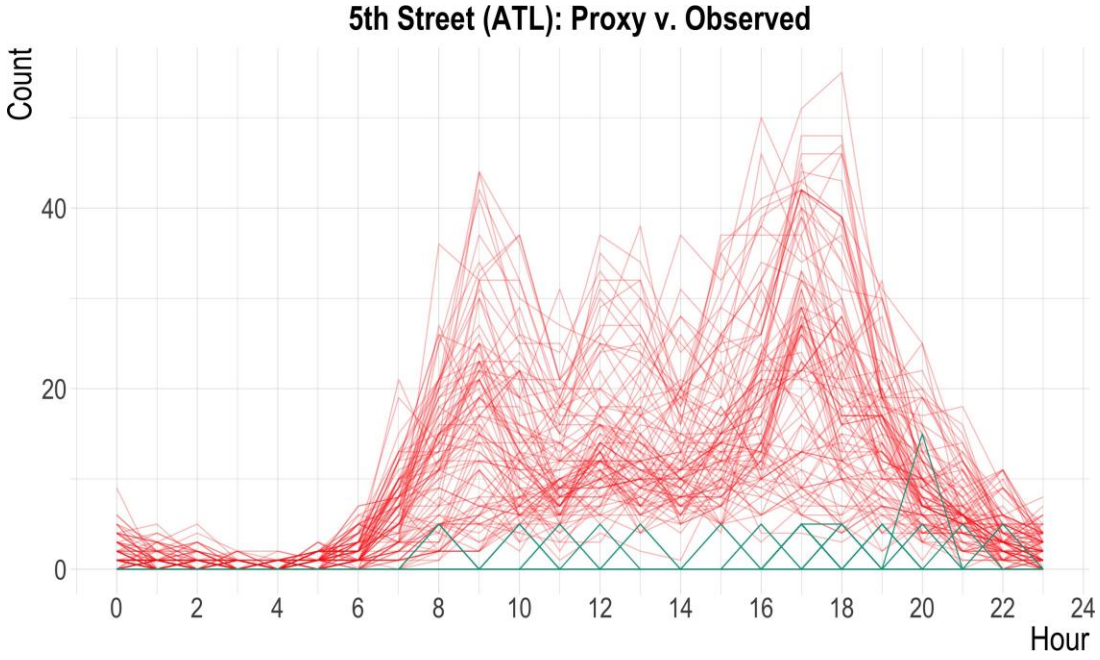


Figure 44: Graph. 90 days of observed bicycle counts (red) and Strava data (blue) at 5th Street in Fulton County

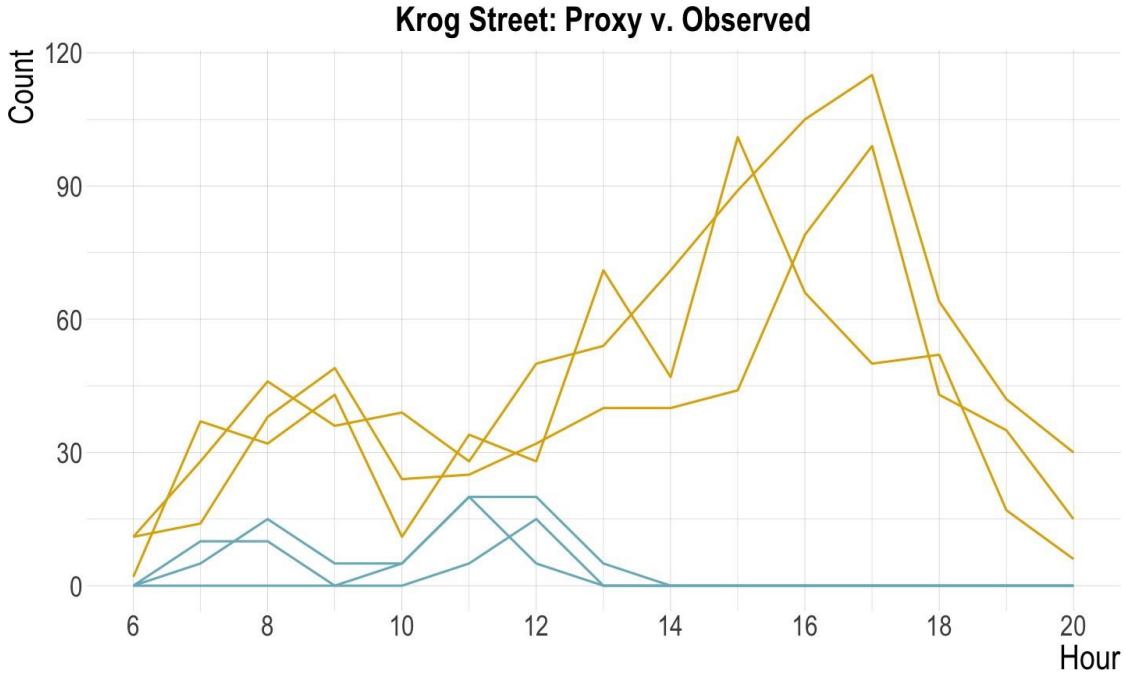


Figure 45: Graph. Three days of observed bicycle counts (orange) and Strava data (blue) at Krog Street and DeKalb Avenue in Fulton County

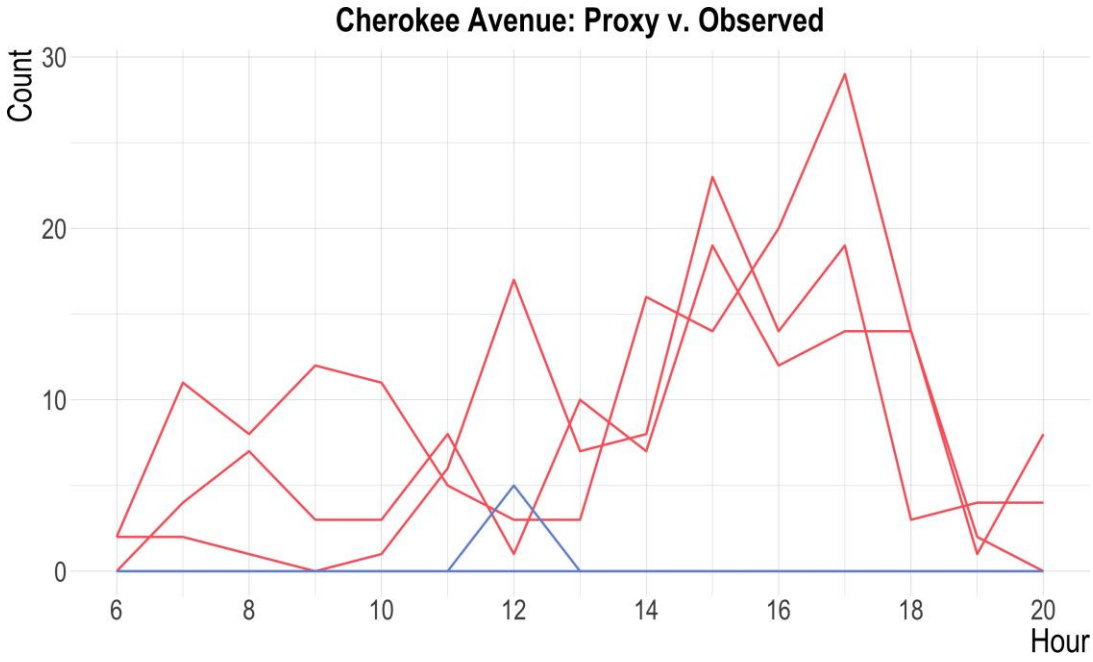


Figure 46: Graph. Three days of observed bicycle counts (red) and Strava data (blue) at Cherokee Avenue in Fulton County

The following two sections describe the results from model predictions of pedestrian and bike volumes using different models but similar specifications, starting with the pedestrian models and predictions.

Pedestrian Volume Estimation Results

Base Specification Predictions and Performance

The base specification for the pedestrian models is $\lambda\ \alpha + \beta_p P * L + f(H)_L$, where λ is the predicted arrival rate based on the ATSPM proxy data (P1 metric), β_p is the proxy mutlipier which is allowed to vary by location L, and where $f(H)_L$ is a spline variable that varies by hour and location, regressed on the observed data. This is the simplest model form used in this work. Figure through Figure show predictive posterior predictions (in purple) at several locations with the corresponding observations for each location in light green. These figures were made from the base specification trained on only the training subset of data, and the observations were those

that correspond to the same time period as the training data. Also note that data shown in the figures are specific to one location but the predictions are made using data from all locations that were included in the pedestrian models. In addition, uncertainty is shown in each figure. Like all other elements of Bayesian models, uncertainty is generally expressed in terms of probability distributions. Uncertainty is typically reported as a "credibility interval," which is the proportion of the posterior probability mass for the predictions. This work used a 90% credibility interval, which is common practice. This means that 90% of the posterior probability mass was captured within certain boundaries, which are reported in each figure showing predictions. Note that table of regression coefficients are not reported here. This is because in Bayesian models the parameters are not independent of each other so they cannot be independently interpreted [60]. Instead, posterior predictions were used to interpret the models results and predictive abilities.

Figure shows a distribution of 78 prediction lines (that correspond with the 78 days of observations included in the training data) plotted on top of the training data from the same location, where the 90% credibility interval is shown in translucent purple. At this location, the predictions shrunk to the mean of each hour and clearly capture the hourly trend throughout the day. The predicted trend was not affected by the more highly variable observations in the evening hours at this location. Also, many of the observations fall within the credibility interval, which indicated a more realistic posterior predictive distribution.

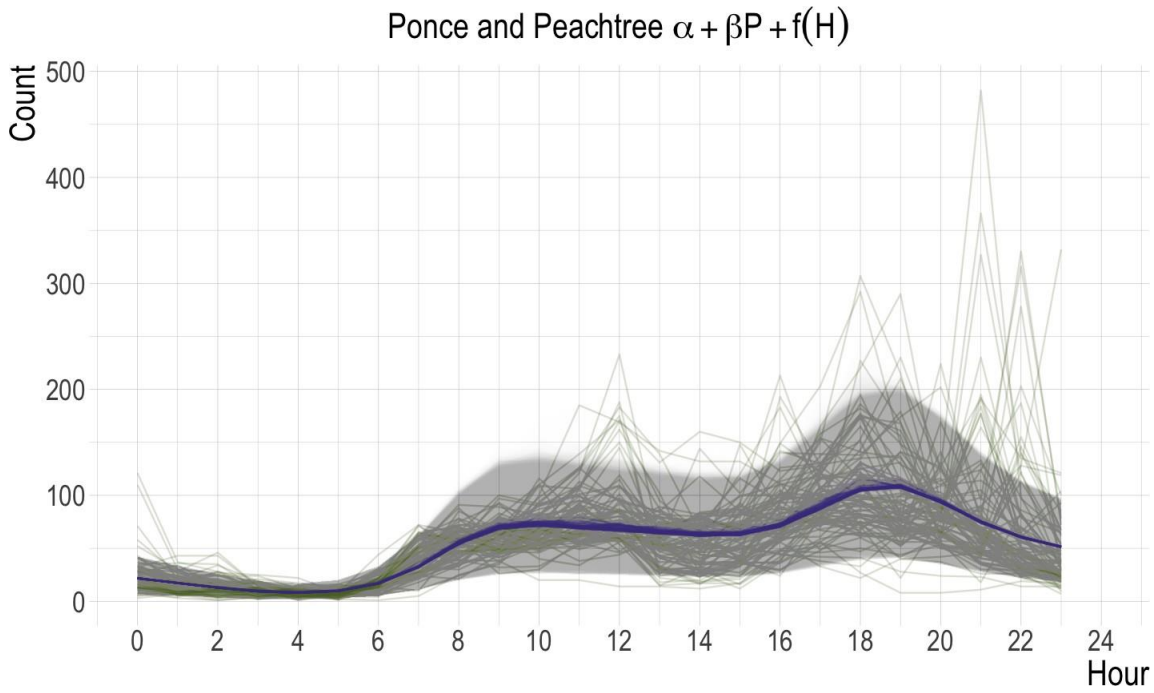


Figure 47: Graph. Ponce de Leon Avenue in Fulton County predicted (purple) versus observed (light green) pedestrian volumes using the base model specification with a 90% credibility region (translucent purple)

Figure and Figure show the same visualization and base specification predictions applied to the Krog Street scramble in Fulton County and the Handcock Street location in Milledgeville, respectively. While the location in Figure shows 78 days of observation at one location, these locations only have three days of observation (and one of the days, Friday, is held out of the model as testing data). The two days were plotted in green, the two corresponding predictions were shown in purple, and the uncertainty was shown in translucent purple. While the predictions appeared to follow the trend of the observations and appear to be capturing the average of the activity at each location, the credibility intervals were much wider in these instances, reflecting much more uncertainty in these predictions. That was expected in part, as there was much less evidence in this location compared to the Ponce de Leon location, but the credibility interval is slightly larger than the range of observed values in most cases, indicating less certainty around the predictions and posterior prediction that may not be precise enough for application in all cases.

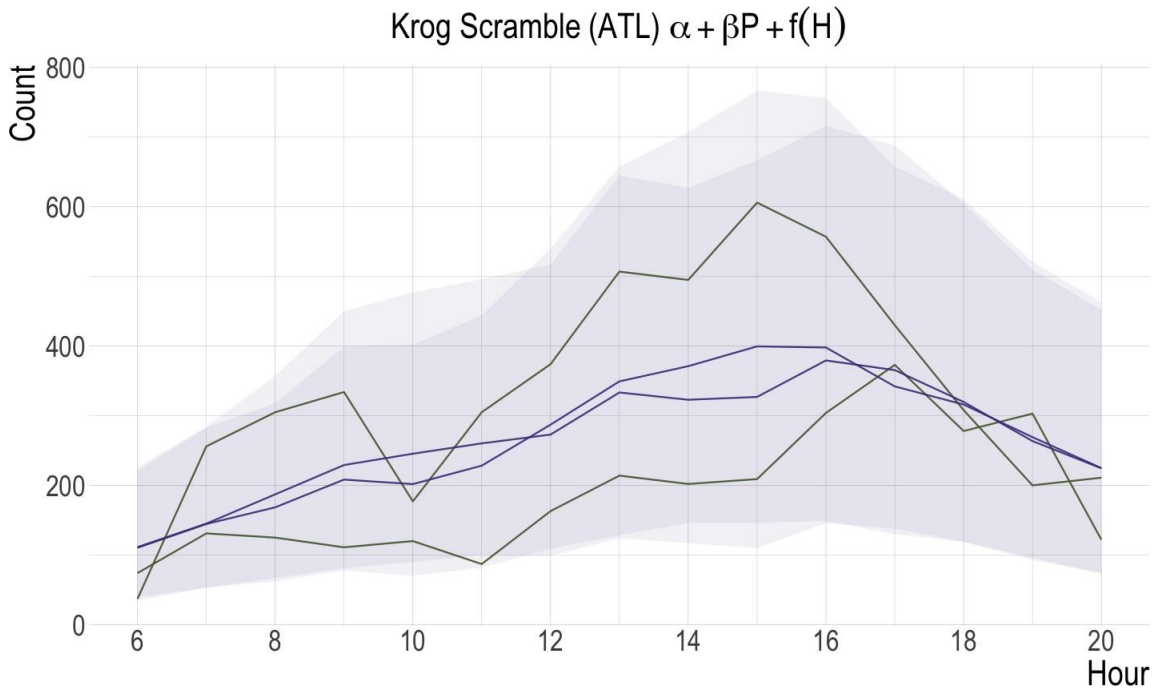


Figure 48: Graph. Krog Street scramble in Fulton County predicted (purple) versus observed (light green) pedestrian volumes using the base model specification with a 90% credibility region (translucent purple)

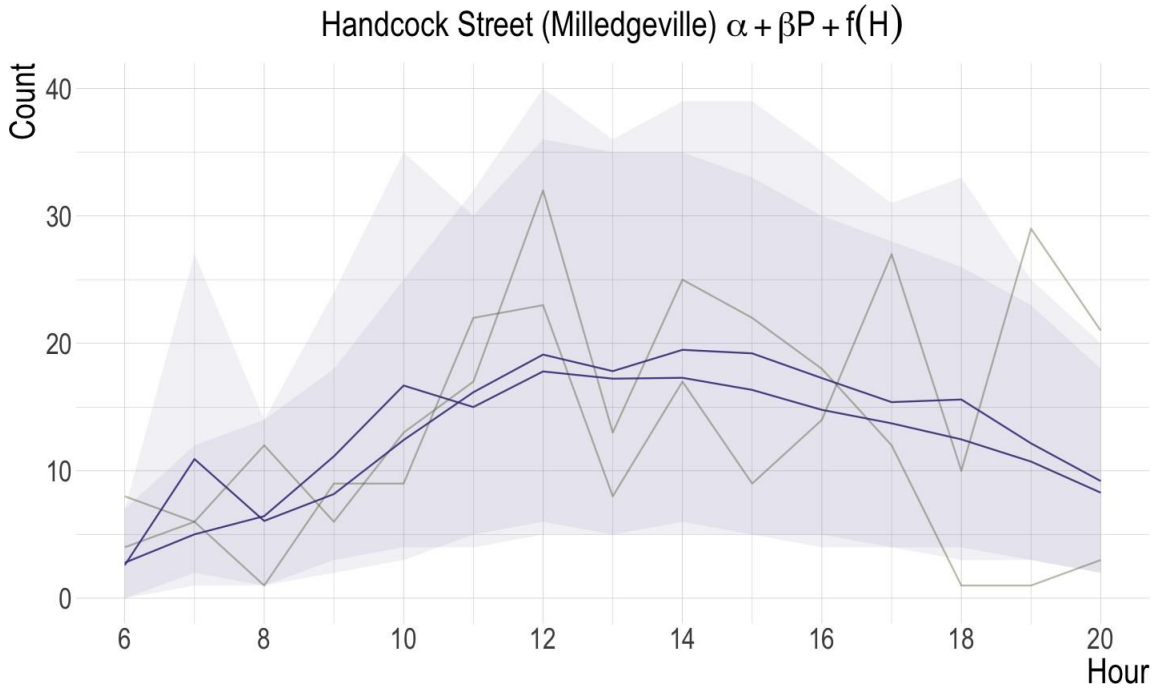


Figure 49: Graph. Handcock Street scramble in Milledgeville predicted (purple) versus observed (light green) pedestrian volumes using the base model specification with a 90% credibility region (translucent purple)

Next, this work more directly assessed the relationships among the average hourly trends of the predicted and observed data. The first two figures, Figure and Figure , show the predicted hourly means compared to the training data's hourly means, where the high-volume locations and low-volume locations were separated for ease of visualization. Some of the locations in Figure were locations where 90 days of data were collected (so there were observations represented at all 24 hours), and most of the locations were places where only 15 hours per day of data were collected. As expected, the high-volume locations' predictions clearly match the observed predictions, and for the most part, this was the case for the low-volume locations, with one exception. Location 9, which is the Krog Street location, has quite different trends of activity among two training days of data, so the predictions do not seem to well match at that location.

The next two figures illustrate the predictions averaged to the hourly level compared to the testing data. Figure showed the high-volume locations, and Figure showed the low-volume lo-

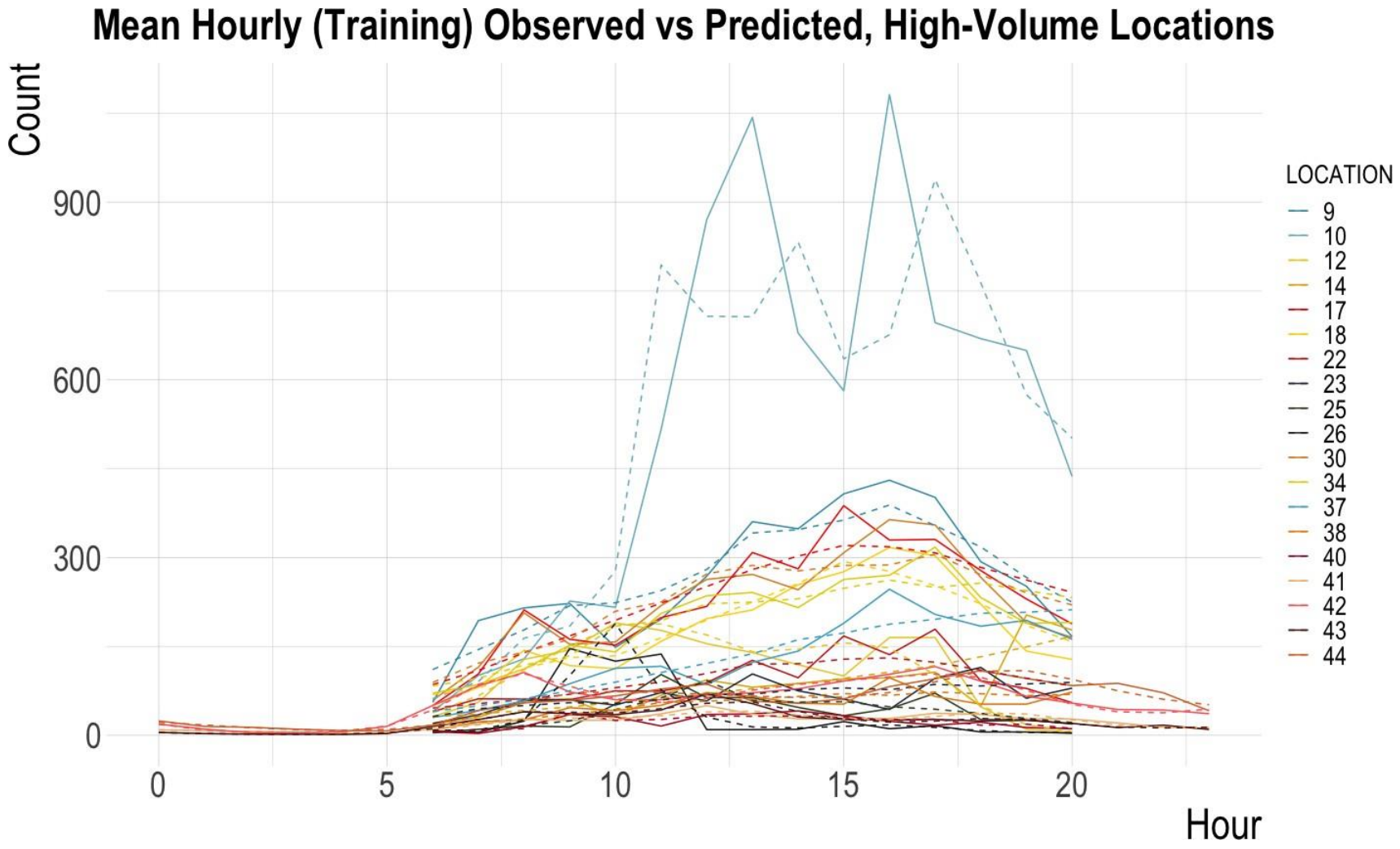


Figure 50: Graph. Average hourly predicted (dashed line) and Observed (solid line) trends at high-volume locations, using training data only

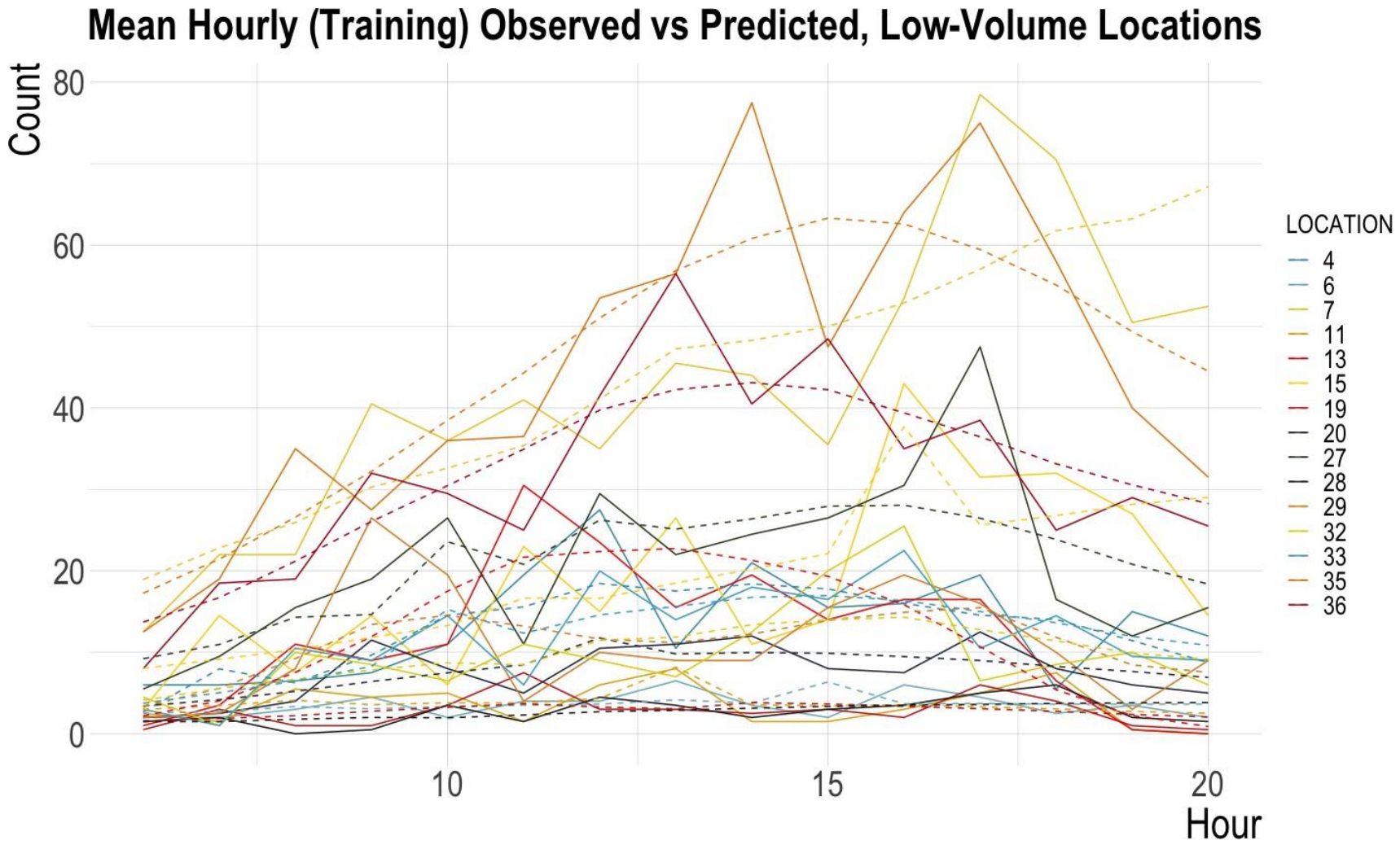


Figure 51: Graph. Average hourly predicted (dashed line) and observed (solid line) trends at low-volume locations, using training data only

cations. Again location 9, Krog Street, did not match well given the high variability among the days of observation. The same was true for locations 7 and 11 in Figure , which were the 14th Street HAWK signal in Fulton County and the North Point Parkway intersection in Fulton County, respectively.

Finally, Figure shows the in-sample and out-of-sample deviance ($-2 * LPPD$) for base specification at each location. The in-sample deviance for each location is represented by the smaller, translucent circles, and the out-of-sample deviance is represented by the larger, solid circles. Because the deviance is a measure of distance from predicted values at each location, this figure illustrates the predictive power at each location for both in-sample predictions and out-of-sample predictions.

The results showed a clear trend: that the strength of predictive power for in-sample versus out-of-sample predictions are relatively similar for most locations, but that predictive power generally is not consistent across each location. It was expected that locations 41-44 would have very similar deviance values between training versus testing, as these were locations where there were much more data than the others. Interestingly, though, these were not the locations with the lowest deviance measures overall. Locations 19 (and Eisenhower Parkway intersection in Macon) and 28 (Russel Parkway in Warner Robins) have low deviance values in- and out-of-sample compared to the rest of these locations. When comparing the ATSPM data and the observed data at each of these locations, it appeared that there were similar levels and trends of activity among all days monitored, and that the trends of ATSPM data were also the same at each location. So, in places where the trends are similar over time, it appears that less data is needed to reduce deviance.

Another clear trend from this is that deviance appeared to be more related to the location itself than the base specification model's predictive power. In each case, the difference between the in-sample deviance and out-of-sample deviance were relatively small, but the magnitude of deviance varies greatly by location. So, some locations appeared to be more difficult to predict than others, perhaps due to characteristics that are not captured in the base specification. The next section investigates iterations of the base specification to see if introducing new variables or

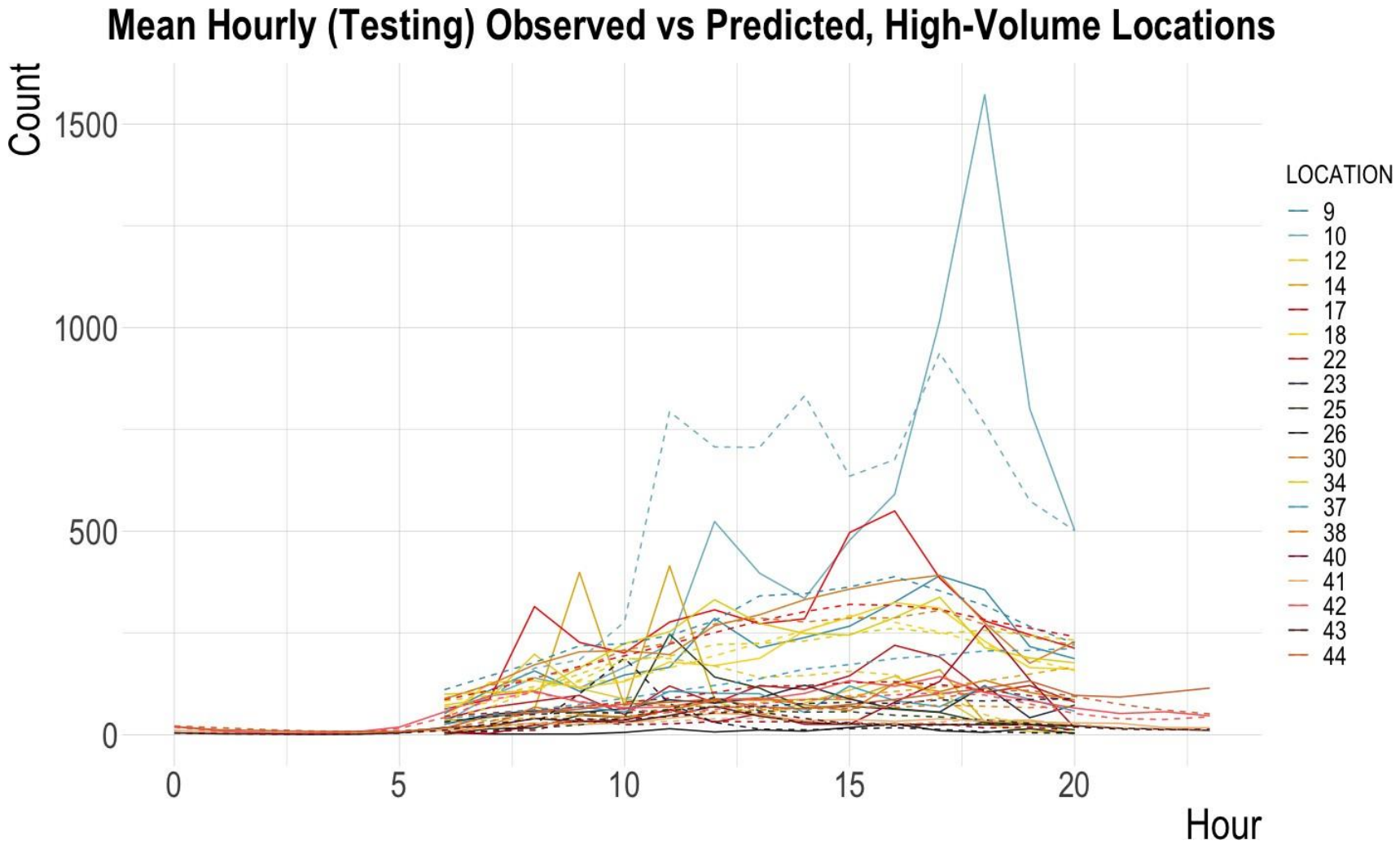


Figure 52: Graph. Average hourly predicted (dashed line) and observed (solid line) trends at high-volume locations, using testing data observations

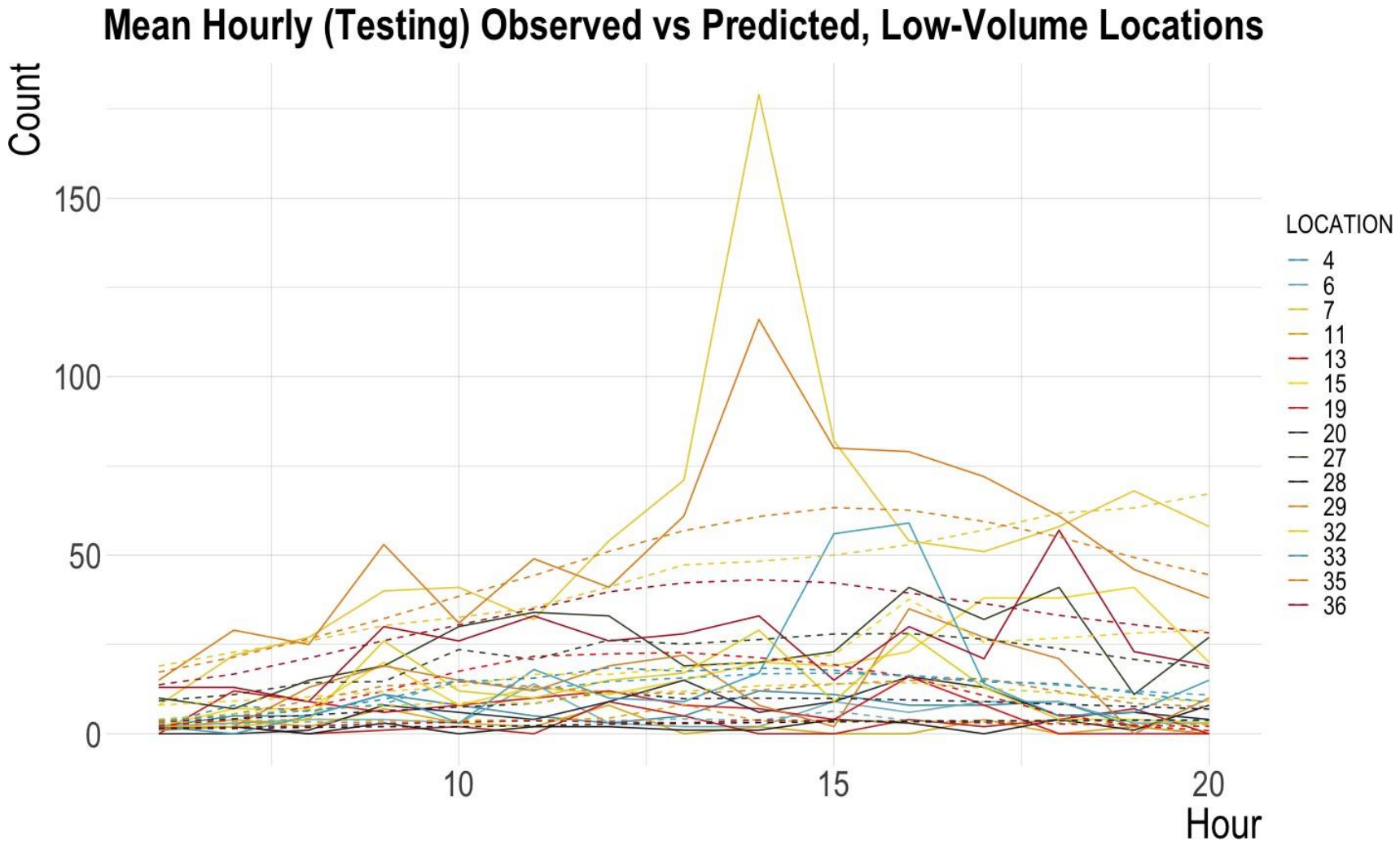


Figure 53: Graph. Average hourly predicted (dashed line) and Observed (solid line) trends at low-volume locations, using testing data observations

hyperparameterization can reduce deviance at some locations.

Also, the MAE (which was selected over RMSE in this case because of its resilience to outliers) was reported for each location in Figure , where the smaller translucent circle is the in-sample MAE and the larger circle is the out-of-sample MAE. While the deviance graphic shows that the base specification's fit is similar in-sample and out-of-sample, the MAE graphic shows how well the model predicts in terms of absolute error at each location. Noting the scale break in the y axis, the results show that in only two places does the model have extremely large error – location 10, which is Marietta Street NW in north suburban Atlanta, and location 25, Virginia Avenue in south suburban Atlanta. In many cases, both in-and out-of-sample MAE were quite low, less than 25 pedestrians per hour. There were also several instances of MAE between 25 and 75. It appears that, like the deviance graphic above, in most cases (except locations and 10 and 25) MAE is similar for in-sample versus out-of-sample and appears to vary more by location.

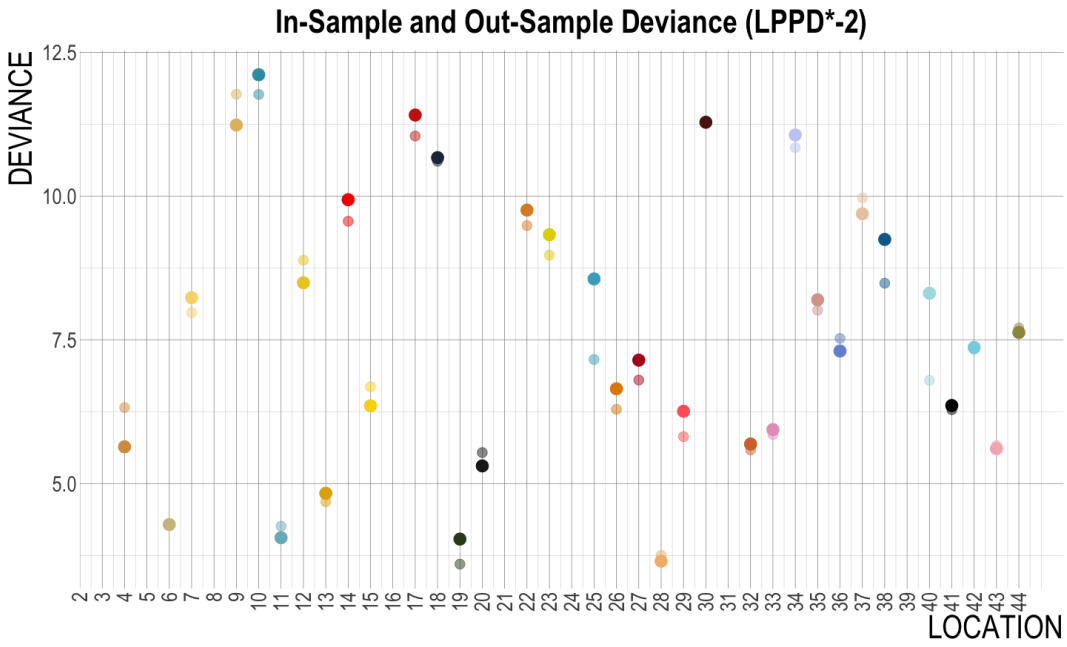


Figure 54: Graph. Base specification in-sample deviance (smaller, translucent circles) and out-of-sample deviance (larger, solid circles)

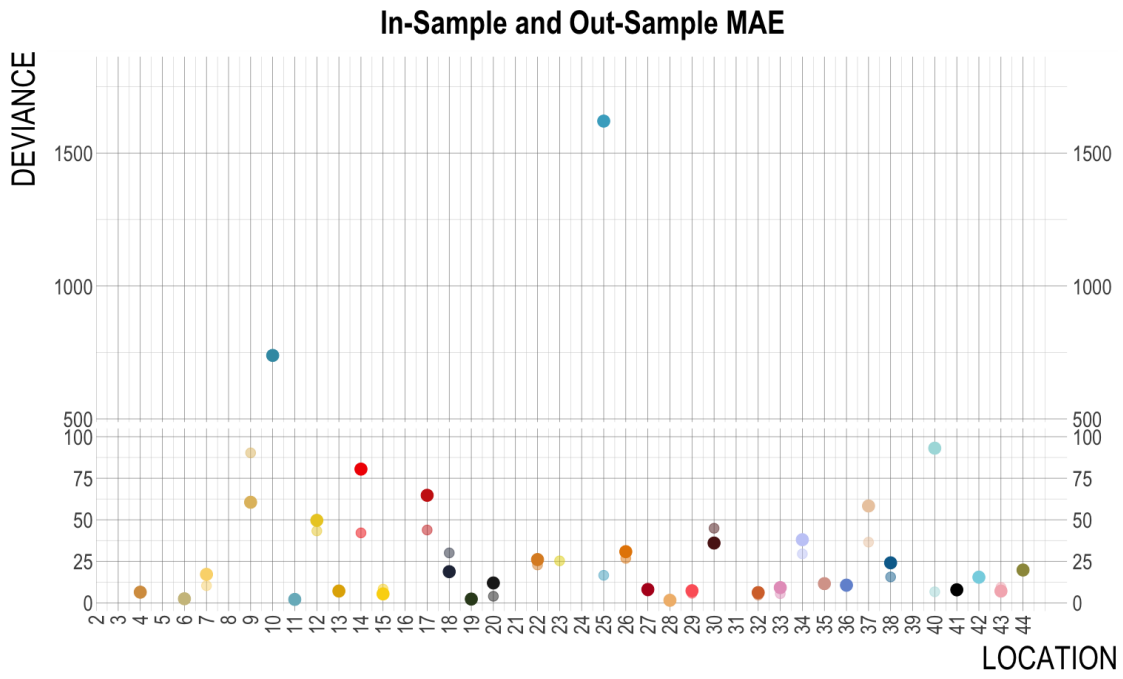


Figure 55: Graph. Base specification in-sample MAE (smaller, translucent circles) and out-of-sample MAE (larger, solid circles)

Additional Specifications and Predictive Performance

This chapter also explored other variables and different specifications and their ability to improve predictive performance from the simplest of specifications. The specifications that were tested on the pedestrian data were summarized in Table . PM1 added a categorical variable for the high-volume versus low-volume locations, which was determined subjectively from viewing the data. PM2 iterated on the base specification by adding a variable for cluster assignment based on the high-performing clusterings described in Chapter 3. In this case, clustering assignments for each intersection were assigned based on the clustering result of event #90, or pedestrian metric 1 (P1). The fourth pedestrian metric's feature-based clustering actually performed better than the first metric in terms of cluster cohesiveness and compactness, but as noted in Chapter 3, not all signals capture event #45, so in this case, the first metric was chosen to ensure that all models were assessed on the same data. This clustering resulted in four clusters, which were added to PM2 as a categorical variable. PM3 introduced a hyperparameter $B_P \sim Normal(3, \sigma_P)$, where

σ_P was assumed to have the distribution *Exponential*(1) [60]. PM4 added one more hyperparameterization in addition to PM3, which was applied to the intercept $\alpha \sim Normal(0, \sigma_a)$, where σ_a was also given the distribution *Exponential*(1).

Table 15: Model specifications applied to the pedestrian data

Model Name	Specification	Description
Base Specification M0	$\lambda = \alpha + \beta_P * L + f(H)$	Only includes variable for proxy plus intercept and temporal correlation control
Model 1 PM1	$\lambda = \alpha + \beta_P * L + \beta_V + f(H)$	Adds a categorical variable for high vs. low-volume locations
Model 2 PM2	$\lambda = \alpha + \beta_P * L + \beta_C + f(H)$	Adds a categorical variable for cluster assignment
Model 3 PM3	$\lambda = \alpha + \beta_{P\ hyperprior} P * L + f_i(H)$	Adds a hyperparameter to the the proxy multiplier
Model 4 PM4	$\lambda = \alpha_{hyperprior} + \beta_{P\ hyperprior} P * L + f_i(H)$	Adds a hyperparameter to the intercept and the proxy multiplier

Each of these models were trained on all data. For each of the additional models tested in Table), several metrics of predictive performance – LOOCV-LPPD and WAIC were assessed. Additionally, LOOCV-LPPD standard error, effective number of parameters based on LOOCV-LPPD, WAIC standard error, and the effective number of parameters based on WAIC were also reported in Table , starting with the base specification (that is used in both bicycle and pedestrian models) M0, to pedestrian model 4 (PM4) which is the most complex specification.

Table 16: LOOCV-LPPD and WAIC results from each model specification applied to the pedestrian data

Model	LOOCV-LPPD	LPPD Standard Error	LPPD Eff # Parameters	WAIC	WAIC Standard Error	WAIC Eff # Parameters
M0	-40123.7	144.7	260.6	80231.4	289.1	252.6
PM1	-40111.5	144.6	272.9	80206.4	288.7	264.3
PM2	-40111.9	144.6	273.1	80204.1	288.8	263.3
PM3	-40098.2	144.7	274.3	80176.2	288.8	264.2
PM4	-40097.6	144.8	274.2	80174.7	289.0	263.9

The model performance metrics indicated several findings that elucidated what kinds of specification improved predictive performance for Bayesian models applied to pedestrian volume esti-

mation. First, as model complexity grew, WAIC decreased and LPPD became less negative, both

of which indicate an increase in predictive performance. Between M0 and PM1, adding a categorical variable for high-volume versus low-volume segments did slightly improve model predictive power, but this improvement was small. PM2 did improve predictions over the base specification slightly, indicating that cluster assignment may be used to improve predictions of walking, but the performance improvements were overshadowed by PM3 and PM4. These models notably improved upon the base specification's information metrics. When comparing PM3 and PM4, which were differentiated by a hyperprior applied to proxy variable (PM3) and to both the proxy variable and the intercept (PM4), PM4 showed modest increases in LOOCV-LPPD but an increase in WAIC. These increases are modest, though. Depending on the application of the models in practice or future research, it may be beneficial to stay with the simpler specification (M0) than to introduce the model complexity of hyperparameters. Given the slight increase in PM4's WAIC value, likely PM3 would be the preferred specification.

Based on the best performing model specification, PM3, Figure presented the full model deviance at each location. When compared to Figure that shows the in- and out-of-sample deviance at each location for the base specification, clearly the best performing model reduced overall deviance at each location, but it did not change relative deviance among the locations; the locations with the largest deviance values in the base specification still had the largest deviance values in the best performing specification in terms of model fit. The more complex specification improved prediction overall by pooling information among locations, but additional information would be needed to reduce location-specific deviance. RMSE and MAE are illustrated in Figure . While there were small differences in these measures of absolute error compared to the base-specification, in general the results were comparable to the base-specification. This was somewhat surprising; it was expected that an increase in model fit would result in lower error per location. In most cases, error in the pooled specification were only modestly better than the base-specification. This indicates that a better model fit overall may not indicate consistently better predictions at each specific location.

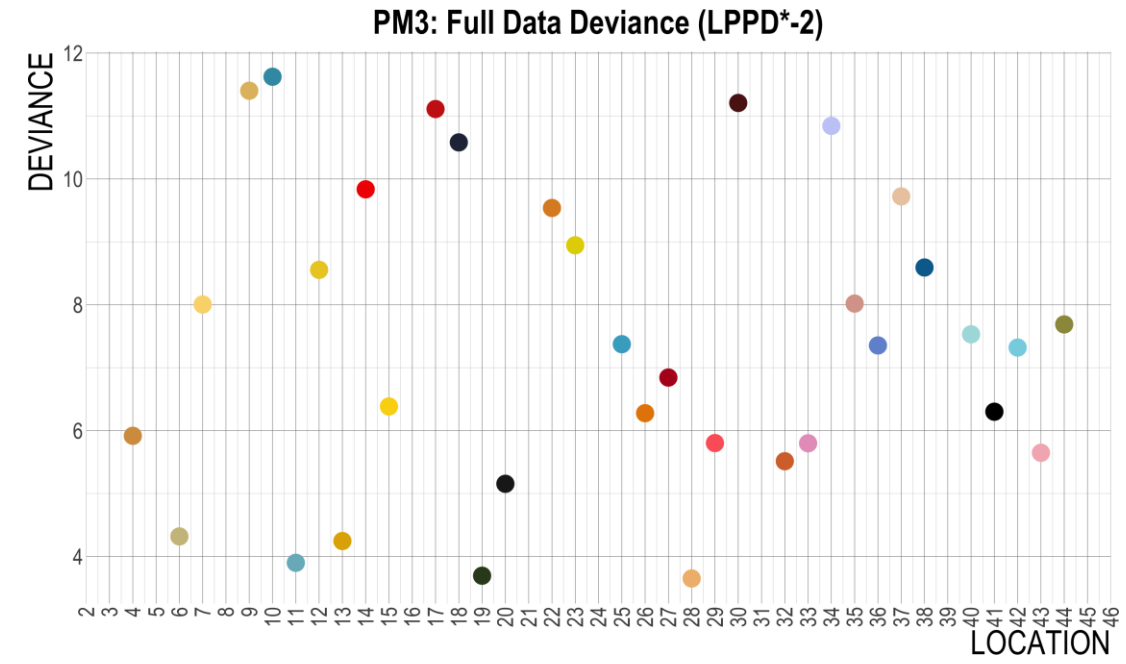


Figure 56: Graph. PM3 specification full model deviance by location

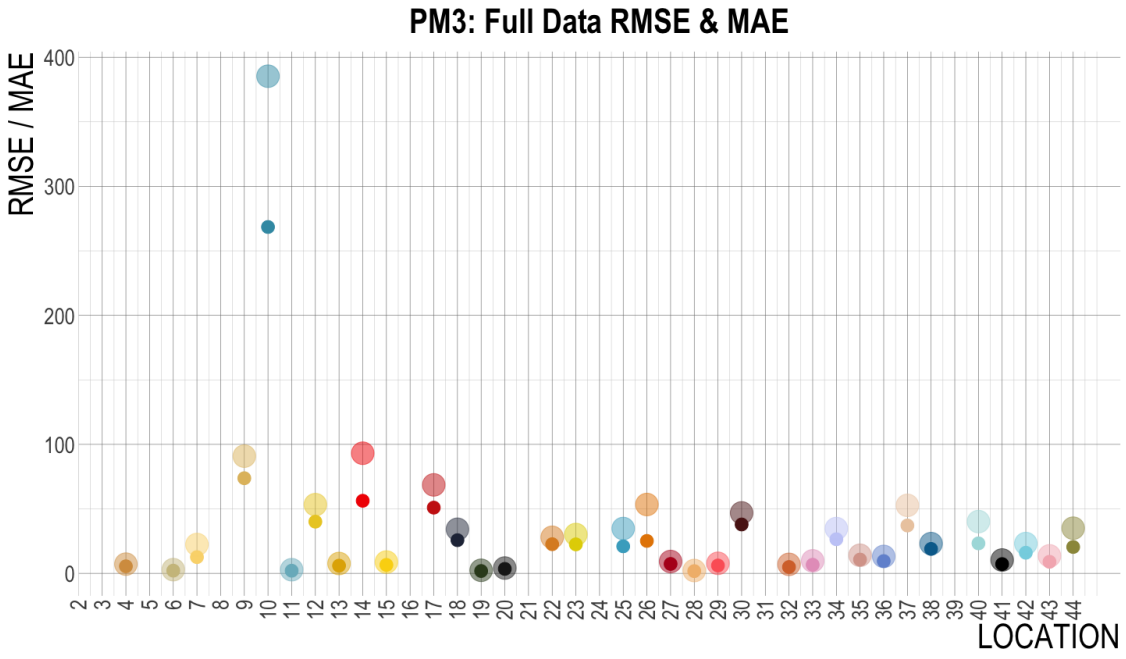


Figure 57: Graph. PM3 specification full model RMSE (large translucent circle) and MAE (small solid circle) by location

It was worth viewing predictions at locations with the highest and lowest deviance. Figure

and Figure shows the two locations with the greatest deviance values, at the intersections of Marietta Street NW and Baker Street NW, and MLK Drive NW and Joseph E Lowery Boulevard NW, respectively. Both of these locations were in Fulton County. The next two figures, Figure and Figure show the two locations with the smallest deviance values, which were at the intersections of Russel Parkway and Davis Drive in Warner Robins, and Eisenhower Parkway and C Street in Macon. Note that the magnitude of pedestrian activity is much lower at these locations. Interestingly, the locations with the highest deviance values were those with large ranges of pedestrian activity throughout the day. The Marietta Street location ranged from 0 to nearly 2000 pedestrians per hour during the 15 hours of observations, and the MLK Drive location ranged from 100 pedestrians to over 500 pedestrians per hour. The locations with low deviance values had much smaller ranges that the model could more accurately predict, ranging from 0 to 10 pedestrians per hour at each of the low deviance locations. The first two locations had much clearer trends in the observed data, however, and their means more closely followed those trends than those in the low-deviance locations.

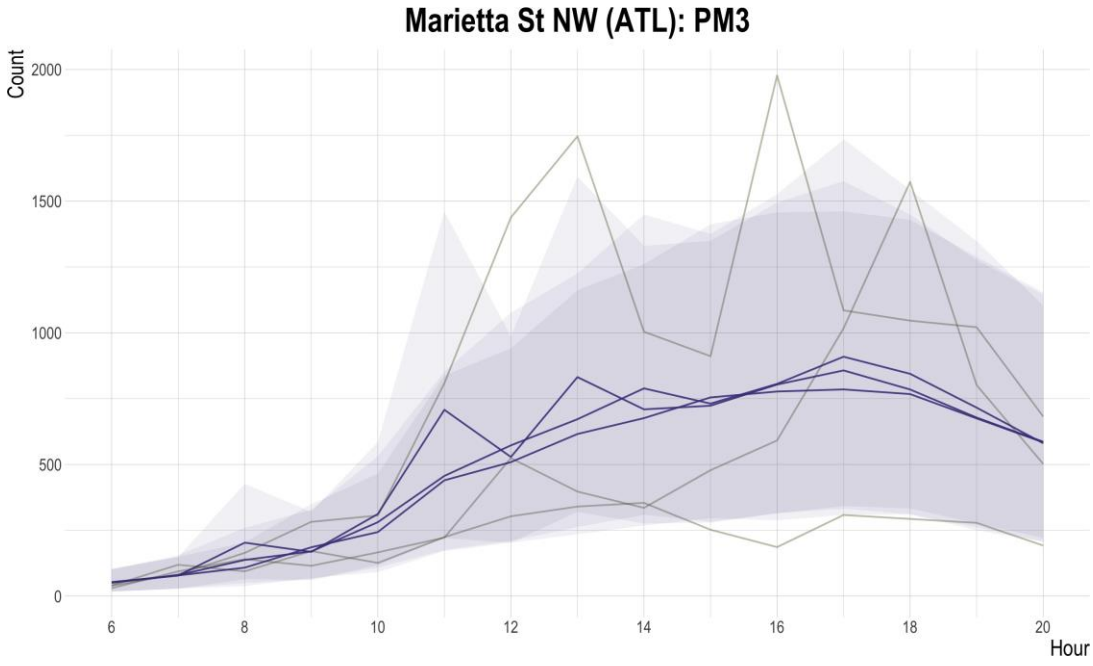


Figure 58: Graph. PM3 specification applied to make predictions at the intersection of Marietta Street NW and Baker Street NW in Fulton County

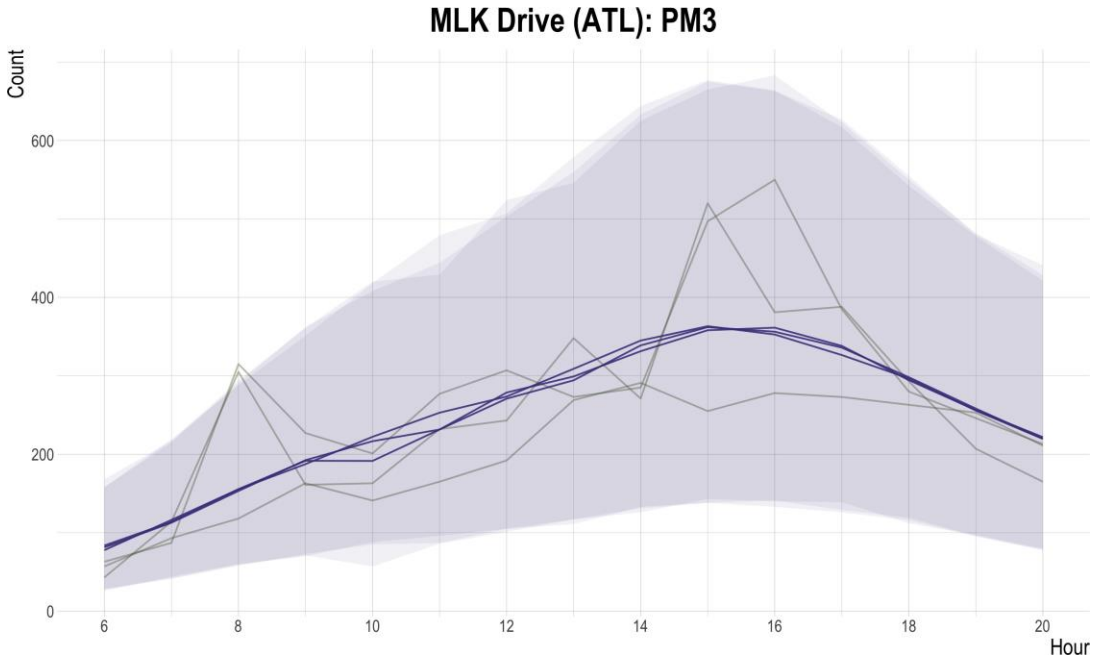


Figure 59: Graph. PM3 specification applied to make predictions at the intersection of MLK Drive NW and Joseph E Lowery Boulevard NW in Fulton County

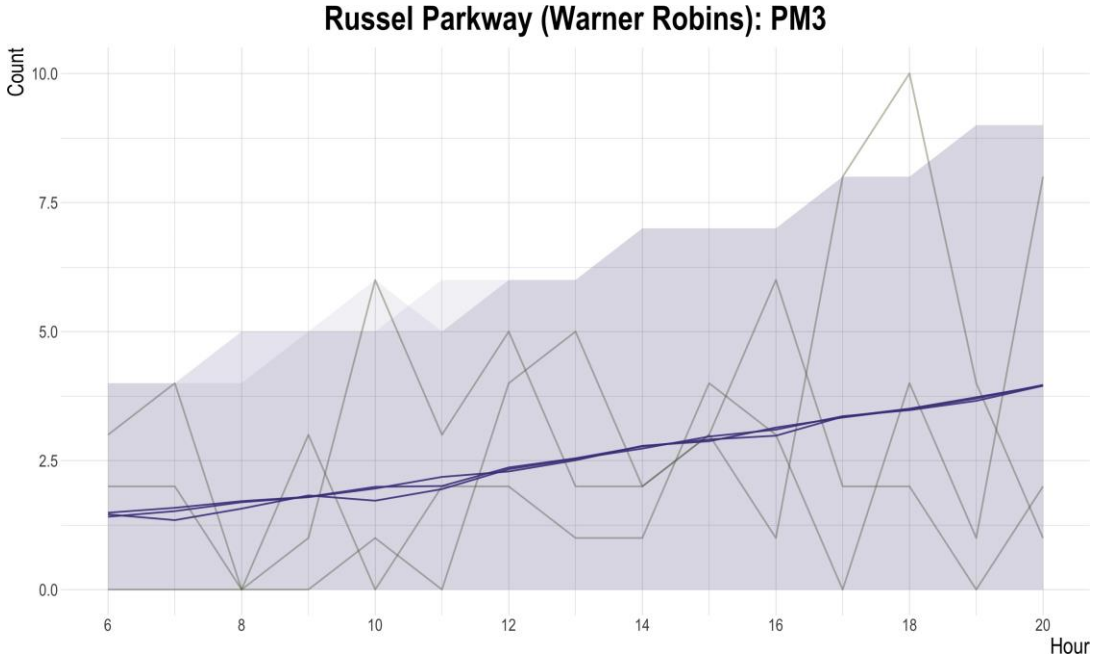


Figure 60: Graph. PM3 specification applied to make predictions at the intersection of Russell Parkway and Davis Drive in Warner Robins

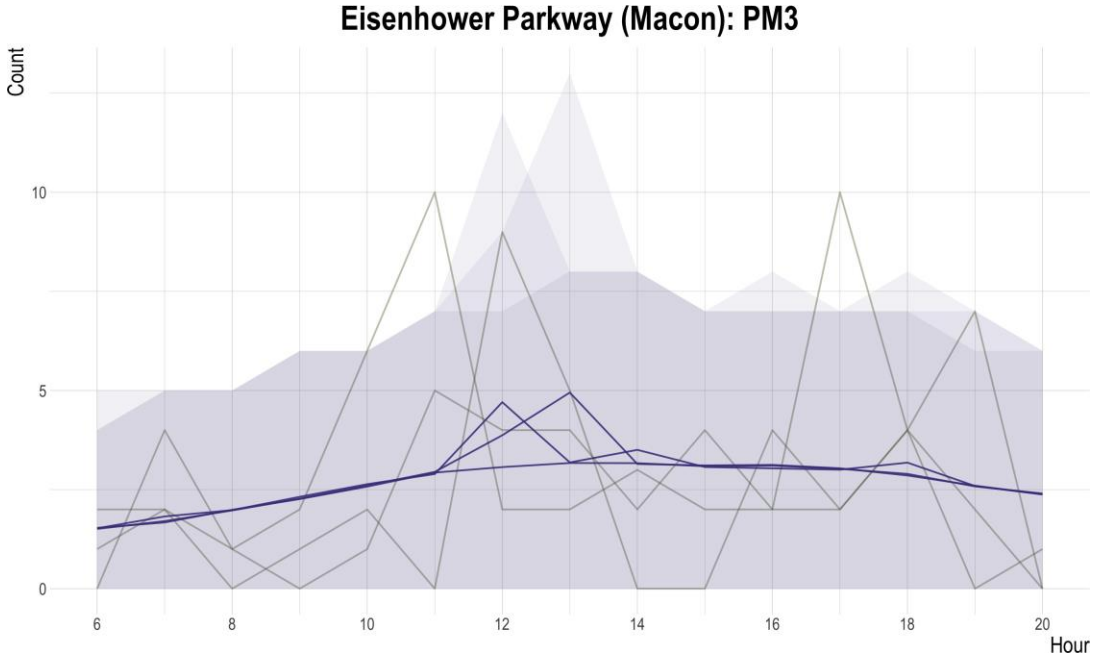


Figure 61: Graph. PM3 specification applied to make predictions at the intersection of Eisenhower Parkway and C Street in Macon

Bicycle Volume Estimation Results

Base Specification Predictions and Predictive Performance

In this section, this work applied the same set of specifications as applied to the bike data, starting with the base specification: $\lambda\ \alpha+\beta_p P*L+f(H)_L$, where λ is the predicted arrival rate based on the Strava proxy data (B1 metric), β_p, which is the proxy (i.e., Strava) multiplier that is allowed to vary by location L, and where $f(H)_L$ is a spline variable that varies by hour and location, regressed on the observed data. Results from the base specification's predictions are shown in Figure through Figure . First, Figure shows predictions based on one of the locations where 90 days of data were collected. Like the predictions from the 90-day observation locations in the pedestrian models, the predictions at this location appear to both follow the hourly average and daily trend, and the 90% credibility interval's boundaries appear plausible. The same is true for Figure , which is another location where 90 days of data were available.

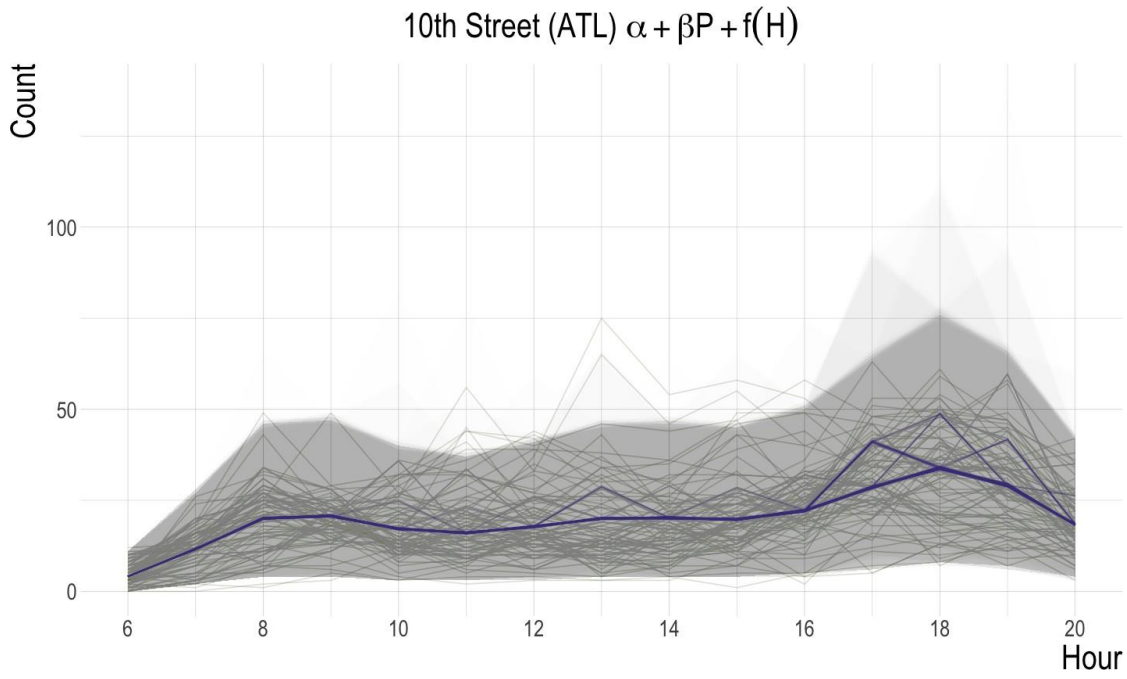


Figure 62: Graph. 10th Street in Fulton County predicted (purple) versus observed (light green) bicycle volumes using the base model specification with a 90% credibility region (translucent purple)

Figure through Figure show predictions at some locations were only three days of data were collected, and one day of that data was held back for testing. Several things were noteworthy from these predictions. First, the predictions still appear to align with the hourly mean despite the limited data. Likely this is a result of the hourly spline variable (which allows for different slopes at each hour) and the presence of longer observation locations also being present in the model.

Second, the uncertainty boundaries are much larger, which is reasonable considering how little data (both in terms of observations and of Strava reported rides) were used to make these predictions. The boundaries of the uncertainty area are quite high in some instances. For example, Figure shows that in the afternoon, the uncertainty interval ranges from 30 bicyclists per hour to less than 5 bicyclists per hour, while the observations only range from 15 to 20 bicyclists per hour. Greater uncertainty in the afternoon seemed consistent across these locations; it was apparent in Figure (Krog Street) between 2:00PM and 5:00PM, and in one of the predictions in Figure (17th Street) at 3:00PM. Likely, this result is because biking activity appears to be much more variable

during afternoon hours at most locations in this study and nearly all of the locations in the model. Third and finally, the credibility interval at locations where three days of data were collected in general were much larger (relative to the observations) in the bike volume predictions than the pedestrian volume predictions; the predictions of bike volumes are more uncertain than the prediction of pedestrian volumes with the same specification. This could be due to fewer locations and fewer overall observations in the models, more variable trends of bicycling activity generally, and less meaningful proxy data in the bike models than the pedestrian models (or a combination of all three).

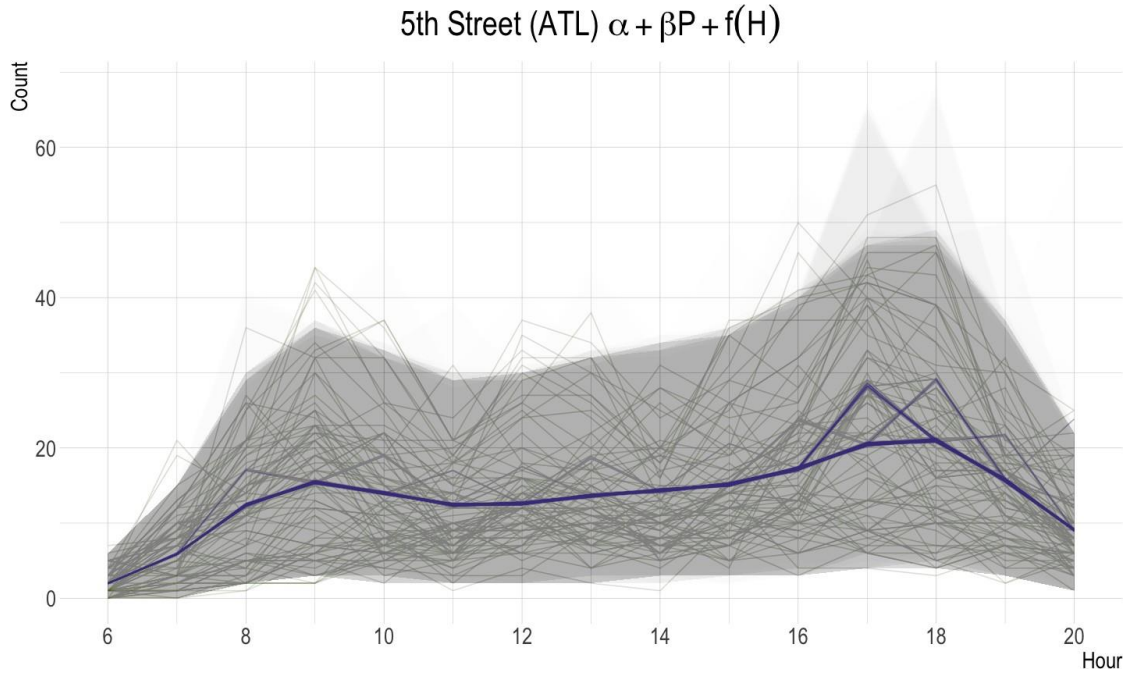


Figure 63: Graph. 5th Street in Fulton County predicted (purple) versus observed (light green) bicycle volumes using the base model specification with a 90% credibility region (translucent purple)

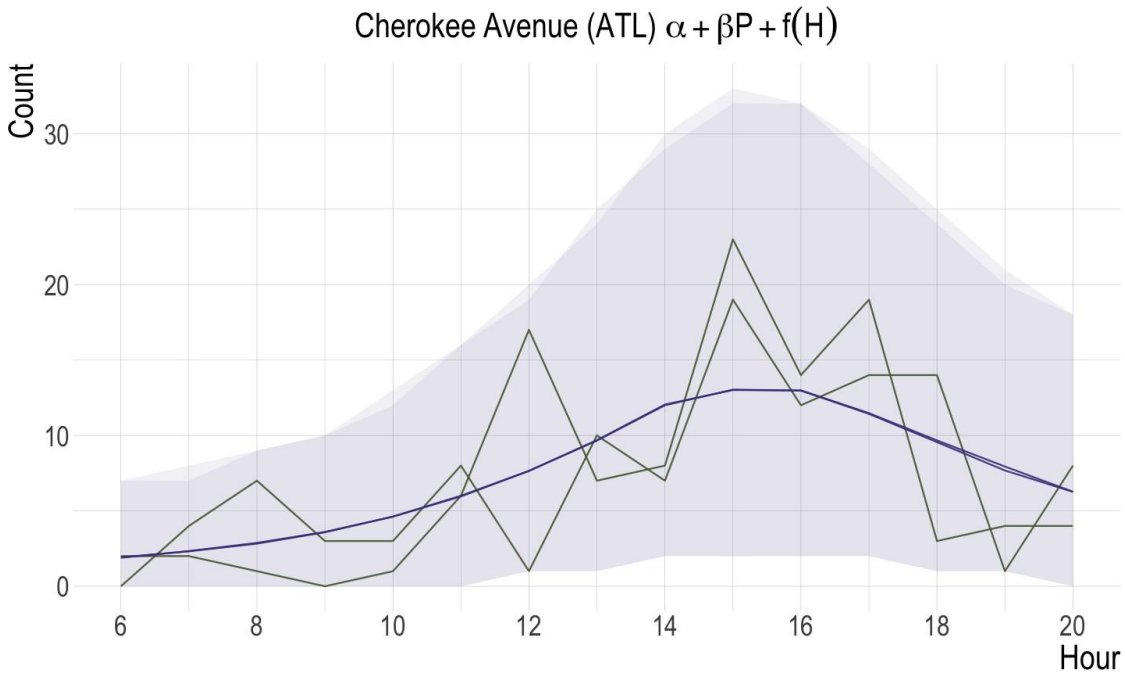


Figure 64: Graph. Cherokee Avenue in Fulton County predicted (purple) versus observed (light green) bicycle volumes using the base model specification with a 90% credibility region (translucent purple)

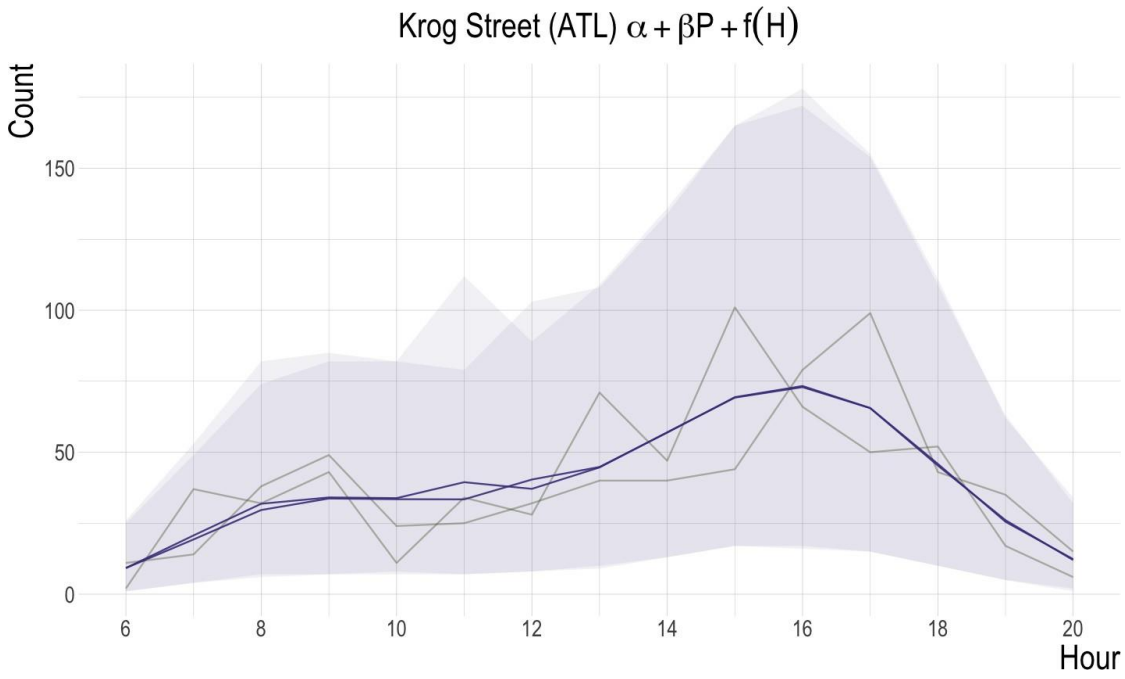


Figure 65: Graph. Krog Street in Fulton County predicted (purple) versus observed (light green) bicycle volumes using the base model specification with a 90% credibility region (translucent purple)

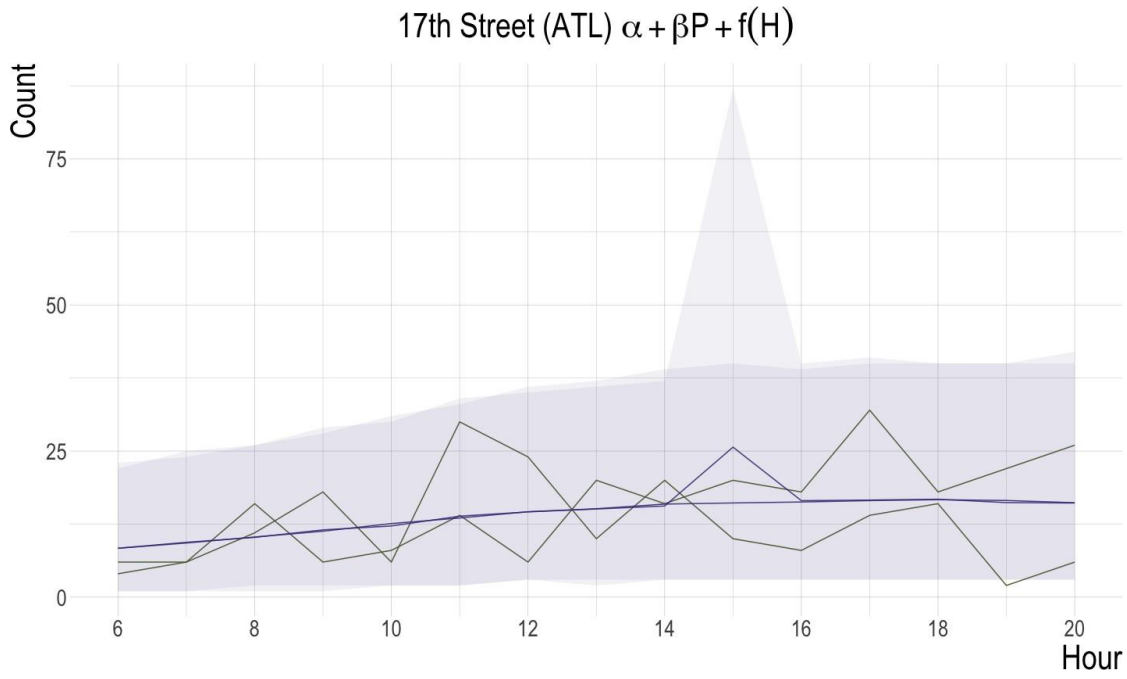


Figure 66: Graph. 17th Street in Fulton County predicted (purple) versus observed (light green) bicycle volumes using the base model specification with a 90% credibility region (translucent purple)

Next, the following figures show how the base specification model's predictions compare to the observed data. Figure through Figure show the mean hourly predicted volumes (in dashed lines) compared to the mean hourly observed volumes that were included in the model at high- and low-volume locations, respectively. The high-volume location predictions matched the in-sample trends well, and the low-volume locations predictions seemed plausible but not very specific. out-of-sample, however, the predictions were not as good in the low-volume locations when compared to the testing data (Figure). While the general trends of the predictions seem plausible at the high-volume locations (Figure), the low volume locations did not seem to be capturing the same trends.

The model's prediction deviance from the observed data by location for in sample predictions and out-of-sample predictions were illustrated in Figure . the in-sample deviance values were shown as smaller, translucent circles, and the out-of-sample deviance values were larger, smaller circles. In general, the bicycle data predictions had lower overall deviance than the pedestrian

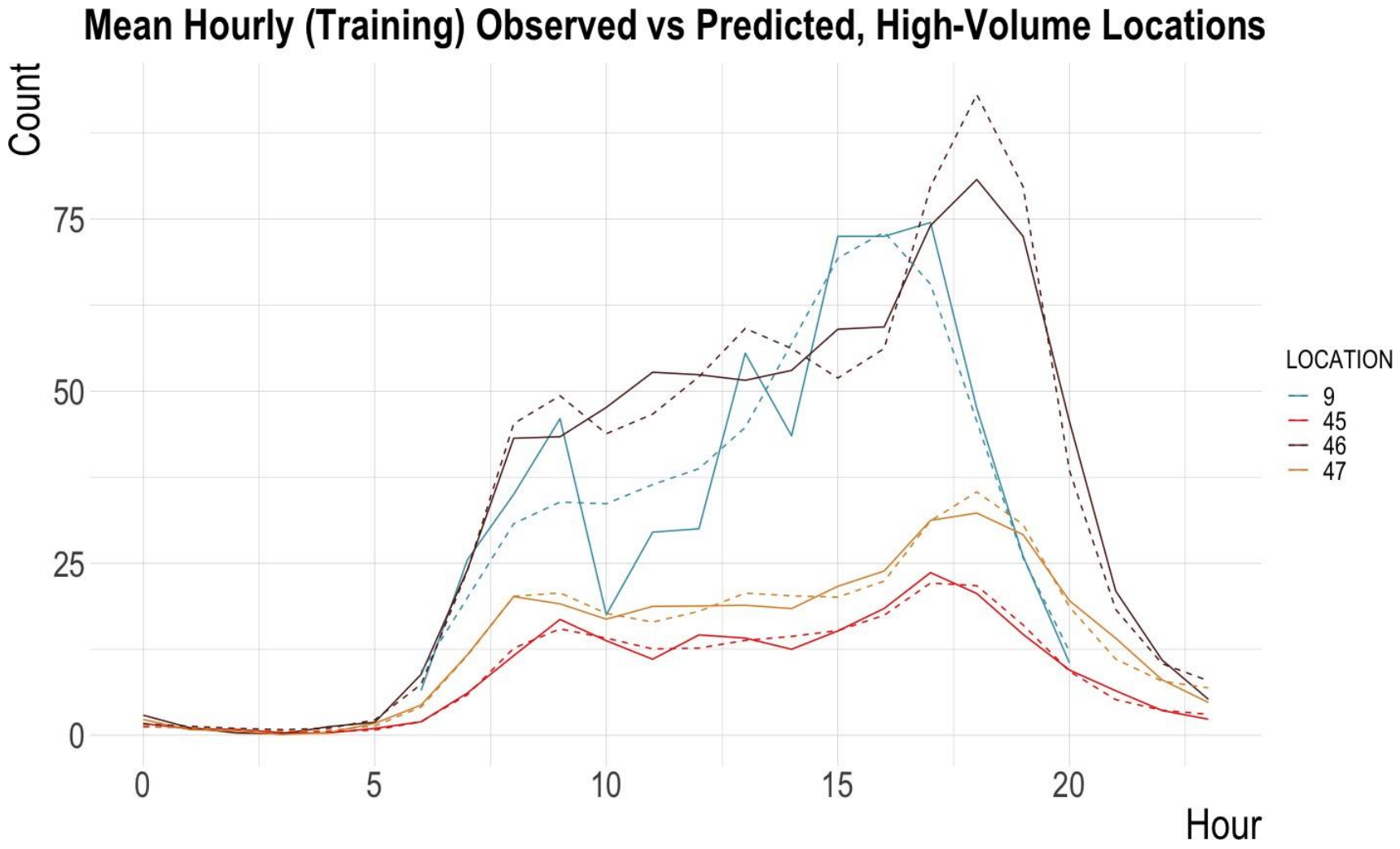


Figure 67: Graph. Average hourly predicted bike volumes (dashed line) and observed (solid line) trends at high-volume locations, using training data only

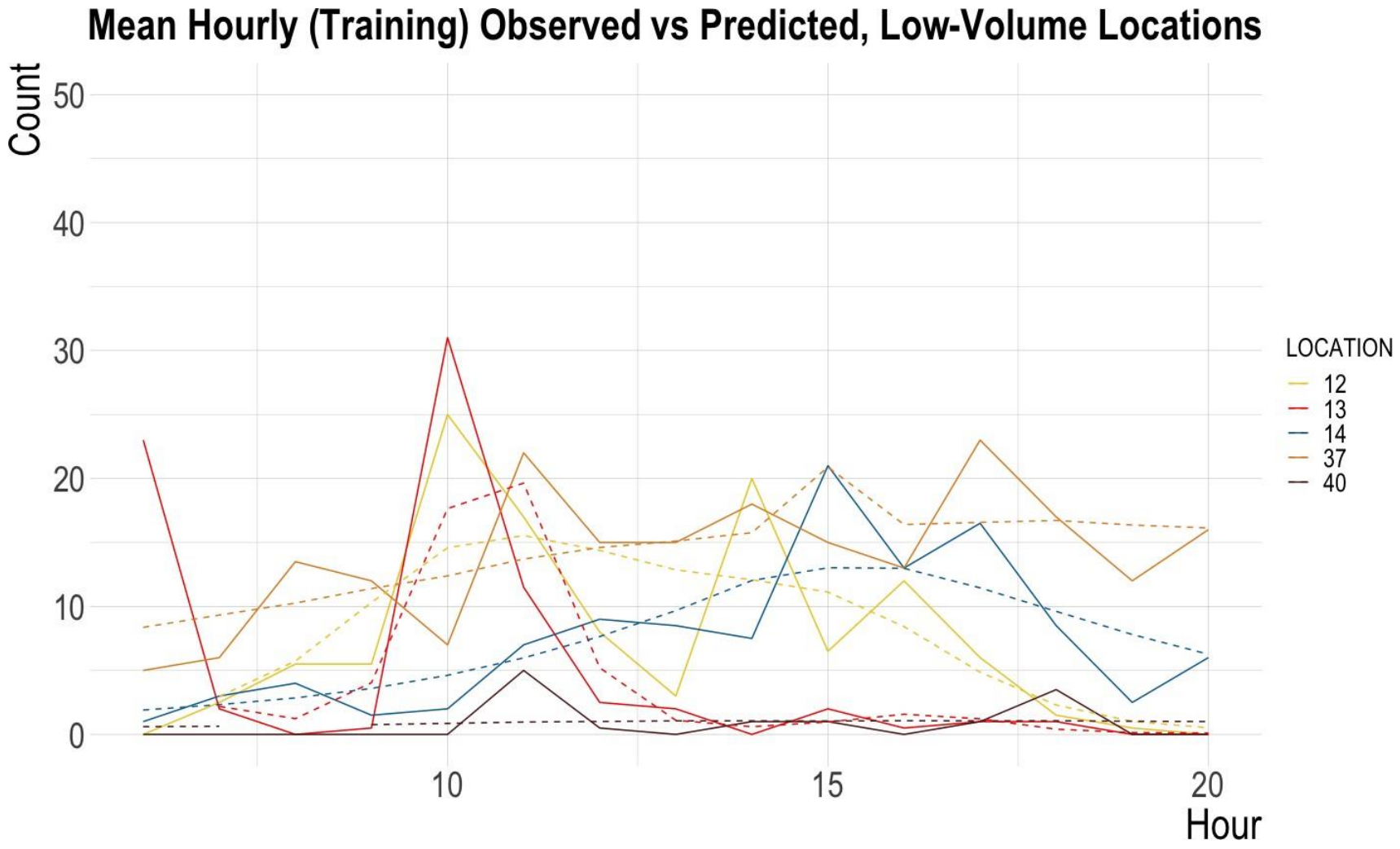


Figure 68: Graph. Average hourly predicted bike volumes (dashed line) and observed (solid line) trends at low-volume locations, using training data only

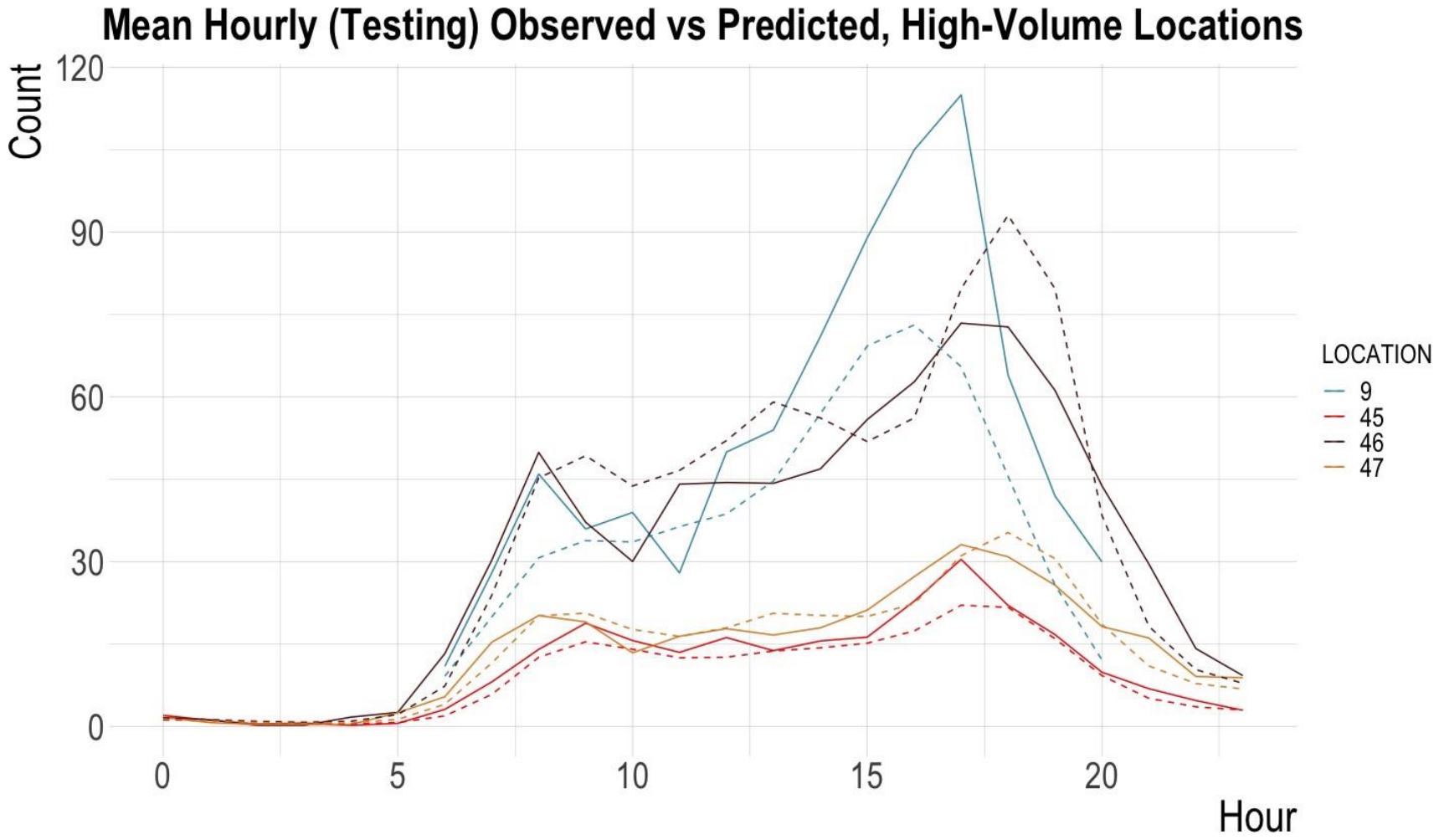


Figure 69: Graph. Average hourly predicted bike volumes (dashed line) and observed (solid line) trends in the testing data at high-volume locations

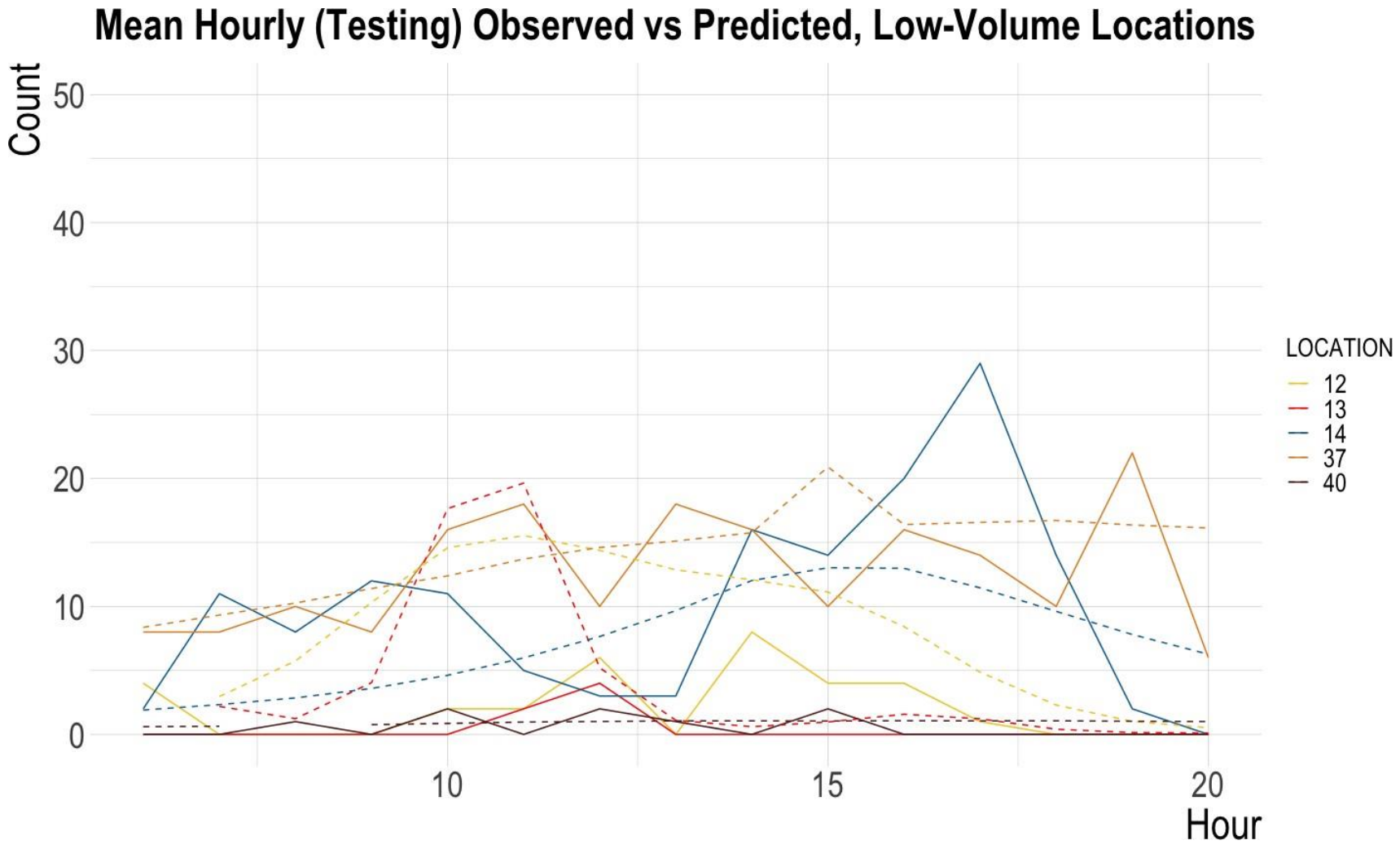


Figure 70: Graph. Average hourly predicted bike volumes (dashed line) and observed (solid line) trends in the testing data at high-volume locations

data deviance (see Figure) when comparing the same specification applied to each dataset. This was somewhat surprising, as the proxy bike data (Strava Metro data) is poorer quality than the proxy pedestrian data (ATSPM actuations). The same trend in the bike model's deviance was also present in the pedestrian model's deviance: in general, in-sample deviance and out-of-sample deviance values are quite close, meaning that the model makes nearly as good (or sometimes better) predictions on the out-of-sample data as the in-sample data. The difference in divergence from observed values versus predicted values appears to vary more by location (and presumably the unique characteristics of that location) than by in-sample versus out-of-sample predictions. This was further supported by the finding that locations where the most data were available (45-47) were not locations with the lowest deviance values; the in-sample and out-of-sample predictions were quite close at these locations, which makes sense, but locations 13 (Dunwoody Road) and 40 (East Point Street) have the lowest overall deviance values. These were low-volume locations as well, which means that this simple specification had relatively good predictive accuracy at places where there are relatively low-volumes of bicycle traffic and minimal observations.

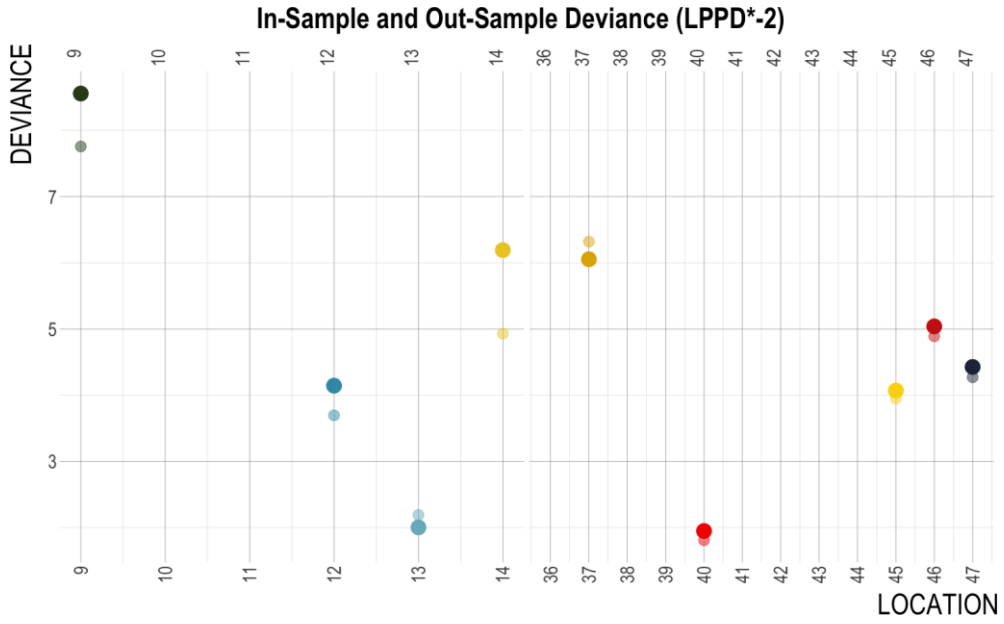


Figure 71: Graph. Base specification in-sample deviance (smaller, translucent circles) and out-of-sample deviance (larger, solid circles)

In addition, the RMSE and MAE of the model applied to the bicycle data was summarized in

Table and (partially) plotted in Figure . The absolute error varied notably between locations, with the largest at location 13 and the smallest at location 14. Location 13's deviance was one of the lower deviance values in the sample, but it had by far the worst measure of absolute error in the sample. Location 12 and 40 are similar. At each of the locations with high values of RMSE and MAE, there were very few observed bicyclists *and* very few self-reported rides in Strava, so it is not surprising that these locations would have poor performance. It was also interesting, however, that some of the locations have quite small error, including some locations where only three days of data were collected. Notably, the error-related results give clearer context to the deviance results shown in the Figure : deviance appears to be more of a measure of how variable the data were compared to the predictions, where as the error shows how close the predictions are relative to the observed data. Both were important to understand how the model performed.

Table 17: RMSE and MAE for bicycle base specification model

Location ID	**RMSE**	**MAE**
9	14.2	11.4
12	657.7	126.3
13	26,390,055.8	4,818,146.8
14	4.4	3.4
37	7.2	5.8
40	790.7	145.6
45	6.5	4.3
46	20.9	12.7
47	8.1	5.3

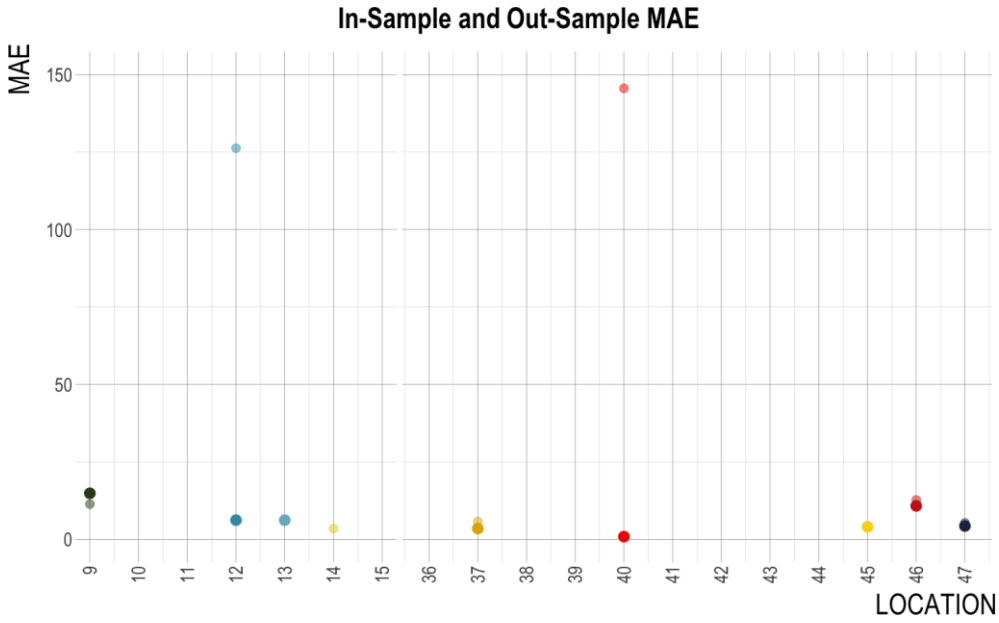


Figure 72: Graph. Base specification in-sample MAE (smaller, translucent circles) and out-of-sample MAE (larger, solid circles), where locations 12, 13, and 40 are not shown due to large values

Additional Specifications and Predictive Performance

Like the iterations shown in Table , additional model specifications were tested on the bike data to attempt to improve predictions. The predictions that were tested are shown in Table . The same iteration of specifications were tested: adding a categorical variable for high-volume versus low-volume locations, adding a cluster assignment variable as determined from Chapter 3 (which suggested 8 clusters), and hyperparameterizations of the Strava data multiplier and the intercept. Table shows WAIC and LPPD leave-one-out cross validation results.

Table 18: Model specifications applied to the bicycle data

Model Name	Specification	Description
Base Specification M0	$\lambda = \alpha + \beta\ P * L + f(H)$	Only includes variable for proxy plus intercept and temporal correlation control
Model 1 BM1	$\lambda = \alpha + \beta\ P * L + \beta\ V + f(H)$	Adds a categorical variable for high vs. low-volume locations
Model 2 BM2	$\lambda = \alpha + \beta\ P * L + \beta\ C + f(H)$	Adds a categorical variable for cluster assignment
Model 2 BM3	$\lambda = \alpha + \beta_{P\,hyper} P * L + f_i(H)$	Adds hyperparameter to the proxy variable
Model 3 BM4	$\lambda = \alpha_{hyper} + \beta_{P\,hyper} P * L + f_i(H)$	Adds hyperparameters for the proxy variable and the intercept

Table 19: LPPD and WAIC results from each model specification applied to the bicycle data

Model	LOOCV-LPPD	LPPD Standard Error	LPPD Eff. # Parameters	WAIC	WAIC Standard Error	WAIC Eff. # Parameters
M0	-20834.8	121.4	96.1	41663.9	243.1	93.2
BM1	-20835.7	121.4	96.8	41665.2	243.0	93.6
BM2	-20837.1	121.7	98.4	41665.1	243.2	93.9
BM3	-20676.5	123.4	103.5	41343.6	246.9	98.8
BM4	-20677.1	123.5	104.1	41347.7	247.3	100.9

There were very little differences between the first three model specification's predictive power. BM1 and BM2, which added a categorical variable for relative volume level and a categorical variable for cluster assignment, respectively, had slightly more negative LOOCV-LPPD and almost exactly the same WAIC – both of which indicated poorer or no better predictions than the base specification M0. There were some improvements, however, when hyperparameters were introduced. BM3 had a slightly better performance in both LOOCV-LPPD and WAIC than BM4, which indicates that the additional variation allowed by the hyperprior added to the intercept's standard deviation did not benefit predictions much in general, and WAIC specifically indicates that the addition did not compensate the cost of additional model complexity. The best performing specification's location-specific deviance is shown in Figure . MAE and RMSE are reported in Table and (partially) plotted in **??**.

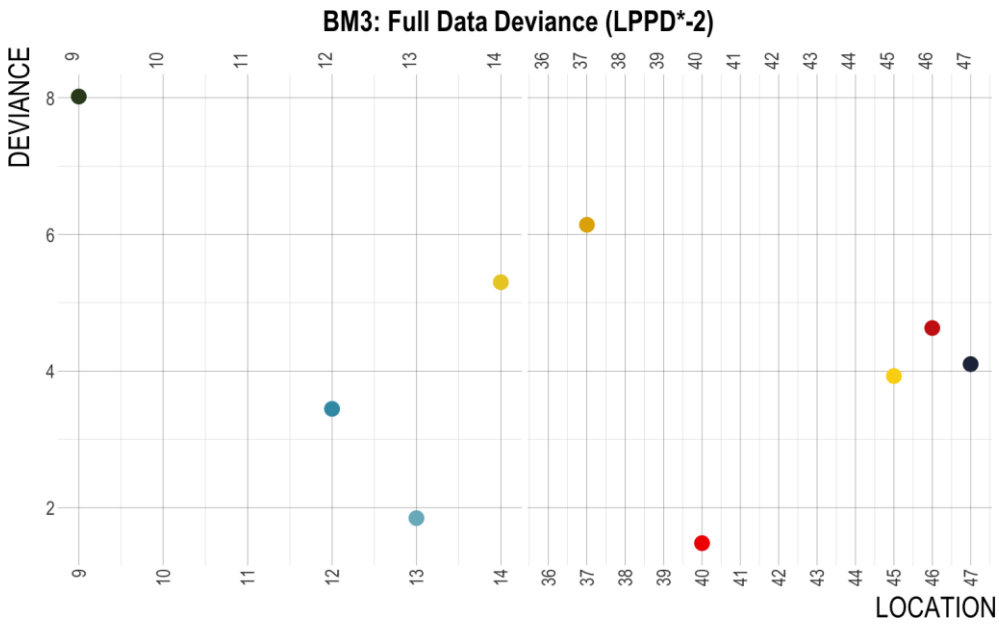


Figure 73: Graph. BM3 full model (all data) deviance by observation location

Table 20: Best performing bike specification RMSE and MAE by location

Location	RMSE	MAE
9	15.0	12.1
12	2.20E+62	3.28E+61
13	4.27E+58	6.37E+57
14	4.78	3.85
37	6.30	4.92
40	1.11E+90	1.65E+89
45	6.44	4.25
46	19.47	12.04
47	7.71	5.07

Like the pedestrian data models, Figure shows that overall deviance is reduced but that the relative deviance value by location does not change. Likely this is a result of the specific benefits that came from pooling across locations; sharing information among all observations in the dataset

improved prediction generally, but it cannot improve prediction where there is not enough information to reduce the uncertainty in the model. Also, Table and **??** show that a duality of results: for locations whose predictions had low error in the base specification, predictive error decreased slightly or remained similar, but for locations where predictions were poor, error increased perhaps exponentially. Locations 12, 13, and 40 had high error in the base specification, but those values skyrocketed in this prediction. These conflicting issues point towards perhaps using either specification, knowing that BM3 is a much better fit for the data generally, but that both specifications perform poorly and should not be used in locations where Strava data *and* observed data are limited.

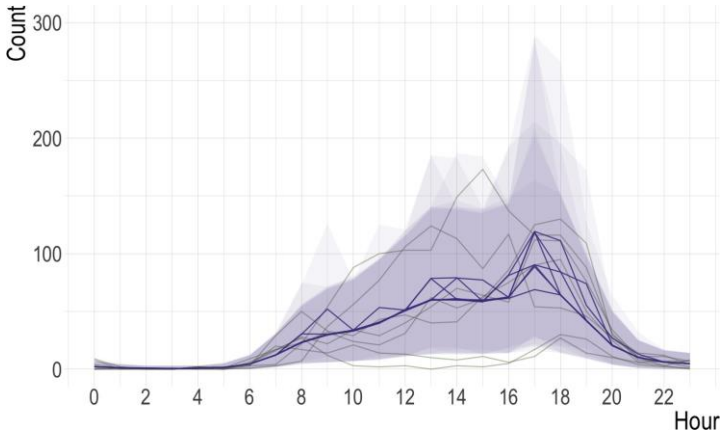
Observation Length and Predictive Performance

Given the relatively few count sites used to build these models, the few Strava observations in the short-term locations, and the resulting wide credibility intervals and poor predictions at some locations detailed in the previous section, this work also explored an additional question related to bicycle volume predictions: how much observation is needed to reduce the uncertainty in predictions? It is possible that observing bicycle activity at locations for longer than three days could improve the predictions and shrink the uncertainty and error in the predictions. Clearly the re- sults above show that predictions with 90 days of observation were generally better than those made with only three days of observation, but how many days between 90 days and 3 days would yield the best predictions with the least amount of observation? Given the available data, this was only possible for three locations (45-47), which are the locations where 90 days of data were available. Creating these predictions involved artificially (and iteratively) reducing the number of observations to 7 days of observation, 14 days of observation, and 30 days of observation, and then observing the how the uncertainty region changed with each prediction. The best performing model, BM3, was used to make these predictions. A bootstrapping approach was employed in this analysis, where every combination of week/two weeks/month in the dataset were modeled, and the error metrics reported here (RMSE and MAE) are the average of each models' statistics.

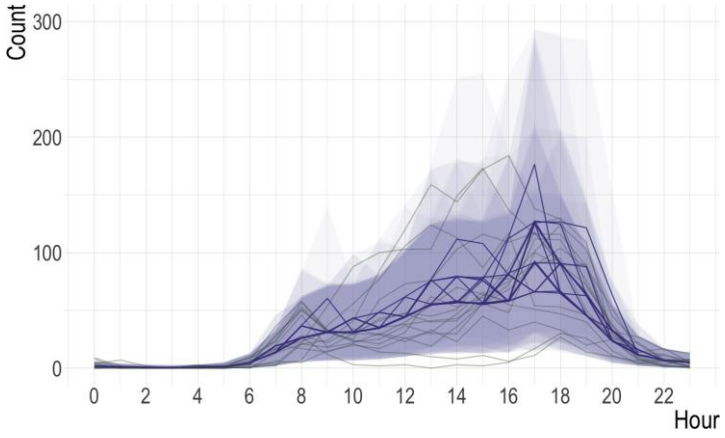
Additionally, the model also includes one location with longer observation periods; for one location, the total number of days in the model were not reduced from 90 days, which was possible because Bayesian models can have observations of different lengths within the model [60]. The rationale for doing this was based on the traditional methods of non-motorized counting [24, 25, 27, 28]: typically, non-motorized counting in practice includes continuous counters, which collect hourly data continuously. These are used to create temporal and seasonal adjustment factors to be applied to short-term count locations that have similar profiles of non-motorized activity. While this process cannot be directly recreated here due to differences in modeling approaches (linearly derived adjustment factors versus Bayesian models with pooled learning across locations) and data limitations (only 90 days of data are available), it is possible to create predictions based on an analogous approach: to create these different predictions, one location (45) contained 90 days of data, while the other two locations (46 and 47) were iteratively reduced to 7 days of data, 14 days of data, and 30 days of data. The models' predictions can still benefit from longer-term observations, similar to long-term counters' adjustment factors, through pooled learning across locations. An example of one of the bootstrapped predictions for each observation length is shown in Figure (10th Street and Monroe Drive) and Figure (10th Street and Myrtle). Each predictions' MAE by location was also calculated and shown in Figure .

Comparing 7 days of counts to 14 days of counts at the Monroe Drive location (Figure), more of the many plotted credibility intervals shrinks noticeably, but there is no perceivable difference between two weeks and 30 days. It makes sense that the region would shrink towards the mean as more data indicates what is plausible from the prediction values. Likewise, the MAE at this location shrinks by several pedestrians per hour. So, for this location, likely the optimal amount of observation is between 14 and 30 days to reduce the uncertainty of the estimates. Like other predictions in this work, afternoon hour predictions are still more highly variable and more uncertain.

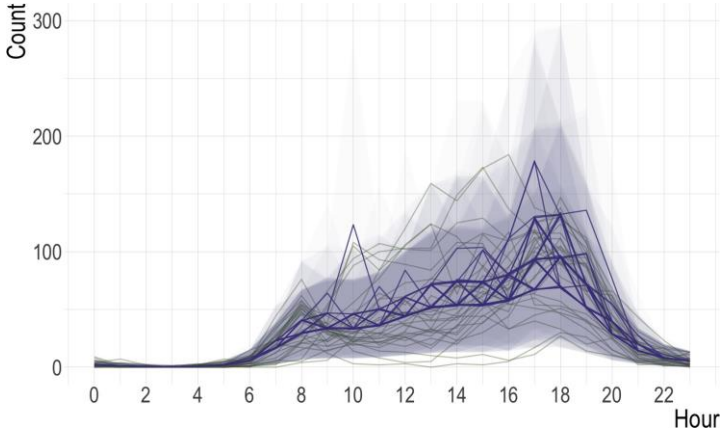
The second short-term location, however, had a different trend. At the Myrtle Street location (Figure), as more observations were added, the uncertainty region became tighter during morning



(a) 7 days

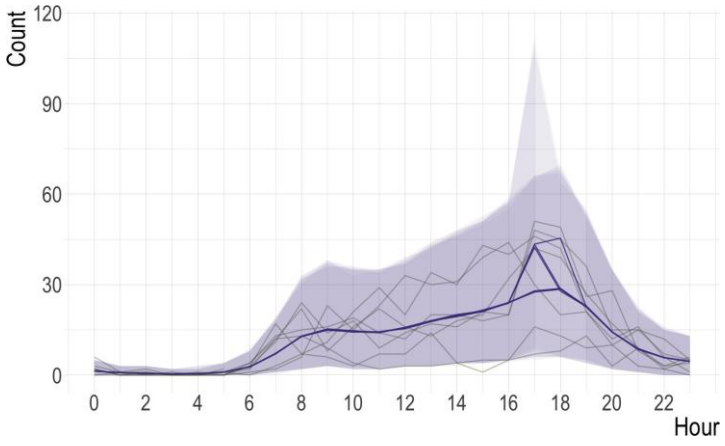


(b) 14 days

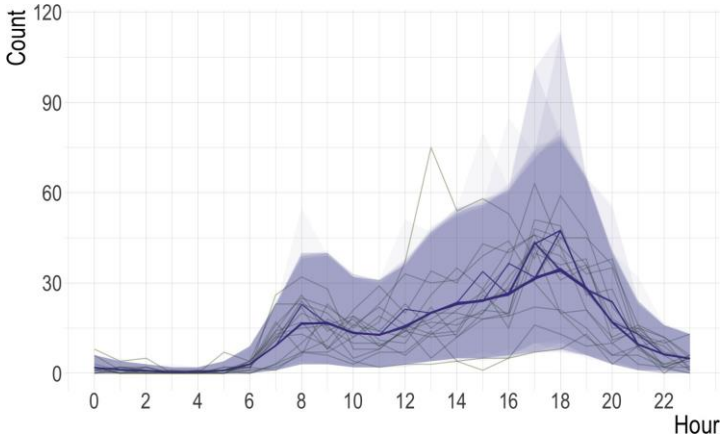


(c) 30 days

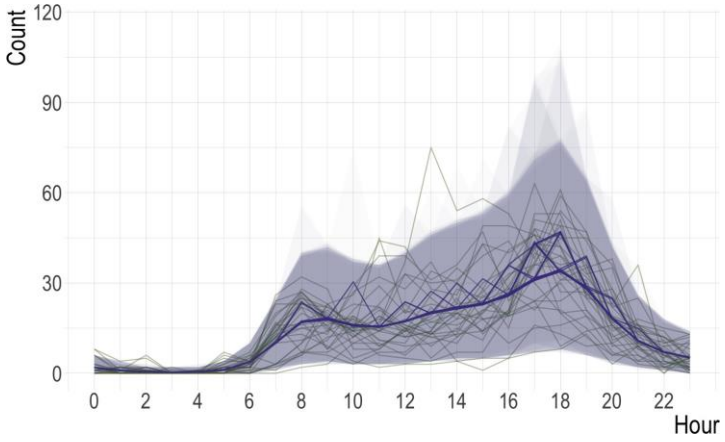
Figure 74: Graphs. Differences in prediction quality when making predictions at 10th Street and Monroe Drive with 7, 14, and 30 days).



(a) 7 days



(b) 14 days



(c) 30 days

Figure 75: Graphs. Differences in prediction quality when making predictions at 10th Street and Myrtle Street with 7, 14, and 30 days).

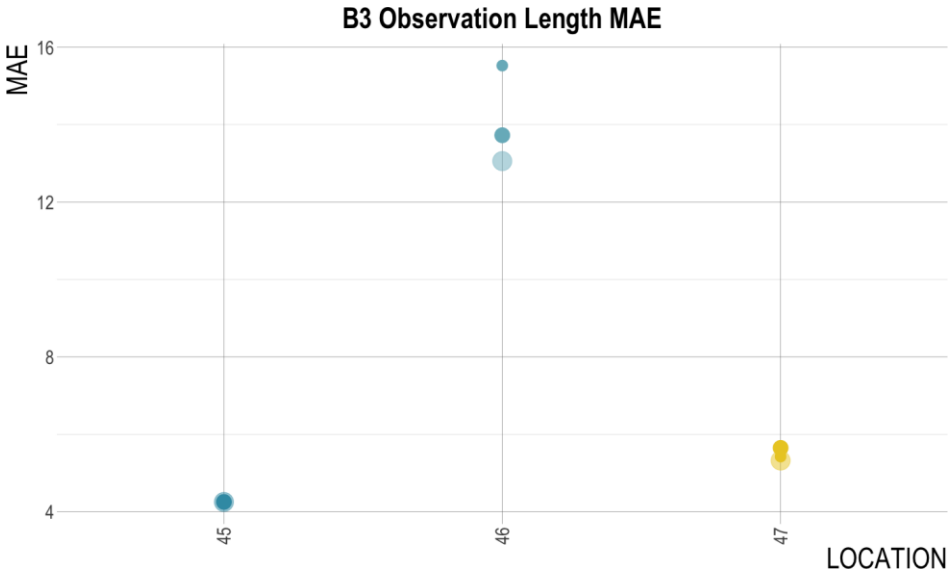


Figure 76: Graph. MAE for different observation lengths for the example predictions shown in previous figures, with 7 days (small circle), 14 days (mid-sized circle), and 30 days (large translucent circle) of observation at short-term count locations.

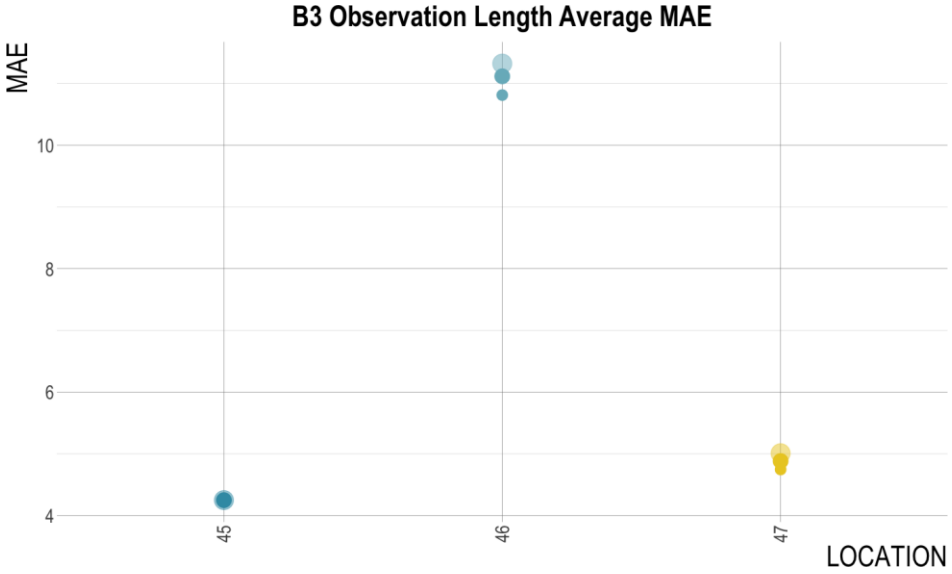


Figure 77: Graph. Differences in full model in sample MAE with 7 days (small circle), 14 days (mid-sized circle), and 30 days (large translucent circle) of observation at short-term count locations.

commute hours but became wider in afternoon hours. This was likely because the additional days of observation varied notably from the first seven days of observation, making the range of possible values actually larger, which in this case appears to be more realistic. There is no notable difference between 15 and 30 days, however, except for a slight decrease in the peaks of some afternoon uncertainty regions. The MAE results are quite consistent for all observation lengths. Collectively, these findings point to only 7 days of observation needed at this location.

The differences between these two locations make sense when considering the observed data: the Monroe Drive location had a wider range of possible values at any hour than the Myrtle Street location. For example, in the 7 day observation graphic in Figure , there were both approximately 150 bicyclists per hour *and* approximately 15 bicyclists per hour *and* 50 bicyclists per hour. This is quite a wide range of possible values at this hour, which is reasonable given the location. This is near the recreational bike trail (Atlanta Beltline) and a large park (Piedmont Park), so it makes sense that there may be many bikes per hour on weekend afternoons and very few during weekday afternoons. It makes sense, then, that more observation is needed to reduce the error in predictions, and that the error in predictions would be larger than a place with a more limited possible range of hourly rates. In contrast, Figure observations range from around 25 to 30 bicyclists per hour at 2:00PM during the 7 days of observation. As more days are added to the observations in the observations begin to "fill in" towards a mean.

Given this information, the findings in Figure and Figure are reasonable. Figure shows that the average MAE shrinks more notably as observations are added

DISCUSSION OF ESTIMATION RESULTS

This Chapter used proxies of biking and walking activity to estimate hourly volumes of biking and walking (separately) at locations around Georgia. This work showed that the two datasets used here, ATSPM actuations and self-reported Strava rides, are generally useful sources of information for understanding biking and walking, despite their incomplete representation of the population and of activity. This finding aligned with existing literature in some ways and expanded the literature

in others. First, ATSPM data has been shown to create good estimates of AADP in some contexts [38, 33, 34, 35, 36, 37]. Similarly, Strava data has been shown to create meaningful estimates of AADB when conflated with many observations [31, 42, 43]. The findings in this work support the literature: these sources of proxy data – ATSPM and Strava – contain meaningful information for predicting non-motorized volumes, but this work also showed that these data can predict better than annualized values in some cases. In cases where there were enough observations, this work showed that it was possible to create hourly estimates of activity at the intersection (pedestrian) or segment (bike) levels.

This finding may be important for quantifying risk in future work. Bicycling and walking rates are highly variable such that an annualized average value does not have much physical meaning for the number of cyclists or pedestrians in a location at a specific time. And because travel volumes are confounded with crash rates, it has previously been difficult or impossible to control for cycling volumes with a physically meaningful metric. While annualized values allow for comparison of volumes across different locations, annualized values limit decision-making around risk that varies by time of day at different locations. If a goal of a transportation agency is to prioritize safety (or other) infrastructure interventions by those that reduce risk or those that serve the most people, using annualized values cannot correctly inform that decision-making process. This work demonstrated that it is possible to quantify a higher resolution with easily-accessible data, so future work should leverage this finding to better inform safety analyses. For example, the posterior predictions of activity could be used a prior distribution of a crash prediction model at the intersections and segments. Also, the predictions could be used to create hourly risk profiles at different locations that control for the number of people using the system.

This work also revealed several important findings related to modeling non-motorized count volumes. First, this is the first work that attempted to use a Bayesian model to predict non-motorized volumes (to the author's knowledge). Hopefully this work can introduce Bayesian statistics, which are well suited for the challenge of quantifying non-motorized volumes, to the research and practice areas of exposure and risk calculation in the non-motorized transportation

field. There are many more opportunities to expand this modeling methodology in practice. For example, what would it look like to introduce these kinds of models to an existing non-motorized count program? How do they fit into the existing regime of permanent counters and then direct placement for short-term counters? Future work should address these questions.

Second, this work explored what kinds of model specifications can best improve prediction of Bayesian count prediction models applied to non-motorized volumes. The best performing specification for both the bicycle and pedestrian models were those that introduced hyperpriors to the proxy multipliers, which follows the finding of Nelson et al.: that predictions of bike volumes can leverage Strava data most effectively when the relationship between Strava volumes and the observed data are allowed to vary by location [31]. This work expands that finding further: it is also true for pedestrian models that allowing the relationship between ATSPM actuations and observed values to be different at each location improves predictions. All models tested in this work interacted the proxy variable with the location variable, and the best specifications leveraged multilevel effects across all locations to make better predictions overall. So the use of multilevel models and/or interaction terms with location appeared to be a universal recommendation across modes when using proxy data.

It is worth discussing the trade-off between predictive power and model complexity. As noted in the preceding paragraph, the best performing specification for both sets of predictions included multilevel effects applied to the proxies' priors' standard deviation. It out-performed the base-specification in both the LOOCV-LPPD and WAIC metrics, which is notable as WAIC disadvantages models as they add more parameters. For example, the pedestrian models had 412 parameters in the base specification and 413 parameters in the best performing specification, and the number of effective parameters increased from 96 in the base specification to 104 in the multilevel specification, pointing to the fact that this specification could more effectively use the parameters in the model despite only adding one additional parameter. However, it is important to consider the practical use of the model. The additional model complexity made only subtle decreases in prediction error in both the bicycle and the pedestrian models. So while the multilevel models performed

better in terms of fit and information criteria, likely either specification – the base specification or the hyperparameterized specification – would be suitable for use. Depending on the use case, it may be preferable to have a simpler model, in which case it would be suitable to use the base specification.

Third, these predictions made use of a spline, or smoothing term, to capture the non-linear, time dependent relationship among observed data and the proxy variable. The spline is applied to the hourly categorical variable, *H*, in all models. When time is included in a variable for modeling arrival rates in transportation, more often it is added to a generalized additive model as a categorical term on its own, or perhaps it is used as an interaction term. To the author's knowledge, the use of a spline term in estimating non-motorized travel is a unique contribution to the literature and to the research area of traffic volume predictions.

Fourth, this work used a clustering categorical variable that was derived from time series clustering of the proxy data. The rationale behind this approach is based on how non-motorized count programs create factor groups based on temporal trends, which are then used to create estimates of annualized non-motorized travel volumes. This work intended to use factor-group-like clusters as a categorical variable in the prediction models. While this variable did create modest improvements in predictive power in both the bicycle and pedestrian models, the greatest improvements in predictive power came from multilevel models, not cluster assignment. While this may have been expected in the bicycle prediction models since the clusterings were not as high quality gen- erally as the pedestrian models), it was surprising that it was not a very meaningful addition to the pedestrian prediction models. It is possible that this is due to using proxy data to form the clusters instead of observed data; perhaps clustering large volumes of observed data (that were not available for this work) would produce different results. It could also be possible, however, that there is simply too much variation, among both the actual non-motorized activity and the relationship of proxy data to that location, for the cluster assignment to improve predictions; there may be differences in the trends of activity in different places that can separated via time series clusterings, but those differences may not be meaningful enough to make hourly predictions much better than they

would be without the clustering variable. Instead, the hyperparameter pools information across all locations and learns the standard deviation from all of the data in the model, thereby generating more information that a single category could. Regardless of the reason, perhaps this work indicated that factor groups may not be the best way to better understand non-motorized travel under a Bayesian regime, especially when multilevel models with interaction effects can be used, as this work demonstrated, to allow all locations to be modeled separately with pooled information across all locations.

The previously listed findings and contributions focused on the modeling approach generally, but it is also interesting to investigate the results of the biking and walking volumes separately, starting with the pedestrian models' results. When reviewing which of the locations had the highest (i.e., worst) deviance values and lowest (i.e., best) deviance values, it seems that the locations with the poorest predictive performance were those that had the greatest range in possible values within the temporal distribution. Predictive performance in this case was based on a (log) average of the probability density of each observation for each sample that makes up the posterior predictive distribution. Given this measure of predictive performance, the finding that larger ranges of values each day made sense; when there is a greater range of possible arrivals at a location in a given hour, there will be greater uncertainty at any hour that the prediction will be true given the data observed. Supporting this theory, predictions with the lowest deviance values were those with the lower maximum values and smaller range of possible observed values – even though the trend in the observed data is not clear, and the differences among the observed values and the predictions were apparent (see Figure and Figure). RMSE and MAE results pointed to similar conclusions; outside of location 10, there were quite similar trends among deviance, RMSE, and MAE (see Figure and Figure). Perhaps this points to the need to consistently compare error metrics, deviance values, predictions, *and* uncertainty intervals; one of these metrics alone should not be used to assess overall model performance. The results with the lower deviance values more meaningfully shrank to the mean and followed the observed trend in the data, but by these established measures, they are more difficult to predict.

The last section of the bicycle models' results also contributes a finding that is unique in the literature: when researchers or practitioners decide how much data needs to be collected at short-term bicycle count locations, the findings from this work indicated that this depends on the expected variation at each location. **??** shows that, at Location 46, MAE from the predicted value decreased as more data were collected. This made sense for that location because there were a greater variety of possible arrival rates per hour at this location. For Location 47, however, prediction errors maintained relatively stable, which also is a reasonable finding: as more data were collected, the data naturally accumulated towards the mean values of the observed data. The graphics in Figure shows why: as more data were added to the model, more and different trends in activity became apparent at that location. From these findings clear implications emerged: If it is known that there are very consistent trends of activity in one location (i.e., that the hourly variance among the days of the week is small), 7 days of observation of bike activity may be enough, depending on the level of precision needed; the differences in MAE reported here were small, despite the improvement in predictive power as more data was collected. If it is not known that there are stable trends of biking activity, however, two weeks of observation are needed to inform biking estimation models to capture the variety of temporal trends. In both cases, little benefit will come from observing longer at each location (assuming there is a permanent count location present in the data).

Finally, it was also important to cite the limitations of this work. This Chapter's findings were limited by the challenges of using proxy data, especially Strava given its biases and limited availability in less urban locations. The bike prediction models could only use 8 locations despite data being collected at 40 locations, which created more uncertain results and limits what can be gleaned from this experiment of using Bayesian statistics and Strava data. ATSPM data were not subject to the same biases, but they were subject to their own limitations. First, ATSPM data could only be used in places where ATSPM-equipped signals were installed, and not all events in the data were recorded at all locations (i.e., event 45 is not registered on signals with pedestrian recall). This complicates which signals and events can be used for model comparison, and the lack of ATSPM-equipment in some locations does not allow this approach to be used in much of

rural Georgia except for the rural towns. Also, ATSPM signals were more sparse in south Fulton County and areas south of the City of Atlanta, which limited this work's ability to study pedestrian activity in places where there are more people of color. Importantly, ATSPM data is not available at many mid-block crosswalks (unless there are HAWK signals), and these mid-block crossings are highly dangerous for pedestrians in many contexts. So, this approach and data cannot be used in these locations, prompting need for alternative approaches and data to be explored for these high-risk locations. Alternative datasets would be needed to study mid-block crossings, perhaps just in the form of direct counts. Also, ATSPM actuations and calls are only detected when people push the button (at signals without pedestrian recall), so any variable that changes in button-pushing behavior could also influence future applicability of these results. Examples of this could come from both social forces (i.e., people may actuate less in general after the COVID- 19 pandemic, people who choose to live in certain locations are less likely to push the buttons due to expectations about the the infrastructure's service, etc.) or environmental forces (i.e., the intersection's cycle length is long enough that people do not expect the signal to serve their needs in a timely manner, etc.). The extent to which these forces influence how people use push buttons is unclear from existing research, and these factors are not included in any of these models. Future work should investigate these kind of influences, and it should also investigate how the kinds of models introduced by this report might change if applied to passively collected data from sensor-equipped signals that detect pedestrian presence without actuation. Likely a similar specification could be use, but new priors should be explored.

Perhaps the most overarching limitation to note was related to the small number of locations that were usable for developing the bike models. Despite collecting counts at 40 locations, only 8 locations had enough data to be included in the model. Future work should learn from this limitation by ensuring that bike counts that are attempting to use Strava must observe for much longer than 3 days in order to ensure counts from Strava and observations are actually measured.

Additionally, all of the models in this work were optimized for predictive performance and were not designed to fully model the system. They do not relate contextual variables – like the

presence of bicycle infrastructure and sidewalks, intersection density, traffic volumes, or other variables that are known to correlate with biking and walking activity – to rates of biking and walking. It is possible that introducing these variables could improve predictions in places with higher deviance values. But, the goal of these models instead were to (1) to make the most accurate predictions with as few variables as possible; and (2) to relate readily-available data proxies to observed counts at these locations. This was a notable limitation and an area for future work for academic research purposes, specifically: Can Bayesian models formally describe the relationship between contextual variables and walking and biking rates? Should contextual variables be used in a generalized additive modeling way? Or would more sophisticated modeling approaches be needed? The work completed in this chapter should be used a starting point and guide towards this end.

The clear next steps for this work are two-fold: First, the next iterations of this work should further investigate how well these models predict out-of-sample for locations that were not included in the data. This could be approached in two ways. First, the in-sample versus out-of-sample cross validation should be re-imagined. In this report, this cross validation approach was considered temporally, not spatially, which addressed how well these models could predict for the same location across a different day in time. Future work with these models should also address how well these models predict for different locations. Perhaps this could cross-validation could be done for high- versus low-volume locations, where the model is trained on a sample of high-volume locations and then used to predict volumes at out-of-sample, high-volume locations using the trained models' posterior distribution. Findings from this kind of process would have important for both the academic and practical contributions: GDOT (and other transportation agencies) would benefit greatly if posterior predictions from data collected at one location could be used to predict volumes at another location, and if this is possible, this would be an major academic contribution pointing to the usefulness of Bayesian models to predict even volatile and non-stationary time series.

Second, work that builds upon this report should investigate ways to apply these high-resolution predictions to risk metrics for understanding safety. The purpose behind developing the models

in this report was safety, specifically to normalize crash rates by traffic volumes. When this is done in existing literature, approaches to using volumes to normalize crash rates use averaged and/or annualized values of non-motorized traffic, which, as demonstrated in this report, may not have much physical meaning for non-motorized volumes. Since this work has shown that it is possible to make high-resolution estimates in some places, it opens the door for more meaningful measures of risk. For example, in the locations where the pedestrian volume predictions are high- quality, there is an opportunity to study how pedestrian risk change over the hours in the day, as it is known that pedestrian safety in darkness is a problem in the U.S. Alternatively, predictions could be annualized and use to normalize for volumes, comparing risk per person along high-injury networks and outside of those networks – which could shed light on the usefulness (or the potential misdirection) of high-injury networks for directing planning and infrastructure dollars.

CHAPTER 4: RISK FACTORS ASSESSMENT

EXPOSURE AND RISK

The previous sections have identified exposure measured in units of bicyclist and pedestrian traffic volumes as an important but complex element of bike and pedestrian safety: as more people bike and walk for (more) trips, they are exposed to risk of crash. But the literature also clearly shows that as more people in the system bike and walk, the risk per person in the system actually *decreases* [22, 63, 64, 5, 65]. This is called the "safety in numbers" effect. While the exact reason for this effect is not fully understood, it is likely due to two factors, both of which relate to driver expectations [22, 63]. First, when more people bike and walk for trips, a larger share of the driving population may also walk and bike for trips, meaning that they may drive with increased awareness of cyclists' and pedestrians' needs. Second, when more people bike and walk in an area, drivers may come to expect to see cyclists and pedestrians, which primes their expectations and decreases their perception-reaction time during conflicts. Collectively, these two factors result in a decrease in crashes, injuries, and deaths in places where more people bike and walk.

Understanding exposure is not the only consideration for bicycle and pedestrian safety when considering bicycle and pedestrian infrastructure. *Exposure is an important consideration as the presence of more people biking and walking itself creates a safety benefit that may correlate with or even add to the safety benefits created by other safety countermeasures.* Therefore the goal of reducing crashes can be paired with the goal of increasing usage of bicycle and pedestrian facilities.

There are many factors that also create (or mitigate) risk for people biking and walking. This chapter explores these factors through the following approach:

- First, built environment risk factors for bicyclists and pedestrians are summarized, citing sources from academic literature and governmental guidance.

- Then each risk factor's relationship to the Safe Systems Pyramid (SSP) [3] (repeated in Figure) and exposure are synthesized, resulting in a table that prioritizes safety interventions by population-level effectiveness and relationship to exposure. The SSP's novel framework can be summarized in two points: First, the framework states that underlying cause of injury and death during traffic crashes is kinetic energy exceeding the human body's allowable tolerance (which is especially relevant for vulnerable road users like cyclists and pedestrians). Second, safety countermeasures should be prioritized by the effectiveness of those interventions at a population level and by minimizing personal initiative needed for the safety benefit to be realized. The table in this chapter relates each level of the SSP to potential safety countermeasures, risk factors, and exposure.

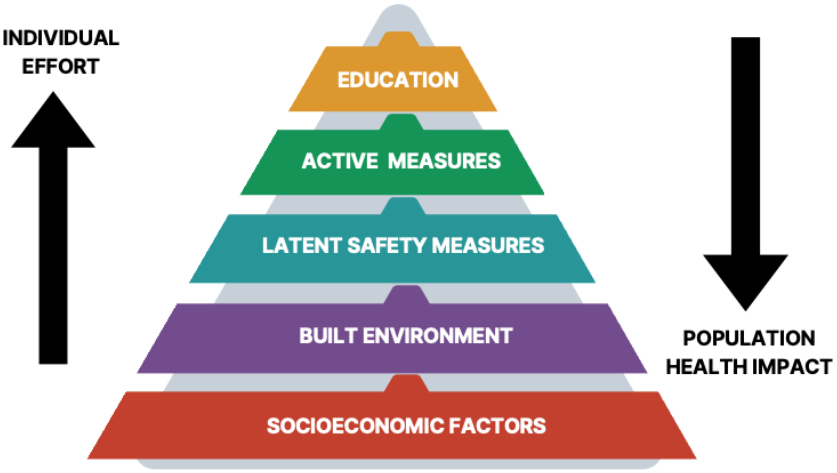


Figure 78: Illustration. Safe Systems Pyramid from Ederer et al. (2023)

Bicycle and Pedestrian Risk Factors

The following summary of risk factors has been developed from a synthesis of government guidance [12, 66], the Safe Systems Pyramid [3], and academic review papers [67, 68, 69, 70].

- *Context:* The type of development context (i.e., urban, rural, and suburban) results in different levels and types of risks for bicyclists and pedestrians [71, 72, 73]. This is because certain

kinds of development have differences in typical roadway/infrastructure design, prevailing speeds, and road-user expectations. For example, urban environments may see more crashes due to higher traffic volumes [71] and more frequent interactions among all roadway users [74]; the presence of retail, parallel parking, pedestrian crossing, conditional yielding, bus stops, and other street elements commonly condensed often require road users to make more complex judgments or estimations about other road users' behavior. Similarly, the same elements may draw the road users' attention away from a more critical risk, creating additional opportunity for a crash to occur [18, 20, 75]. But the severity of crashes (i.e., likelihood of fatality) is often greater in rural areas due to factors like higher prevailing speeds and higher collision speeds given limited driver expectations of bicyclists and pedestrians. Because development context is more of a "catch-all" for more specific risk factors, this review will focus more on the specific risk factors that correlate with development.

- *Speed:* Roadways with high prevailing speeds are dangerous for people biking and walking. Speed represents an important part of the epidemiological cause of injuries and death [76]: people die in roadway crashes when the kinetic energy hitting their body exceeds the allowable human tolerance [3], where kinetic energy is a function of speed and mass. Given this fundamental understanding of speed's impact on risk, it is not surprising that the literature consistently finds that driver speed impacts both crash frequency and severity, and high-speed crashes are always more severe regardless of road users and context – in part due to the increased kinetic energy [3] and in part due to decreasing driver vision cones as speeds increase [77, 78]. While speed limits are often used as a proxy for driving speed [73], it is not always the case that the posted speed limit reflects the driving speed, especially if the roadway design characteristics do not align with the posted speed limit. Therefore, designing roads to encourage appropriate speeds for the land use and mix of users (including bikes and pedestrians) as well as signing the appropriate speed are both critical to a safe system.

- *Lighting and visibility:* In a transportation system with vehicles that are much larger and

obscuring than the people who travel in and around them, an important risk factor is cyclists and pedestrians' ability to be seen by other road users. There are two factors that influence risk within this concept: lighting and visibility. First, bicyclists and pedestrians are at greater risk in low-lighting conditions [16, 65, 79, 68]. Low-lighting at dusk and nighttime when there is no street lighting create risk for pedestrians by obscuring them altogether and/or reducing driver perception-reaction time, as people may not be seen by drivers until they are closer. Also, evening hours may lead to fewer vehicles and thus higher driving speeds [80, 81]. Low-lighting appears to be an important factor in turning-movement-related crashes and failed overtaking crashes. This appears to be of particular concern in rural areas where there is less ambient light and higher travel speeds. Second, visibility hazards due to roadway design also creates risk [82]. Parking close to pedestrian crossings, curves, planting obstructions, and elevation along the roadway can block pedestrians and bikes from view, creating opportunities for collision. These kinds of visibility risks are more common in urban environments.

- *Geometric Roadway Design:* Roadway design elements can create notable risk for bicyclists [83] and pedestrians, including:
  - *Travel lane width, crossing distances, and intersection characteristics:* Wide travel lanes encourage faster driving, contributing to greater risk for bikes and pedestrians. Roadways with wider intersections – due to multiple travel lanes and wider travel lanes – are more dangerous to cross for bicyclists and (especially) pedestrians. For people walking, longer crosswalk distances increase the amount of time that the user must spend in a conflict zone with vehicles. In addition to travel lane width and crossing distances, there are other intersection characteristics that can create poor safety outcomes for bicyclists and pedestrians, including some roundabout designs [84].
  - *Paved shoulders:* The effects of paved shoulders on bike and pedestrian risk are mixed [83, 85]. First, paved shoulders create delineated space for people biking and walking,

especially in urban areas, which may have some safety benefit, but it is not clear if this safety benefit counteracts the general risk associated with walking and biking along rural highways with high speeds. There may be fewer crashes along roadways with paved shoulders, but these shoulders do not reduce (and may even increase) severity of crashes when they occur, perhaps due to increased travel speed due the wider visual cues created by paved shoulders.

  - *Characteristics of urban arterials:* The roadway design elements of urban arterials create quite high risk environments for all road users [69, 86, 87]. They typically have multiple lanes (often 4-6), including through lanes, turn lanes, and sometimes dedicated bus lanes and bus stops. Lane width is generally between 10 to 12 feet, with wider lanes accommodating higher volumes of traffic and larger vehicles such as buses and fleet trucks – and encouraging high driving speeds. On-street parking and many curb cuts are also common along these roadways, creating conflicts and obstructions. Individually each of these design features creates risk for people biking and walking, and collectively these environments create great risk.

- *Vehicular Traffic Volumes:* Roadways with higher traffic volumes create more risk for all roadway users, including bicyclists and pedestrians [3, 76]. Higher traffic volumes increase the number of interactions among road users generally, and also higher traffic volumes represent an increase in kinetic energy in the system (i.e., more heavy vehicles). Together, locations with turning movements and other conflict points *and* higher traffic volumes create great risk for people biking and walking. It is worth noting that vehicle miles traveled (VMT) *alone* may not be associated with an increased risk of collisions [3, 68], but increased VMT alongside other known risk factors, like obstructions and turning movements, likely will increase risk for bicyclists and pedestrians.

- *Bus stops:* Especially in urban environments, bus stops complicate risk for people biking and walking [88, 89]. For cyclists, bus stops often produce conflicts, and large passenger

buses often have poor visibility of users directly in front or behind. The bus stop creates many conflicts among all road users: bus-vehicle conflicts, especially when cars try to go around buses; bus-pedestrian conflicts, as people may be crossing the street to catch the bus; bus-bike conflicts, where buses may need to cross bike facilities to enter bus stops; vehicle-pedestrian conflicts, as the large bus may obscure pedestrians who are crossing the street; and pedestrian-bike conflicts, as cyclist infrastructure may conflict with pedestrian bus-boarding. But, the presence of bus stops can increase bike and pedestrian activity, which may lead to a change in driver expectation and thus increase their safety. Thus, the *design and implementation* of the bus stop is important to determine potential increases or decreases in risk for people biking and walking [90, 91].

- *Roadway and sidewalk maintenance:* The quality of travel way surfaces also creates risks for bicyclists and pedestrians, and this is especially true for single-bike crashes [68]. Trash, leaves, or other debris left in bike lanes or along sidewalks can create slipping or crash hazards, as can poor concrete or asphalt quality. Cracking due to roots or wear and potholes can causes several opportunities for harm, including cyclists and pedestrians swerving/entering the roadway to avoid the hazardous surface.

- *Density of development:* There is also a complex relationship among density of development and bike and pedestrian risk [88, 92]. More urban areas (i.e., more densely developed areas) are more likely to generate biking and walking travel, which may lead to an increase in crashes in some cases, but a decrease in crashes in others given the "safety in numbers effect" [63, 22]. These complex relationships also interact with other risk factors; there are more turning movements in more densely developed areas and more distractions among road users due to increased mental load [93].

- *Road user and vehicle characteristics:* While road user behaviors are not directly in the control of transportation agencies, they are an important part of roadway safety outcomes. These behaviors include traffic violations in the form of accidental errors [80, 94] and intentional

violations [94, 87]. These errors could be the result of substance use or perhaps more often turning maneuver mistakes [67].

Risk Factors, Exposure, and the Safe Systems Pyramid

Given the previously described risk factors, Table summarizes how each risk factor relates to the Safe Systems Pyramid's recommended countermeasures and exposure (measured in terms of bicycle and pedestrian volumes). In summary, this table indicates several takeaways for GDOT's investment in safety projects to reduce risk:

- *Potential bike and pedestrian countermeasures do not all have the same impact on safety.* As identified in the Safe Systems Pyramid [3], safety countermeasures that have the largest impact on population-level outcomes (i.e., reduce system-wide risk) are those that require the least amount of individual effort for the safety benefit to accrue. Those countermeasures in the "Education" category are less impactful for overall safety, and importantly, are not notably related to exposure, indicating another limitation of these countermeasures: they do not contribute the additional benefit of the safety in numbers effect.

- *Some countermeasures recommended by the Safe Systems Pyramid do not directly relate to risk factors as commonly discussed in academic and professional literature.* The SSP is a novel framework that prioritizes countermeasures by reduction of kinetic energy in the system and population-level effectiveness – a wholly different approach to risk factor measurement. Risk factors that are commonly mentioned in academic reviews and guidance about bike and pedestrian safety do not explicitly assume kinetic energy to be the agent of injury, which means that some risk factors identified in the literature do all not neatly fit into the SSP; in some cases, like examples from the "Socioeconomic Factors" level of the SSP, there are no risk factors identified in this review that align with these measures, as they are considered outside of the purview of typical engineering practice (i.e., affordable housing near transit). Importantly, however, these countermeasures are likely to see the safety ben-

efits generated by increased biking and walking volumes *and* system-wide kinetic energy reduction.

- *Countermeasures that influence exposure measured as bicycle/pedestrian volumes likely have additional safety benefits outside of the countermeasures themselves.* As GDOT chooses to invest in safety projects and develop policies for better roadway safety outcomes, they should consider those projects' and policies' ability to encourage more people to bike and walk, as these will be more effective than other investments.

Table 21: Relationship among risk factors, exposure, and the Safe Systems Pyramid [3]

Level of SSP	Potential Bicycle and Pedestrian Countermeasures	Risk Factors Addressed	Relationship to Volume as Exposure
Education	Texting and walking campaigns	Distracted walking	N/A
Active measures	Bike safety courses	Bicyclist compliance with road rules	N/A
	Yield signage and striping	Visual obstructions, speed	Drivers' participation in active measures like yielding can improve based on their expectations: as drivers see more people in crosswalks, the more primed they are to yield.
Latent measures	Bicycle helmets	Helmet use	N/A
	Leading pedestrian intervals	Visual obstructions, speed, wide crossing distance	LPIs give priority to pedestrians, which may encourage more walking in those places.
Built environment	Bike activated signal detection	Visual obstructions, speed, wide crossing distance	Bike-specific signals and other upgrades to bike infrastructure may encourage more cycling, increasing the safety in numbers effect.
	Separated bike infrastructure	Arterial characteristics, speed	Separated infrastructure encourage more people to cycle who may not feel comfortable on delineated facilities (i.e., painted bike lanes).
Socioeconomic factors	Curb extensions	Visual obstructions, speed, wide crossing distance	Curb extensions as a part of context designed for pedestrians can create safer and more attractive crossing locations.
	Affordable housing near transit	N/A	When more people take pedestrian-transit trips instead of driving, more people walk and increase the safety in numbers effect around transit stations.
	Zoning and development that reduces driving	Reducing system-wide kinetic energy	Fewer car trips may increase the number of people biking and walking for trips.
	Reducing evening shift work	Low-lighting risk, reducing system-wide kinetic energy	Less travel during low-lighting hours reduces risks for all travelers, including bicyclists and pedestrians.

CHAPTER 5: SAFETY METRICS – TRANSFERABLE ESTIMATION

Chapter 3 of this report developed a Bayesian framework to estimate hourly non-motorized volume estimations for locations where data was collected (i.e., in-sample predictions). These in-sample predictions were made from a model development process that used a *temporal* training-testing regime in two ways: (1) datasets with all Fridays removed were used to build (i.e., train) the models and then Fridays' volume predictions were compared against observed Friday volumes using error calculates (mean absolute error and root mean squared error); and (2) leave-one-out cross validation calculations were estimated and compared using estimation criteria. These approaches were helpful in determining the models' ability to estimate future volumes at locations where data had been collected. But it is also of interest to GDOT to understand how these models can be used to estimate volumes for locations where data is *not* collected (i.e., out-of-sample predictions).

This section of the report explores the potential for out-of-sample prediction for the pedestrian data collected in this report. Note that only pedestrian volume predictions are explored in this chapter; while it is also important to do the same for bike volumes in Georgia, several limitations preclude the possibility to do that in this study. First, out of the 40 locations where observed data were collected, only 9 locations had bicycle observations, which limits the amount of data that can be used in the models and likely will not result in meaningful out-of-sample predictions. Second, the bicycle models in Chapter 3, as discussed in the final section of the chapter, were generally poor quality in any locations outside of urban areas and in locations where the variance of average hourly volumes were quite high. Given that these models did not preform as well on in-sample predictions, it seemed unlikely that they would perform well for out-of-sample predictions. More data – both more total counts and longer observations at each location – are clearly needed to make out-of-sample predictions (see the subsection "Observation Length and Predictive Performance.)"

This chapter explores out-of-sample volume predictions using a generalized approach applied

to several specific iterations. Because it is not possible to directly measure the effectiveness of out-of-sample predictions (because we do not have data for out-of-sample locations), this chapter uses the same approach for temporal training-testing but instead applies it to a spatial dimension; here the authors "simulate" out-of-sample testing by removing some locations from the data, building models on the remaining locations, and then testing volume predictions' error on the withheld locations. The variable on which the data are separated is called the cross-cutting variable in this report; it is the metric used to determine similarity and difference among the locations, and random locations within this cross-cutting, or stratifying variable, are selected. This location-specific cross validation illuminates how well the models predict volumes for data that they have not "seen" before, thus replicating the experience of out-of-sample predictions using the data available.

The following sections explore the possibility of out-of-sample hourly prediction using two cross-cutting variables: context (i.e., urban, suburban, and rural town), and clustering assignment (see Chapter 2).

GENERAL APPROACH: MODEL SPECIFICATION AND TESTING SITES

To determine the Bayesian time series models' ability to predict out-of-sample, this report presents a cross validation using a generalized approach applied to two different cross-cutting variables. First, the data were separated into training and testing by location and by cross-cutting variable. Within each category of the cross-cutting variable, locations were randomly selected to be in the training or testing datasets so that approximately 70% of all data were used as training data to calibrate the model and 30% of all data were withheld for testing. Then, the models were calibrated with the training data, creating posterior predictions using MCMC sampling (see Chapter 3 for a more complete explanation of this process). Then, more samples from the posterior predictive distribution of the response variable (Observed counts per hour) *given new data* (the testing data) were collected by drawing from the posterior distribution of the model parameters estimated during model fitting [60]. Like the previous models, estimates are made in the form of *distributions*, not point estimates, but point predictions are generated as the mean of the posterior parameters'

distribution can be calculated to create a single point estimate.

The locations withheld as testing data in the following analyses are shown in Table below.

Table 22: Locations withheld for testing and their cross-cutting variables

Location ID	Location Name	Jurisdiction	Context	Cluster Assignment
4	Handcock Street	Milledgeville	Rural town	1
17	MLK Drive NW	Fulton County	Suburban	1
23	Cherry Street	Macon	Rural Town	1
37	17th Street	Fulton County	Urban	2
38	Mercer University Drive	Macon	Rural town	1
41	Spring Street	Fulton County	Urban	2

The models used in the spatial cross-validation follow the same structure as the previously developed models, but they are modified slightly for this application. Generally, given the following definitions:

- N_i = number of observed pedestrians per intersection per hour;
- P_i = count of pedestrians from proxy data per hour;
- $f(H)_C$ = a spline fit to the hourly data, where the spline is indexed by cross-cutting variable C and captures the temporal autocorrelation for each observed hour H;
- α_C = variable-specific intercept, capturing the population-level effect (i.e., average effect) per cross-cutting variable;

the chosen model specification is:

$$N_i \sim GP\ (\lambda,\ \phi)$$

$$\lambda = \alpha_C + \beta_P P * C + f(H)_C$$

$$\alpha_C \sim Normal$$

$$\beta_P \sim Normal$$

$$C \sim Normal$$

$$f(H) \sim Student's\ T$$

$$\phi \sim Exponential$$

While the structure of this model is the same as those in Chapter 3, there are two main differences: First, instead of location-specific interactions with β_P and the spline $f(H)$, these terms were instead interacted with the cross-cutting variable, C – which is either the locations' context or the cluster assignment. Second, the intercept term has been allowed to vary by cross-cutting variable to improve predictions; the location-specific predictions in Chapter 3 captured more location-specific information in its predictions, but since location could not be used as a predictor variable here, these models' cross-cutting-specific intercepts attempt to capture more information from each category.

Daily Models

In addition to this chosen method for out-of-sample predictions, the authors also considered an additional approach: collapsing the hourly data into daily data to make predictions at the daily level. The rationale behind this decision is that daily models may have a smaller error given the lower-resolution predictions (temporally). The predictions from these models, however, were quite poor quality, even when making in-sample predictions. Likely this is because many of the locations were reduced to just three data points (i.e., three days of data) instead of the hourly models $3\ days * 15\ hours = 45$ data points. So by reducing the information in the model, the predictions actually became poorer generally with much higher error. Given this finding, the authors chose

to continue with hourly models. Also, as stated in previous sections of the report, hourly models were originally selected as the unit of time intentionally: if the goal of understanding walking (and biking) travel trends is to improve safety outcomes, it is important to understand how volumes vary by time of day because cyclists risk also varies by time of day (see the previous chapter's discussion of traffic volumes, lighting and visibility, etc.). Given this understanding of the importance of hourly variation, it is better for GDOT to have these hourly models for future predictions of walking.

CONTEXT CROSS-VALIDATION

The first cross-cutting variable used in this research is context: urban, suburban, and rural town. Given the model specification and testing data described in the previous section, out-of-sample predictions and their uncertainty from select locations are shown in Figure through Figure . This includes one location with 90 days of observations (all other testing data has three days of observation).

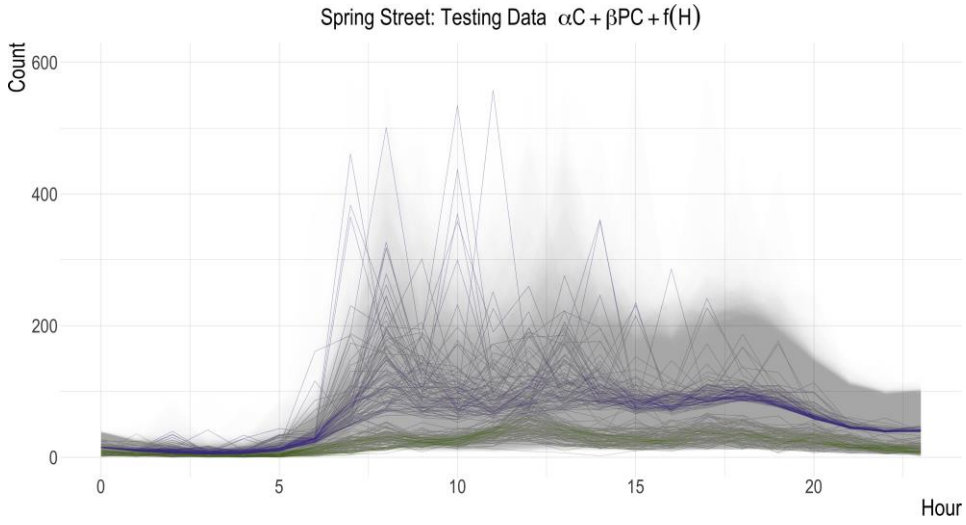


Figure 79: Graph. Spring Street in Atlanta, Fulton County, out-of-sample predictions (purple) versus observed (light green) pedestrian volumes using the context model with a 90% credibility region (translucent purple)

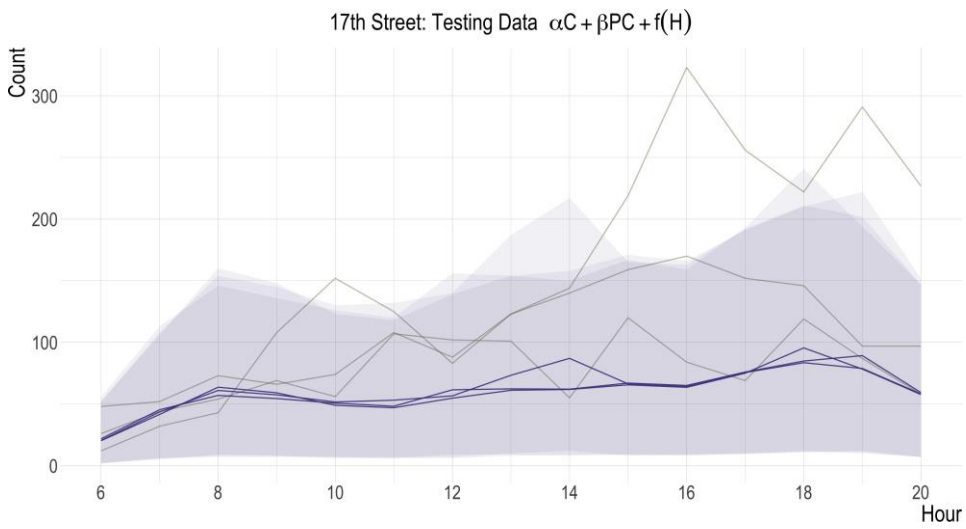


Figure 80: Graph. 17th Street in Fulton County, out-of-sample predictions (purple) versus observed (light green) pedestrian volumes using the context model with a 90% credibility region (translucent purple)

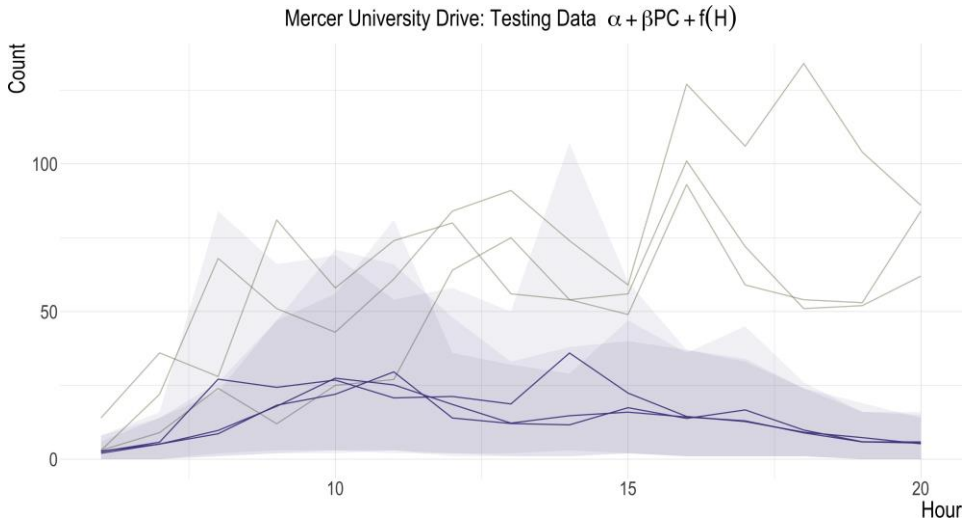


Figure 81: Graph. Mercer Street Street in Macon out-of-sample predictions (purple) versus observed (light green) pedestrian volumes using the context model with a 90% credibility region (translucent purple)

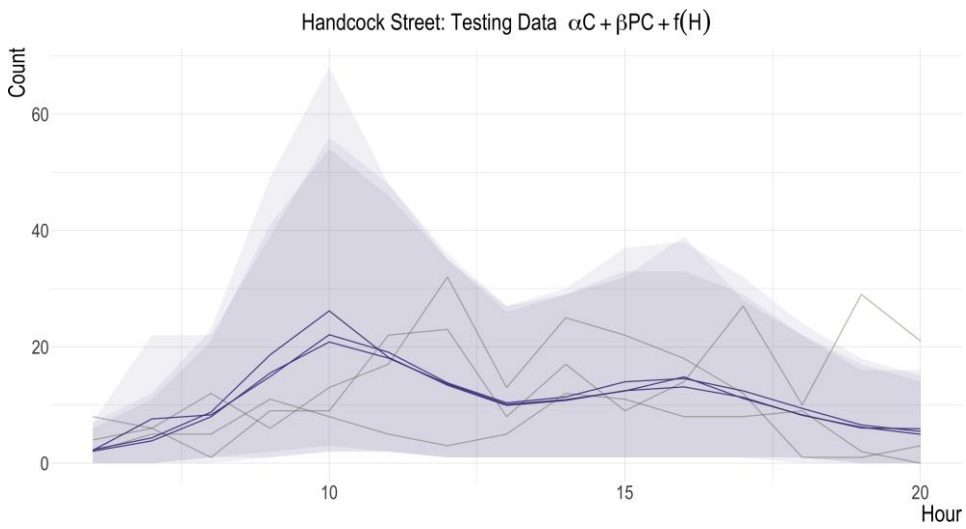


Figure 82: Graph. Handcock Street in Milledgeville out-of-sample predictions (purple) versus observed (light green) pedestrian volumes using the context model with a 90% credibility region (translucent purple)

From each of these predictions, it is important to note the following findings: First, the out-sample predictions are on the same order of magnitude as observed data. This is good news; in general, there are similar scales of predictions as the observed data for those locations. These predictions are not, however, as high-quality as the predictions in Chapter 3. For example, in Figure .The predictions have a high degree of uncertainty (up to 400 pedestrians per hour), more randomness in the mid-morning predictions, and in general the average of these predictions are higher than the observed data. In most other cases, the predictions are under-predicting for each hour at each location (like Mercer Street in Figure and 17th Street in Figure . The Handcock Street (Milledgeville) prediction, the highest quality of the testing data, with the only issue being the high levels of uncertainty around mid-morning predictions.

The lower quality of these predictions are expected, to an extent. Most of the cases in this data are based on only three days of observation, which is an excellent beginning to calibrate models and, in some cases, presents enough information to make in-sample predictions. But out- of-sample predictions typically require much more information to have high levels of accuracy, as well as more specific cross-cutting variables. The fact that these predictions in some cases follow

similar trends as the observed data and are on the same order of magnitude given such limited observed data in these locations showcases the strength of Bayesian time series modeling.

The next two figures – Figure and Figure – show two different ways of calculating error for these models for both in-sample error and out-of-sample error: error by location, and error by context type. Both figures show MAE and RMSE. The color codes relate differences in training versus testing, and the transparency differences relate to the measure of error, where the solid circle represents MAE and the translucent circle represents RMSE. Collectively, these two graphics help illustrate model accuracy by specific location and by the cross-cutting variable.

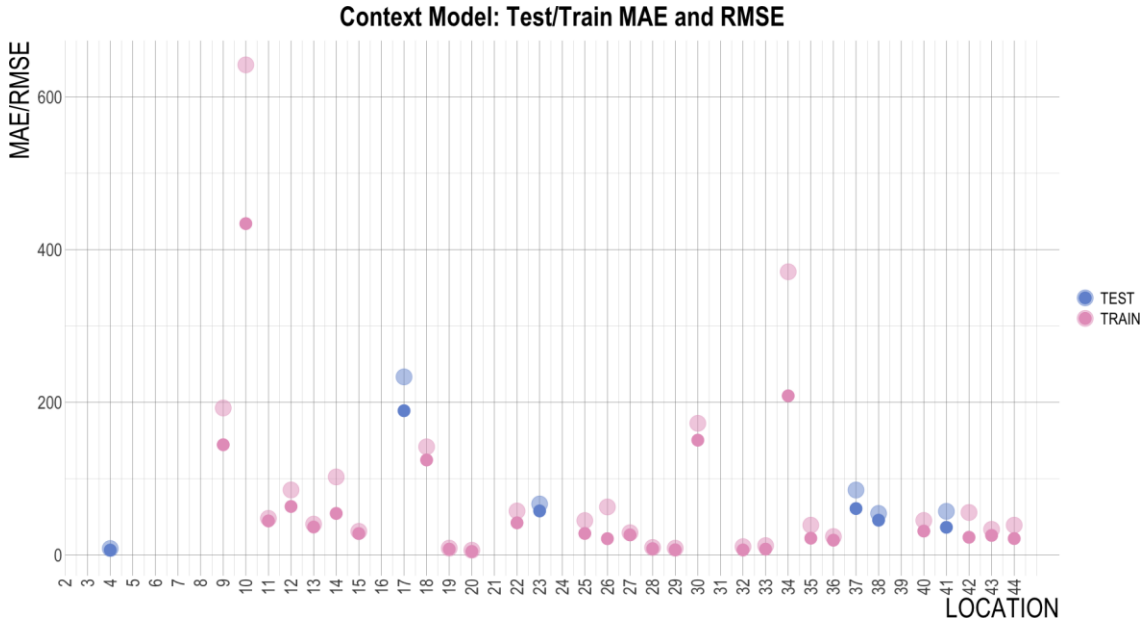


Figure 83: Graph. Context model MAE (solid circle) and RMSE (translucent circle) for testing and training data by location

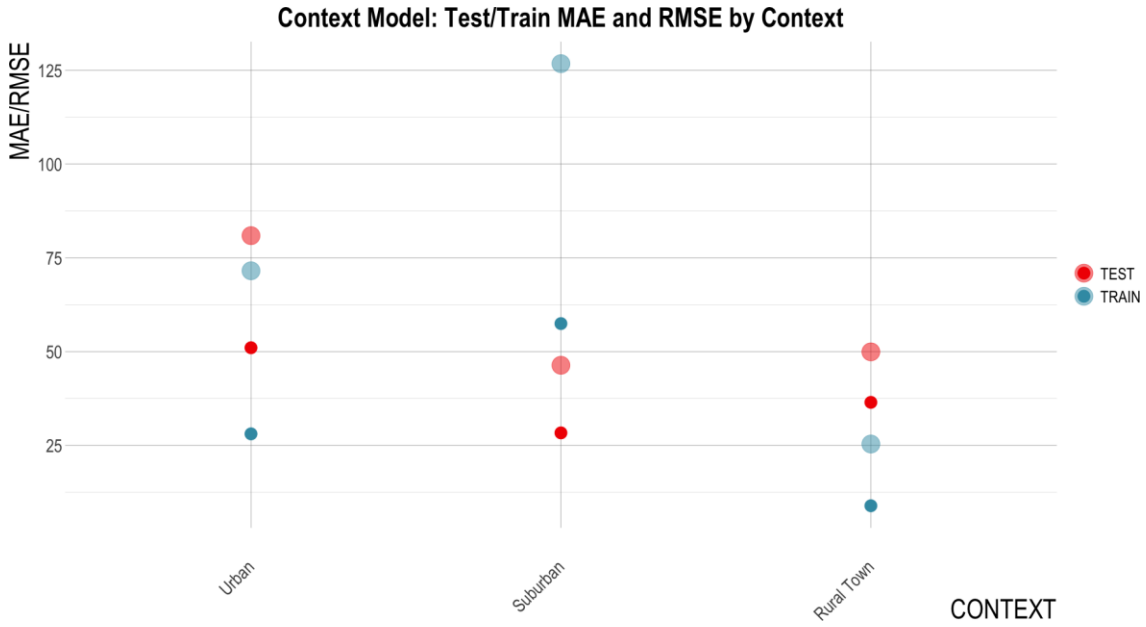


Figure 84: Graph. Context model MAE (solid circle) and RMSE (translucent circle) for testing and training data by context

When looking at the error by location, results show that the location with the highest error is Location 10, Krog Street, which was included in the training data. As shown in Chapter 3, this location had relatively high error even in the location-specific models with in-sample predictions and temporal cross-validation. This location had so much variation among the observed data across the three days that the model was not able to predict this location well, regardless of specification (or mode; the bike models also performed poorly here). To improve predictions at this location, more days of observation should be collected, likely at least two weeks (see the end of Chapter 3 for this discussion).

Within the testing locations, Location 4, Handcock Street, had the lowest error ($RMSE = 8.34$ and $MAE = 6.33$), and Location 17, MLK Drive, had the highest error ($RMSE = 233.0$ and $MAE = 189.0$). Across all testing locations, the mean values of error were $MAE = 65.9$ and $RMSE = 84.1$, and those averages dropped to $MAE = 41.3$ and $RMSE = 54.3$ when Location 17 was not considered in the average, as it appears to be hard to predict out-of-sample and had relatively high error in-sample as well (see Figure). This is notably high error of pedestrians

per hour, especially in places where there are low daily volumes, like in some of the suburban locations. The out-of-sample predictions did well in the rural town location, though, so likely these models would produce relatively accurate predictions of high-resolution pedestrian volumes in those locations.

Figure shows how the average error (both MAE and RMSE) vary by contextual category (urban, suburban, and rural town). Testing error was highest in urban locations, followed by suburban locations, and then rural town locations. Interestingly, the suburban locations training error measures were much higher than testing error, which is unexpected. This appears to be due to a greater number of outlier data points in the suburban training data's observations. This is mirrored in the finding that the average RMSE, which is more sensitive to outliers, is much greater that the average MAE in the suburban training data. The rural town's testing error was the lowest in the group, which is interesting and points to what is perhaps the most important factor in prediction accuracy: even though there were relatively few pedestrian observations in the rural town locations, there was much lower variation overall at those locations, which makes it easier to predict out-of-sample volumes with minimal error.

CLUSTER CROSS-VALIDATION

The second cross-cutting variable used in this research was cluster assignment. For this analysis, we used the the dynamic time warping $K = 2$ solution, as it had high quality CVIs and appeared to be the most separated when plotted in PCA space (see Figure). Results from these predictions are shown below in Figure through Figure , which are the same selection of testing locations as shown in the previous sections.

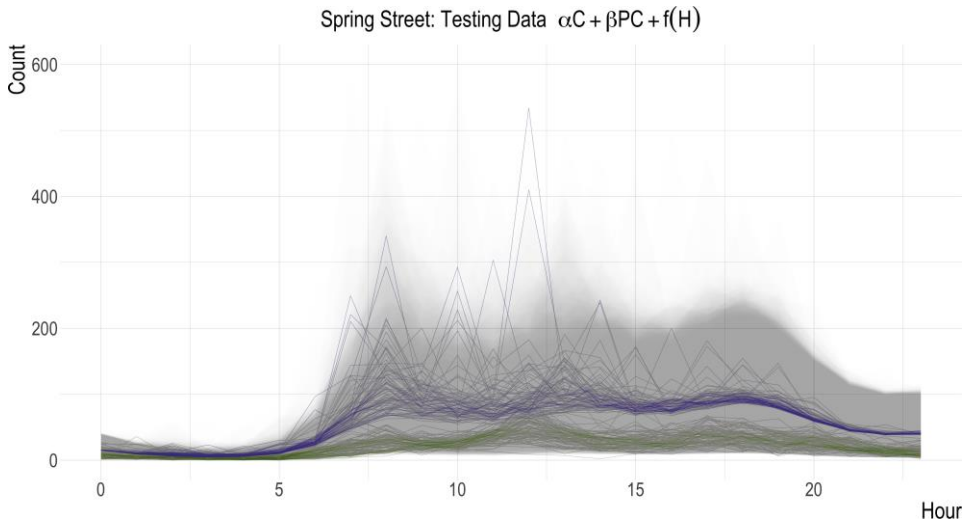


Figure 85: Graph. Spring Street in Atlanta, Fulton County, out-of-sample predictions (purple) versus observed (light green) pedestrian volumes using the cluster model with a 90% credibility region (translucent purple)

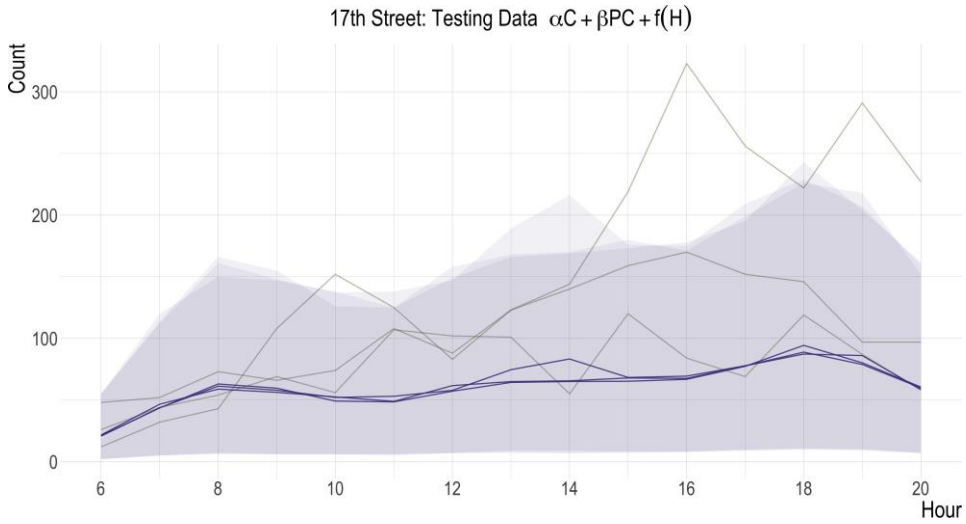


Figure 86: Graph. 17th Street in Fulton County, out-of-sample predictions (purple) versus observed (light green) pedestrian volumes using the cluster model with a 90% credibility region (translucent purple)

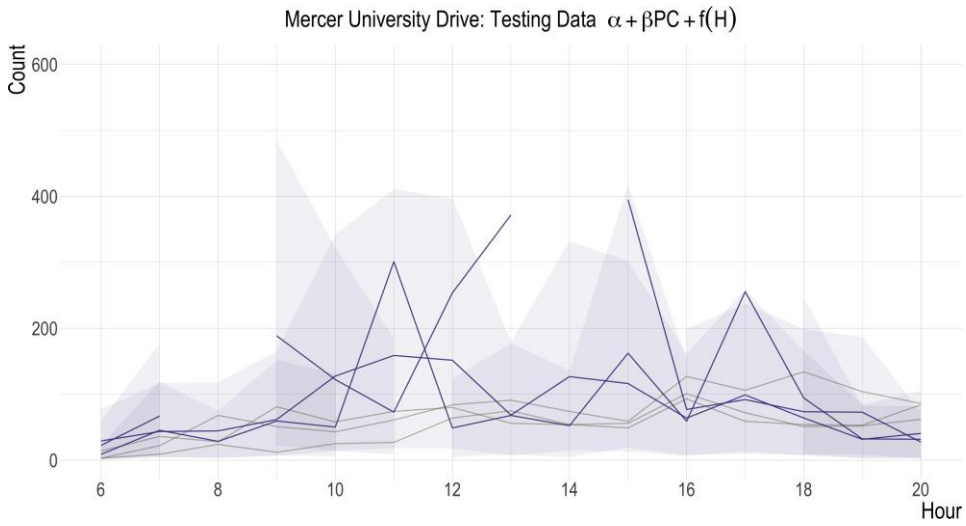


Figure 87: Graph. Mercer Street Street in Macon out-of-sample predictions (purple) versus observed (light green) pedestrian volumes using the cluster model with a 90% credibility region (translucent purple)

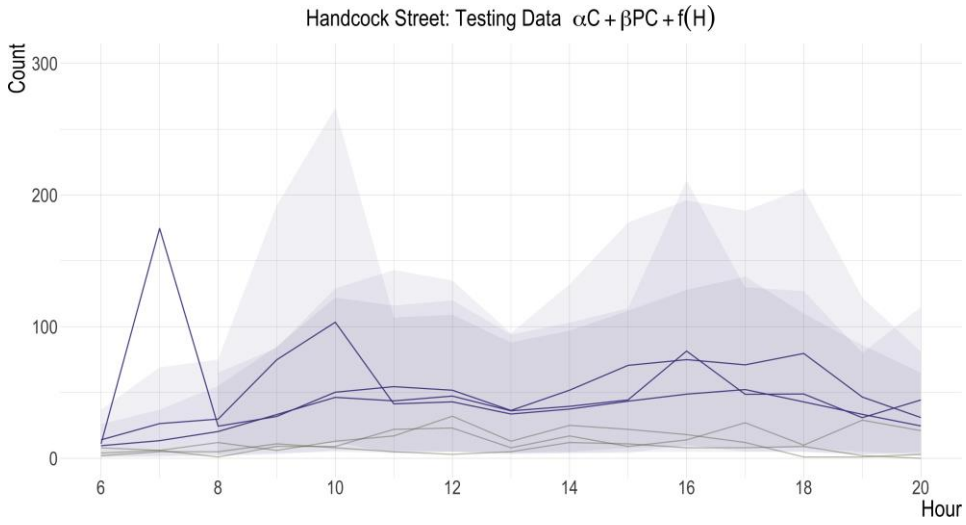


Figure 88: Graph. Handcock Street in Milledgeville out-of-sample predictions (purple) versus observed (light green) pedestrian volumes using the context model with a 90% credibility region (translucent purple)

In many instances, the predictions using this cluster cross-cutting variable look similar to the predictions made when the cross-cutting variable was context. For example, the 17th Street prediction looks nearly the same, except context predictions have slightly smaller ranges of uncertainty.

The same is true with the Spring Street location, although there are fewer extreme predictions in the cluster assignment predictions. For Handcock Street and Mercer Street however, the cluster assignment predictions were much worse than the context predictions. These are both rural town locations, which indicates that the clustering assignment is not as helpful of a cross-cutting variable as context in these locations, pointing to a likely weakness in the time series clustering approach. This is a reasonable finding; in general, there are fewer actuations per hour and more hours that were zero in the ATSPM data in rural towns, so perhaps clustering on ATSPM data in more rural places cannot meaningfully separate time series trends.

Next, RMSE and MAE are summarized by location and by cluster assignment in Figure and Figure , respectively. While the predictions in this section in some cases show similar accuracy as those generated from the context cross-cutting variable, these two figures show that in general the time series clustering was a poorer means of describing location. In general, the error is much higher at these locations, especially for Location 17, which had RMSE and MAE values of 1222 and 372, respectively. Similarly, Figure shows much higher error by the cross cutting variable categories than the context error (Figure), especially for the first cluster.

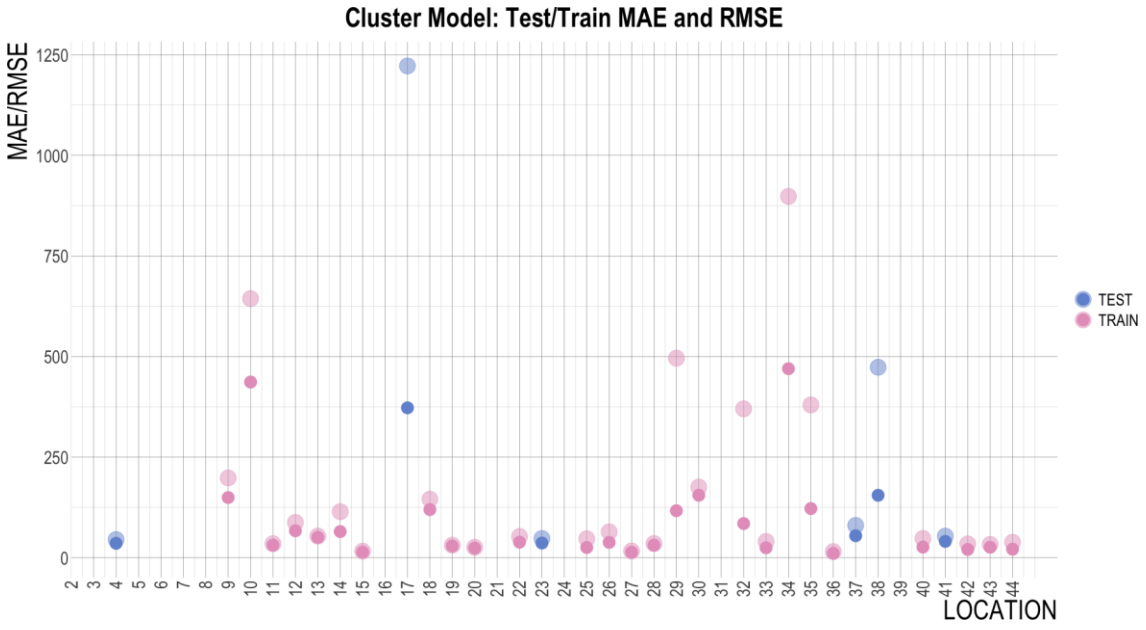


Figure 89: Graph. Cluster model MAE (solid circle) and RMSE (translucent circle) for testing and training data by location

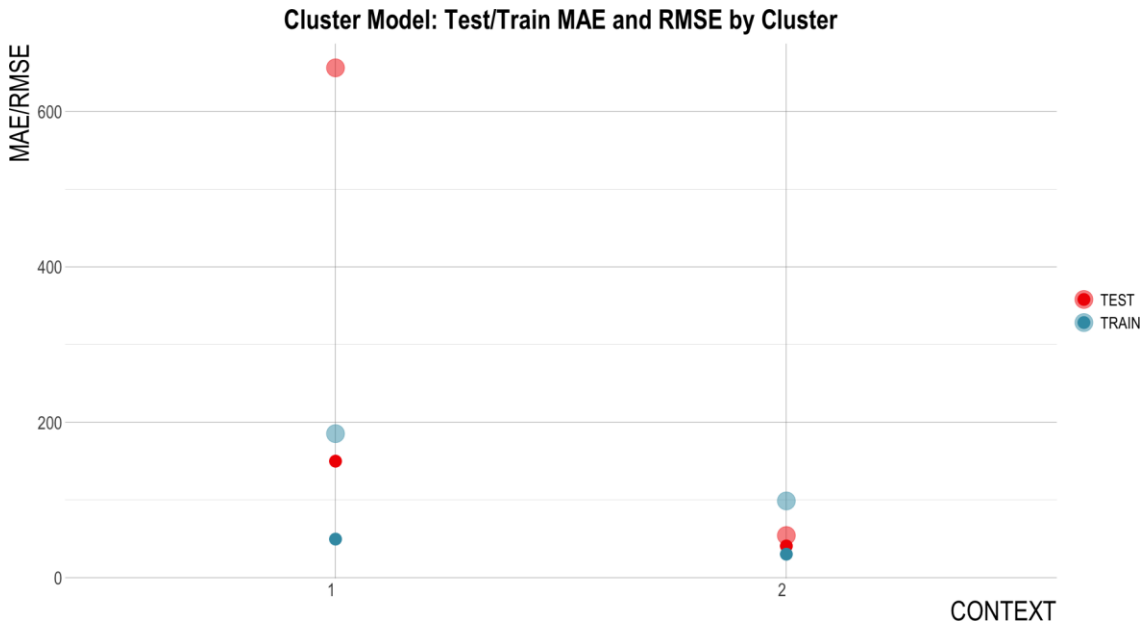


Figure 90: Graph. Cluster model MAE (solid circle) and RMSE (translucent circle) for testing and training data by cluster assignment, using PAM 1 and the DTW, K = 2 solution

CONCLUSION

This chapter explored the possibility of using the modeling framework developed in Chapter 3 to predict hourly pedestrian volumes at "new" locations – places where the model has not been trained. In general, the findings from this chapter point to a clear reality: even sophisticated models that can handle large amounts of variation and zero inflation need to be trained on more data or need to be calibrated with at least some location-specific observations. While findings in the first cross-validation approach (using a context cross-cutting variable) showed that in some locations, predictions could could be precise enough to be meaningful – an error of 5-10 pedestrians per hour – in locations where the hourly variance is large *relative to the number of observations*, predictions both in-sample and out-of-sample are poor, meaning that more observations are needed. See this illustrated in Table . In these cases, the mean is smaller than the MAE, which points to limited value of these estimates; for example, if the average of number of pedestrians per hour at a location is 10, but the MAE is 15 pedestrians per hour, then that means the range of possible pedestrians per

hour with these estimates are 0-30 pedestrians per hour, which does not even indicate if pedestrians are present at the location at all.

Table 23: Relationship among error, hourly average, and variance

Location	Hourly Average	Hourly Variance	RMSE	MAE	Average-MAE	Test/ Training Data
28	2.5	5.6	9.7	8.2	-5.7	TRAIN
19	2.7	6.6	8.9	7.6	-4.9	TRAIN
11	3.2	10.1	47.9	44.5	-41.4	TRAIN
20	7.3	25.9	6.1	4.6	2.7	TRAIN
13	10.2	97.8	40.3	36.6	-26.4	TRAIN
32	10.5	73.1	10.8	6.5	4.0	TRAIN
4	10.9	65.6	8.3	6.3	4.6	TEST
29	11.8	92.1	8.8	6.7	5.1	TRAIN
33	13.0	150.8	12.2	7.6	5.5	TRAIN
15	18.8	154.3	30.9	28.0	-9.1	TRAIN
27	21.9	143.8	29.0	26.2	-4.4	TRAIN
41	22.2	253.7	57.1	36.2	-14.0	TEST
43	23.0	459.8	33.5	25.6	-2.6	TRAIN
26	27.0	3987.8	62.8	21.3	5.8	TRAIN
36	29.3	253.5	24.0	19.3	10.1	TRAIN
40	35.6	2067.7	45.1	31.2	4.4	TRAIN
25	47.4	2273.9	45.0	28.1	19.3	TRAIN
35	47.5	541.2	39.0	21.9	25.6	TRAIN
44	56.3	2194.1	38.9	21.4	34.9	TRAIN
42	58.0	1683.1	55.6	23.1	34.9	TRAIN
38	59.4	960.2	54.3	45.5	13.9	TEST

Table 23: Relationship among error, hourly average, and variance

Location	Hourly Average	Hourly Variance	RMSE	MAE	Average-MAE	Test/ Training Data
23	67.9	1188.4	66.8	57.6	10.3	TEST
14	94.6	9649.9	102.2	54.3	40.2	TRAIN
22	98.0	2118.2	57.5	42.0	56.0	TRAIN
12	99.8	5318.2	85.1	63.5	36.4	TRAIN
37	113.4	4820.8	85.0	60.7	52.7	TEST
18	183.3	6784.5	141.6	124.4	58.9	TRAIN
34	202.2	5459.3	370.9	208.4	-6.2	TRAIN
30	230.5	8356.3	172.4	150.3	80.2	TRAIN
17	245.8	12952.4	233.3	188.9	56.9	TEST
9	247.5	17977.5	192.5	144.3	103.2	TRAIN
10	508.0	235947.0	642.1	434.0	73.9	TRAIN

A notable reason for the limited ability of these models to perform well out-of-sample compared to the models in Chapter 3 is that the models in in temporally-specified models (Chapter 3) allowed location-specific predictions in two ways: the relationship between the proxy variable (ATSPM or Strava) and the location were allowed to vary by location, and the hourly spline was also allowed to vary by location. This modeling approach is not possible when using out-of-sample predictions, as the model treats these locations as categorical variables and thus cannot allow new levels of prediction in an update; said another way, when the model is specified to vary by location in its time and proxy relationships, it cannot define a new relationship for locations that were not seen in the training data. And, as shown in this chapter, without this location-specific dependence in the model (i.e., when other variables must be used to determine the relationship among variables, like context) the predictions are less accurate and more uncertain; the predictions

shown in the Chapter 3 have much lower uncertainty and generally lower error that those in this

chapter. Predicting in-sample requires less information than predicting out-of-sample when there is meaningful variation among the locations in the sample, which is true in this case.

Also, in general, predicting future values for locations where some trends have already been observed is also more difficult that predicting values for locations that are not monitored, as the trends on which the model are trained must be present and well captured in the model. Likely in this case, low accuracy out-of-sample predictions would be improved with more days of observation (especially in the three-day locations) *and* more specific cross-cutting variables, which can be explored in future work.

CHAPTER 6: CONCLUSION AND RECOMMENDATIONS

In summary, this research project explored the following facets of non-motorized volumes, proxy data sets, and model estimation in Georgia:

- *Overview of biking and walking safety in the US and in Georgia:* The first chapter investigated the relationship between safety and exposure for VRUs. It also summarized non-motorized counting practices at the agency level and research that leverages proxy data to estimate biking and walking volumes.

- *Categorization of proxy-based non-motorized travel trends:* Chapter 2 of this report explored the potential of unsupervised machine learning to create categories of non-motorized traveler proxy data, using Strava Metro data as a proxy for bikes, and ATSPM pedestrian actuations as the proxy for pedestrians. This work found that it is possible to create distinct clusters of travel trends from this proxy data, although the number of distinct clusters is small in both datasets. It also found that if GDOT is interested in pursuing further categorization of biking and walking trends, simpler forms of clustering and feature extraction (using features commonly used to describe traffic volumes) is an equal substitute for more complex forms of unsupervised machine learning.

- *Bayesian time series models to estimate non-motorized volumes at locations where data was collected:* The next section, Chapter 3, developed a generalized framework for modeling non-motorized counts' relationship to proxy data (ATSPM data for pedestrian volumes and Strava data for bicycle volumes) based in Bayesian time series modeling. Bayesian models are more amenable to limited data than other modeling approaches, and they are also well suited for modeling sparse data – both of which are true for the challenge of modeling biking and walking volumes in Georgia. Results from this section point to several important

findings for research and non-motorized counting in Georgia. First, hourly Bayesian time series can effectively model the relationship among proxy data and observed non-motorized counts at locations where there is enough observed data *and* there is limited hourly variation; these models worked well in rural town settings – where there are consistent trends of activity across days – and in urban settings – where enough data was collected and hourly variations were low. In some cases, there is a threshold to how much prediction error can be reduced because of the underlying large amount of variation across hours and days generally. The ATSPM-pedestrian models worked well in many settings, and the Strava-bike models worked best in urban locations.

- *Risk factors for bicyclists and pedestrians and their relationship to travel volumes:* This chapter synthesized the literature of bicycle and pedestrian risk factors, exploring how each factor relates to exposure and safety countermeasures categorized by the Safe Systems Pyramid [3]. Findings from this chapter should direct GDOT towards the types of projects that will most effectively reduce biking and walking injuries and deaths and encourage more people to bike and walk.

- *Bayesian time series models to estimated volumes at locations where data was not collected (i.e., out-of-sample predictions):* The final chapter of this report explores how the modeling framework developed in this research can be used to predict volumes at locations where data is not collected. This work explored this possibility through a testing-training framework, where the data were separated into training and testing data based on "cross-cutting variables," context (i.e., urban, rural town, and suburban) and clustering assignment (from Chapter 2). This chapter showed the difficulties of predicting out of sample with limited observational data; while the context cross-cutting approach did produce estimates with reasonable error in some places, at locations where variation was high relative to the number of hourly observed values, then predictions were too poor for meaningful use. To improve out-of-sample spatial predictions, more observed data is needed and more sophisticated cross-

cutting variables should be developed. Project results show that data of 14 days or more provides the information needed to improve on volume estimates.

Synthesizing the findings from each of these sections, this report makes the following recommendations for the Georgia Department of Transportation as it continues to explore bicycle and pedestrian travel trends and improve their safety outcomes. The authors have outlined these recommendations into two categories: non-motorized monitoring and safety investments.

NON-MOTORIZED MONITORING RECOMMENDATIONS

The authors make the following recommendations about biking and walking monitoring in Georgia:

- *Make use of short-term counts to improve on developed models:* A major benefit of using the modeling structure proposed in this report is that these models are "living" in a way and able to be continually updated based on new information: new data can be added to the model parameters' posteriors at any point to create predictions for future volumes at those locations. This points to the opportunity for GDOT to make use of its archived and future counts that it collected/collects as a part of other projects. For example, GDOT has archived counts of bike and pedestrian volumes at every location where they have conducted traffic studies. At locations where GDOT and Strava data are available (i.e., Fulton County for Strava data and the over 7,000 intersections where ATSPM data are collected), these counts can be paired with their respective proxy data and incorporated into the Bayesian model, regardless of how long or short the counts are conducted. Similarly, when new counts are conducted, they should be paired when their respective proxies and likewise can be used to produce future count estimates at those locations. This is an important opportunity for GDOT; because of the models' ability to "learn" across locations, less data is needed to make high-quality, high-resolution predictions at locations where only small amounts of observed data are available. Based on the results of this project, the authors recommend that

the Bayesian models developed in this work be used as a basis for updating with additional data sources to provide high-resolution hourly estimates of volumes at specific locations.

- *Develop seasonal adjustment factors via long-term counters:* The scope of this project used short-term counts to create high-resolution estimates of biking and walking volumes. While the authors successfully explored this possibility, an important factor that was not explored given the scope is seasonal adjustment factors [25]. As described in the literature review, these factors accommodate for seasonal changes in travel volumes, but these factors must be developed from long-term counts conducted at key locations in the network. It is important, then, to conduct more long term counts and use the methods described in the Travel Monitoring Guides to develop these adjustment factors, which can then be incorporated into the time series models to improve predictions. To do this, GDOT should explore the possibility of investing in long-term counters, or partnering with other organizations in the region (e.g., Midtown Alliance) to leverage their long-term counts.

- *Explore the possibility of a count program in GDOT:* Like most other states in the U.S., GDOT currently does not have a non-motorized count program embedded in the agency, and understandably so – these programs require investment into counting equipment and staff to design the program, maintain counters, and move them around the network systematically. As GDOT prioritizes its investments, this project recommends that GDOT consider a statewide non-motorized count program as one of its possible investments. While systematic count programs do require consistent investment, they create immense value: they provide justification for bike and pedestrian projects, inform safety decisions, and provide feedback on the impact of infrastructure investments that can be used to apply for federal funding. GDOT should consider the possibility of starting one of these programs with the support of other local agencies, like Atlanta DOT or the Atlanta Regional Commission. There are high-quality examples of these programs in the U.S. that GDOT could use for reference for creating one of these programs, including Minnesota DOT [95] and Colorado DOT [96].

Based on the results of this project, it is shown that even counts of 14 days at a specific location can be used to increase confidence in volume predictions at those locations.

SAFETY INVESTMENT RECOMMENDATIONS

Finally, the authors make the following recommendations about roadway safety projects to improve biking and walking safety in Georgia:

- *Prioritize projects that increase volumes and separation:* Chapter 4 showed that the projects that increase traffic volumes, like new bike infrastructure that connects to an existing network, create additional benefit from the safety in numbers effect [22]: as more people bike and walk in an area, the likelihood of risk per person decreases. When prioritizing safety and/or non-motorized projects, then, GDOT should invest in those that *both* increase separation for people biking and walking (i.e., trails, separated bike paths, etc.) – which decreases crash likelihood and is more attractive to a wider range of users – *and* those that generate new ridership – which decreases the risk per user. Collectively, these two powerful approaches to safety projects can improve safety outcomes in Georgia.

- *Begin exploring metrics for higher resolution safety assessments:* This research generated hourly estimates of biking and walking for several reasons. Annualized values of activity are problematically reductive in that they (1) imply more accuracy than they actually contain, (2) they do not capture travel actual variability, and (3) losing this information precludes understanding how non-motorized travelers' risk varies with the time of day. As shown in Chapter 4, low-lighting hours and high vehicular traffic volume hours may correspond with higher crash rates in Georgia, but it is not clear if that is the case or in which contexts this occurs. Understanding how crashes relate to hourly volumes will help inform more specific safety countermeasure selection, which can help GDOT more strategically prioritize safety projects. Given this justification, GDOT should investigate new ways of measuring risk-exposure at higher resolution, perhaps at the hourly level as presented in this work. This can

be done through additional research projects or learning from academic literature.

ACKNOWLEDGMENTS

This report is funded and directed by the Georgia Department of Transportation through RP 21-09. Supplemental support for this research came from Dwight David Eisenhower Transportation Fellowship Program (DDETFP) of the Federal Highway Administration. This work has also been advised by Michael Rodgers, PhD (Georgia Institute of Technology) and Jonathan Rupp, PhD (Emory University).

REFERENCES

[1] K. Nordback, S. Kothuri, D. Johnstone, G. Lindsey, S. Ryan, and J. Raw, "Minimizing annual average daily nonmotorized traffic estimation errors: How many counters are needed per factor group?" *Transportation Research Record*, vol. 2673, pp. 295–310, 10 Jan. 2019.

[2] M. Wallace, C. Kitson, M. Ormstrup, J. Cherian, and J. H. Saleh, "Pedestrian and light transit accidents: An examination of street redesigns in Atlanta and their safety outcomes," *Case Studies on Transport Policy*, vol. 9, pp. 538–554, 2 Jun. 2021.

[3] D. Ederer, R. T. Panik, K. Watkins, and N. Botchwey, "The safe systems pyramid: A new framework for traffic safety," *Transportation Research Interdisciplinary Perspectives*, vol. 21, 21 Aug. 2023.

[4] "Georgia Traffic Safety Facts: Pedestrians and Bicyclists 2019," Governor's Office of Highway Safety, Apr. 2021, Received from: www.gahighaysafety.org, pp. 1–15.

[5] P. B. Ohlms, L. E. Dougald, and H. E. MacKnight, "Bicycle and pedestrian count programs: Scan of current U.S. practice," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, pp. 74–85, 3 Mar. 2019.

[6] P. Schepers, M. Hagenzieker, R. Methorst, B. V. Wee, and F. Wegman, "A conceptual framework for road safety and mobility applied to cycling safety," *Accident Analysis and Prevention*, vol. 62, pp. 331–340, Jan. 2014.

[7] H. W. Heinrich, *Industrial accident prevention: A scientific approach*. McGraw-Hill., 1931.

[8] Y. Toft, G. Dell, K. Klockner, and A. Hutton, "Models of causation: Safety," *The Core Body of Knowledge for Generalist OHS Professionals*, pp. 1–25, Jan. 2012.

[9] S. Othman, R. Thomson, and G. Lannér, "Identifying critical road geometry parameters affecting crash rate and crash type," *Annals of Advances in Automotive Medicine / Annual Scientific Conference*, vol. 53, p. 155, Oct. 2009.

[10] C. W. Runyan, "Using the Haddon matrix: Introducing the third dimension," *Injury Prevention*, vol. 4, pp. 302–307, 4 Dec. 1998.

[11] L. A. Merlin, E. Guerra, and E. Dumbaugh, "Crash risk, crash exposure, and the built environment: A conceptual review," *Accident Analysis & Prevention*, vol. 134, p. 105 244, Jan. 2020.

[12] S. Turner *et al.*, "Synthesis of methods for estimating pedestrian and bicyclist exposure to risk at areawide levels and on specific transportation facilities," U.S. Federal Highway Administration, Jan. 2017, pp. 1–93.

[13] N. Dong, F. Meng, J. Zhang, S. C. Wong, and P. Xu, "Towards activity-based exposure measures in spatial analysis of pedestrian–motor vehicle crashes," *Accident Analysis & Prevention*, vol. 148, 2020.

[14] "NCHRP 08-36, task 141: Evaluation of walk and bicycle demand modeling practice," American Association of State Highway Transportation Officials, Tech. Rep., 2019.

[15] R. L. Sanders, "Perceived traffic risk for cyclists: The impact of near miss and collision experiences," *Accident Analysis and Prevention*, vol. 75, pp. 26–34, Feb. 2015.

[16] R. L. Sanders, R. J. Schneider, and F. R. Proulx, "Pedestrian fatalities in darkness: What do we know, and what can be done?" *Transport Policy*, vol. 120, pp. 23–39, May 2022.

[17] A. Tarko, G. Davis, N. Saunier, T. Sayed, and S. Washington, "Surrogate measures of safety," Transportation Research Board ANB20(3) Subcommittee on Surrogate Measures of Safety, Apr. 2009, pp. 1–13.

[18] C. Johnsson, A. Laureshyn, and T. D. Ceunynck, "In search of surrogate safety indicators for vulnerable road users: A review of surrogate safety indicators," *Transport Reviews*, vol. 38, 6 Nov. 2018.

[19] A. Laureshyn, M. de Goede, N. Saunier, and A. Fyhri, "Cross-comparison of three surrogate safety methods to diagnose cyclist safety problems at intersections in Norway," vol. 105, Elsevier, Aug. 2017, pp 11–20.

[20] C. Johnsson, A. Laureshyn, and C. D'Agostino, "Validation of surrogate measures of safety with a focus on bicyclist–motor vehicle interactions," *Accident Analysis & Prevention*, vol. 153, Apr. 2021.

[21] C. Hydén, "The development of a method for traffic safety evaluation: The Swedish traffic conflicttechnique," Lund Institute of Technology, 1987, pp. 1–312.

[22] R. Elvik, "The non-linearity of risk and the promotion of environmentally sustainable transport," *Accident Analysis & Prevention*, vol. 41, pp. 849–855, 4 Jul. 2009.

[23] F. Ye and D. Lord, "Investigation of effects of underreporting crash data on three commonly used traffic crash severity models," *https://doi.org/10.3141/2241-06*, pp. 51–58, 2241 Jan. 2011.

[24] N. K., O. S., and K. Blank, "Bicycle and pedestrian count programs: Summary of practice and key resources," *Pedestrian and Bicycle Information Center*, Oct. 2018.

[25] *Traffic Monitoring Guide*. Federal Highway Administration, 2022, pp. 1–288.

[26] S. Kothuri *et al.*, "Bicycle and pedestrian counts at signalized intersections using existing infrastructure opportunities and challenges," *Transportation Research Record: Journal of the Transportation Research*, vol. 2644, pp. 11–18, 2017.

[27] K. Nordback, "Guide to bicycle & pedestrian count programs," in *TREC at Portland State University*, 2014.

[28] P. Ryus *et al.*, "NCHRP report 797: Guidebook on pedestrian and bicycle volume data collection," American Association of State Highway and Transportation Officials, p. 139.

[29] *Traffic Monitoring Guide*. Federal Highway Administration, 2013, pp. 1–300.

[30] M. O'Toole, "The relationship between employees' perceptions of safety and organizational culture.," *Journal of Safety Research*, vol. 33, pp. 231–243, 2 Jun. 2002.

[31] T. Nelson *et al.*, "Generalized model for mapping bicycle ridership with crowdsourced data," *Transportation Research Part C: Emerging Technologies*, vol. 125, p. 102 981, Apr. 2021.

[32] B. Blanc, P. Johnson, M. Figliozzi, C. Monsere, and K. Nordback, "Leveraging signal infrastructure for nonmotorized counts in a statewide program pilot study," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2527, pp. 69–79, 2015.

[33] P. A. Singleton, F. Runa, and P. Humagain, "Utilizing archived traffic signal performance measures for pedestrian planning and analysis," Utah Department of Transportation, 2020, pp. 1–87.

[34] P. A. Singleton and F. Runa, "Pedestrian traffic signal data accurately estimates pedestrian crossing volumes," *https://doi.org/10.1177/0361198121994126*, vol. 2675, pp. 429–440, 6 Mar. 2021.

[35] P. A. Singleton, K. Park, and D. H. Lee, "Utilizing ATSPM data for pedestrian planning and analysis – phase II: Extending pedestrian volume estimation capabilities to unsignalized intersections," Utah Department of Transportation, 2021, pp. 1–64.

[36] P. Singleton and F. Runa, "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*, vol. 93, pp. 1–11, 2021.

[37] P. A. Singleton, M. Mekker, and A. Islam, "Safety in numbers? developing improved safety predictive methods for pedestrian crashes at signalized intersections in utah using push button-based measures of exposure," Utah Department of Transportation, 2021, pp. 1–88.

[38] P. Humagain and P. Singleton, "Advances in pedestrian travel monitoring: Temporal patterns and spatial characteristics using pedestrian push-button data from utah traffic signals," *Journal of Transportation and Land Use*, vol. 14, pp. 1341–1360, 1 2021.

[39] F. R. Proulx and A. Pozdnukhov, "Bicycle traffic volume estimation using geographically weighted data fusion," *Journal of Transportation Geography*, 2017.

[40] M. M. Miah *et al.*, "Challenges and opportunities of emerging data sources to estimate network-wide bike counts," *Journal of Transportation Engineering Part A: Systems*, vol. 148, pp. 1–21, 3 2021.

[41] H. H. Hochmair, E. Bardin, and A. Ahmouda, "Estimating bicycle trip volume for miami-dade county from strava tracking data," *Journal of Transport Geography*, vol. 75, pp. 58–69, Feb. 2019.

[42] B. Dadashova, G. P. Griffin, S. Das, S. Turner, and B. Sherman, "Estimation of average annual daily bicycle counts using crowdsourced strava data," *Transportation Research Record: Journal of the Transportation Research Board*, p. 036 119 812 094 601, Sep. 2020.

[43] D. Bhowmick *et al.*, "A systematic scoping review of methods for estimating link-level bicycling volumes," *Transport Reviews*, vol. 43, pp. 622–651, 4 Jul. 2023.

[44] *Strava Metro Home*, https://metro.strava.com/metro/home. Accessed: 2021-08.

[45] *Strava: About Us*, https://www.strava.com/about. Accessed: 2021-08.

[46] M. D. Garber, K. E. Watkins, and M. R. Kramer, "Comparing bicyclists who use smartphone apps to record rides with those who do not: Implications for representativeness and selection bias," *Journal of Transport and Health*, vol. 15, p. 100 661, Dec. 2019.

[47] K. Watkins, R. Ammanamanchi, J. LaMondia, and C. A. L. Dantec, "Comparison of smartphone-based cyclist GPS data sources," *Transportation Research Board Annual Meeting*, 2016.

[48] R. L. Sanders, A. Frackelton, S. Gardner, R. Schneider, and M. Hintze, "Ballpark method for estimating pedestrian and bicyclist exposure in Seattle, Washington," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2605, pp. 32–44, 1 Jan. 2017.

[49] S. Aghabozorgi, A. S. Shirkhorshidi, and T. Y. Wah, "Time-series clustering – a decade review," *Information Systems*, vol. 53, pp. 16–38, Oct. 2015.

[50] J. Strauss, L. F. Miranda-Moreno, and P. Morency, "Mapping cyclist activity and injury risk in a network combining smartphone GPS data and bicycle counts," *Accident Analysis Prevention*, vol. 83, pp. 132–142, 2015.

[51] D. Li, Y. Zhao, and Y. Li, "Time-series representation and clustering approaches for sharing bike usage mining," *IEEE Access*, vol. 7, pp. 177 856–177 863, 2019.

[52] R. Ewing and R. Cervero, "Travel and the built environment," *Journal of the American Planning Association*, vol. 76, no. 3, pp. 265–294, 2010.

[53] A. Sardá-Espinosa, *Comparing time-series clustering algorithms in R using the dtwclust package*, 2017.

[54] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, pp. 43–49, 1 Feb. 1978.

[55] J. Han, M. Kamber, and J. Pei, "Cluster analysis: Basic concepts and methods," in 4th ed. Morgan Kaufmann., 2022, pp. 379–429.

[56] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "STL: A seasonal-trend decomposition procedure based on loess," *Journal of Official Statistics*, vol. 6, pp. 3–73, 1 1990.

[57] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, Nov. 1987.

[58] R. V. de Schoot, D. Kaplan, J. Denissen, J. B. Asendorpf, F. J. Neyer, and M. A. van Aken, "A gentle introduction to bayesian analysis: Applications to developmental research," *Child Development*, vol. 85, pp. 842–860, 3 May 2014.

[59] R. V. de Schoot *et al.*, "Bayesian statistics and modelling," *Nature Reviews Methods Primers*, vol. 1, 1 Dec. 2021.

[60] R. McElreath, *Statistical Rethinking: A Bayesian Course with Examples in R and STAN*, 2nd ed. Chapman & Hall, Mar. 2020, pp. 1–612.

[61] J. Shorey, "Characterization of pedestrian activity in georgia via clutering of passivley collected traffic signal data," Master's Thesis at the Georgia Institute of Technology, Jul. 2023, pp. 1–215.

[62] A. Gelman, J. Hwang, and A. Vehtari, "Understanding predictive information criteria for Bayesian models," 2014.

[63] R. Elvik and R. Goel, "Safety-in-numbers: An updated meta-analysis of estimates," *Accident Analysis Prevention*, vol. 129, pp. 136–147, Aug. 2019.

[64] R. T. Panik, ""Safety in Numbers" and bicycle safety: A detailed analysis of the Denver metropolitan area," *Clemson University Tiger Prints*, pp. 1–98, May 2018.

[65] J. Schepers, P. Kroeze, W. Sweers, and J. Wüst, "Road factors and bicycle–motor vehicle crashes at unsignalized priority intersections," *Accident Analysis Prevention*, vol. 43, pp. 853–861, 3 May 2011.

[66] C. Monsere, H. Wang, Y. Wang, and C. Chen, *Risk factors for pedestrian and bicycle crashes*, May 2017.

[67] G. Prati, V. M. Puchades, M. D. Angelis, F. Fraboni, and L. Pietrantoni, "Factors contributing to bicycle–motorised vehicle collisions: A systematic literature review," *Transport Reviews*, vol. 38, pp. 184–208, 2 Mar. 2018.

[68] R. Utriainen, S. O'Hern, and M. Pöllänen, "Review on single-bicycle crashes in the recent scientific literature," *Transport Reviews*, vol. 43, pp. 159–177, 2 Mar. 2023.

[69] E. Dumbaugh and W. Li, "Designing for the safety of pedestrians, cyclists, and motorists in urban environments," *Journal of the American Planning Association*, vol. 77, pp. 69–88, 1 Dec. 2010.

[70] R. Aldred, "Pedestrian injury risk: Unanswered questions and a developing research agenda," *Transport Reviews*, vol. 38, pp. 685–688, 6 Nov. 2018.

[71] B. Dai and B. Dadashova, "Review of contextual elements affecting bicyclist safety," *Journal of Transport & Health*, vol. 20, p. 101 013, 2021.

[72] P. Stoker *et al.*, "Pedestrian safety and the built environment," *Journal of Planning Literature*, vol. 30, pp. 377–392, 4 Nov. 2015.

[73] J. Ambros, R. Turek, E. Šragová, K. Petr, M. Šucha, and J. Frič, "How fast would you (or should you) drive here? investigation of relationships between official speed limit, perceived speed limit, and preferred speed," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 83, pp. 164–178, Nov. 2021.

[74] L. Thomas, K. Nordback, and R. Sanders, "Bicyclist crash types on national, state, and local levels: A new look," *Transportation Research Record*, vol. 2673, pp. 664–676, 6 2019.

[75] J. Engström, E. Johansson, and J. Östlund, "Effects of visual and cognitive load in real and simulated motorway driving," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 8, pp. 97–120, 2 SPEC. ISS. Mar. 2005.

[76] D. Arias, D. Ederer, M. O. Rodgers, M. P. Hunter, and K. E. Watkins, "Estimating the effect of vehicle speeds on bicycle and pedestrian safety on the georgia arterial roadway network," *Accident Analysis Prevention*, vol. 161, p. 106 351, 2021.

[77] U. S. D. of Transportation, U. Transportation, and F. H. Administration, *Manual on Uniform Traffic Control Devices for Streets and Highways (MUTCD) 11th Edition, December 2023*

(Complete Book, Hardcover, Color Print) National Standards for Traffic Control Devices. Independently Published, 2023, ISBN: 9781998295517.

[78] Global Designing Cities Initiative, *Design Speed - Global Street Design Guide*, Accessed: 2024-11-01, 2024.

[79] M. S. Myhrmann, K. H. Janstrup, M. Møller, and S. E. Mabit, "Factors influencing the injury severity of single-bicycle crashes," *Accident Analysis & Prevention*, vol. 149, p. 105 875, Jan. 2021.

[80] C. J. Hamann, C. Peek-Asa, C. F. Lynch, M. Ramirez, and P. Hanley, "Epidemiology and spatial examination of bicycle-motor vehicle crashes in iowa, 2001–2011," *Journal of Transport Health*, vol. 2, pp. 178–188, 2 Jun. 2015.

[81] X. Liu, L. Shen, and J. Huang, "Analysis of bicycle accidents and recommended countermeasures in beijing, china," *Transportation research record*, vol. 1487, pp. 75–83, 1995.

[82] M. M. Hoque, "An analysis of fatal bicycle accidents in victoria (australia) with a special reference to nighttime accidents," *Accident Analysis Prevention*, vol. 22, pp. 1–11, 1 Feb. 1990.

[83] J. DiGioia, K. E. Watkins, Y. Xu, M. Rodgers, and R. Guensler, "Safety impacts of bicycle infrastructure: A critical review," *Journal of Safety Research*, vol. 61, pp. 105–119, 2017.

[84] C. A. Mulvaney *et al.*, "Cycling infrastructure for reducing cycling injuries in cyclists," *Cochrane Database of Systematic Reviews*, no. 12, 2015.

[85] J. Daniel and K. Brodtman, "Shoulder and rumble strips and bicyclsits," New Jersey Department of Transportation, Jun. 2007, Received from: https://www.discountpdh.com/wp-content/themes/discountpdh/pdf-course/designing-shoulder-rumbles-strips-compatible-with-bicyclists-and-beneficial-to-motorists.pdf, pp. 1–77.

[86] S. Kaplan and C. G. Prato, "A spatial analysis of land use and network effects on frequency and severity of cyclist–motorist crashes in the copenhagen region," *Traffic Injury Prevention*, vol. 16, pp. 724–731, 7 Oct. 2015.

[87] R. L. Wessels, "Bicycle collisions in washington state: A six-year perspective, 1988–1993," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1538, pp. 81–90, 1 Jan. 1996.

[88] R. A. Chaney and C. Kim, "Characterizing bicycle collisions by neighborhood in a large midwestern city," *Health Promotion Practice*, vol. 15, pp. 232–242, 2 Mar. 2014.

[89] L. F. Miranda-Moreno, J. Strauss, and P. Morency, "Disaggregate exposure measures and injury frequency models of cyclist safety at signalized intersections," *Transportation Re-*

search Record: Journal of the Transportation Research Board, vol. 2236, pp. 74–82, 1 Jan. 2011.

[90] R. O. Phillips, O. H. Hagen, and S. H. Berge, "Bus stop design and traffic safety: An explorative analysis," *Accident Analysis Prevention*, vol. 153, p. 105 917, 2021.

[91] M. B. Ulak, A. Kocatepe, A. Yazici, E. E. Ozguven, and A. Kumar, "A stop safety index to address pedestrian safety around bus stops," *Safety Science*, vol. 133, p. 105017, 2021.

[92] J. Lee, M. Abdel-Aty, and X. Jiang, "Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level," *Accident Analysis Prevention*, vol. 78, pp. 146–154, May 2015.

[93] A. Gadsby and K. Watkins, "Instrumented bikes and their use in studies on transportation behaviour, safety, and maintenance," *Transport Reviews*, vol. 40, no. 6, pp. 774–795, 2020.

[94] K. Kim and L. Li, "Modeling fault among bicyclists and drivers involved in collisions in hawaii, 1986–1991," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1538, pp. 75–80, 1 Jan. 1996.

[95] *Minnesota Department of Transportation: Bicycle and Pedestrian Traffic Counting*, https://www.dot.state.mn.us/bike-ped- counting/index.html. Accessed:2024-11.

[96] *Colorado Department of Transportation: Bicycle and Pedestrian Counts, Program Overview*, https://www.codot.gov/programs/bikeped/bicycle-pedestrian-counts. Accessed: 2024-11.