Estimating Intersection Annual Average Daily Bicycle Traffic from 8-hour Turning Movement Counts

by

Benjamin Allen

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Civil Engineering

Waterloo, Ontario, Canada, 2021

Authors Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

This thesis outlines a set of procedures for estimating annual average daily bicycle traffic (AADB) from one day, 8-hour turning movement counts (TMCs). Factoring methods for annualizing short duration counts have been long established for motor vehicle traffic, and more recent research has adapted many of these methods to pedestrians and bicyclists. However, much of this research has been concerned with estimating AADB on dedicated cycling facilities. Less has been done to transfer these methods to estimating cyclist activity at intersections, even though it would be valuable to measure cyclist exposure for network safety analysis and for broader planning purposes. TMCs represent a valuable potential source of data for this purpose, as it is common for North American jurisdictions to regularly collect them as part of ongoing traffic monitoring programs. Sets of video monitoring unit (VMU) data from Milton, Ontario, and Pima County, Arizona, were used to evaluate whether existing methods could be appropriately applied to 8-hour TMCs. Several updates to conventional estimation methods were proposed to account for the differences between TMCs and "conventional" cyclist counts. Additionally, methods are proposed for filtering VMU data; and for matching short-duration count locations to empirical factor groups using their land-use and physical characteristics. The resulting set of procedures could be implemented by transportation agencies using data which they may already be collecting to generate estimates of cyclist activity at any intersection in their jurisdiction, although further work is likely needed to improve estimation accuracy, especially at low-volume locations.

Acknowledgements

I would like to sincerely thank my academic supervisor, Dr. Bruce Hellinga, P.Eng., for the effort he put into every step of this process. I could not have finished this thesis without his direction and insight, and I'm extremely grateful to have been given the chance to work under his guidance. I owe whatever comes next for me, academically or professionally, to him.

I would also like to thank the other students in Dr. Hellinga's research group, Andrew Orr, Lucas Sobreira, Alan Xaykongsa, Trevor Vanderwoerd, and Mohammad Zarei, who were not only excellent academic colleagues, but also good friends throughout these two years.

Thank you to the entire University of Waterloo Faculty of Environment and School of Planning for giving me the educational background I needed, and specific thanks to Prof. Clarence Woudsma and Prof. Dawn Parker for making it possible for me to pursue this goal in the first place. Thank you also to Dr. Jeff Casello for helping to put me on this path, and to Pedram Izadpanah for his help with this research.

Finally, I would like to say thank you to the support and encouragement provided by my friends and family, especially to my loving mother and father for their unconditional support; and to my partner Kiana who was with me from start to finish.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Background and Need

Cycling as a legitimate mode of commuting has grown substantially in North America over the past 20 years, with the number of cycling commuters in Canada alone growing by 87.9% between 1996 and 2016 [1]. This growth has been accompanied by (or possibly driven by) an increase in policy promoting sustainable transportation and the development of larger and more comprehensive cycling infrastructure networks, with governments at all levels acknowledging the potential health, environmental, social, and economic benefits that cycling provides. With this increase in interest, it is necessary for transportation agencies to adapt their policies and procedures to account for the growing demand. It also creates the need for more comprehensive and detailed cyclist activity data to facilitate an informed, multi-modal transportation planning process.

Although the volume of research relating to cycling activity has increased as well [2], the body of literature is still lacking somewhat compared to that for motor vehicle traffic. Still, there has been a significant number of studies published since 2010 relating to the estimation of non-motorized traffic volumes. Of particular emphasis has been methods for calculating expansion factors, which can be calculated from complete datasets (i.e. data collected by permanent, continuous traffic monitoring equipment) and then applied to short-duration counts (typically manual counts less than 24-hours in length) to produce estimates of annual average daily bicycle traffic (AADB). Although procedures for the development and application of expansion factors for motorized traffic have been long established, equivalent procedures for cyclists are more recent.

Importantly, the scope of research on this subject has been limited to estimating AADB on permanent, dedicated, usually high volume and high priority cycling facilities. While this is very useful, it is also somewhat constraining. It is well established that the presence of dedicated facilities increases cycling activity [3] [4], and although it is more difficult to show that variation in relative temporal patterns can be explained by facility type, it seems reasonable to assume that cyclist activity patterns at a given facility will be most similar to activity patterns at facilities of the same type This would mean that factors calculated using continuous data collected from dedicated cycling facilities would be most appropriately applied to estimate AADB at other dedicated cycling facilities, so that only a very small portion of the overall transportation network in most North American cities could be covered by such an analysis. It would be more valuable, then, to be able to estimate AADB at intersections across the whole network, so that some measure of cyclist exposure is available to use for safety analysis and for broader planning purposes.

Until recently, the key limitation towards this goal had been data availability, with most count data coming from dedicated cycling infrastructure. However, video monitoring units (VMUs) which provide continuous, movement specific traffic counts at intersections are becoming more common. With these data, it should be possible to calculate sets of expansion factors which can be used to estimate AADB from one-day TMCs at any intersection in a network.

1.2 Purpose

This thesis outlines a comprehensive procedure for the development and application of expansion factors to estimate annual average daily bicycle traffic (AADB) statistics from one day, 8-hour turning movement counts (TMCs). Cyclist data from several North American jurisdictions are used to develop, refine, and test methods for quality checking continuous cyclist count data; establish factor groups in sets of count locations; calculate standardized temporal expansion factors; and those factors used to generate estimates of AADB from 8-hour turning movement counts (TMCs) which local jurisdictions are already collecting. These methods could be implemented by transportation agencies as part of ongoing traffic monitoring programs, giving them the ability to estimate AADB at any intersection for which a TMC is available, creating a valuable source of data for numerous transportation planning processes.

1.3 Research Scope and Organization

This thesis is divided into five additional chapters. Each of these chapters has a specific purpose and scope, all contributing to the broader goal of developing a comprehensive procedure for the estimation of AADB from one-day TMCs. A review of practice is split up throughout the thesis, so that the literature most relevant to the work done in each chapter is discussed alongside it.

1.3.1 Chapter 2: Filtering Continuous TMC Data

This chapter is concerned with adapting already established quality control measures for conventional cyclist count data to video monitoring unit (VMU) data. The specific research objectives of this chapter are to:

- Summarize and characterize the datasets used in this research.
- Understand the cyclist count patterns which may indicate a counting error in a conventional count dataset.
- Develop a set of baseline filters for identifying suspicious observations in a set of conventional counts.
- Understand the differences between conventional cyclist count data and cyclist counts collected using a VMU; adjust the baseline filtering logic accordingly.
- Establish rules for simulating the collection of TMCs in real-world practice.

Several different datasets are used in this chapter: a set of conventional cyclist count data from Hamilton, Ontario is used to establish the baseline filters; a VMU dataset from Milton, Ontario is used to adjust the filters; and the adjusted filters are applied to a VMU dataset from Pima County, Arizona, which is the primary dataset used in subsequent chapters.

1.3.2 Chapter 3: Baseline AADB Expansion Methods

Chapter 3 includes a detailed review of the existing literature pertaining to expansion factors for AADB estimation. Several "baseline" methods are identified and applied to the Milton and Pima County VMU datasets. The research objectives for this chapter are to:

- Summarize in detail the state of practice pertaining to expansion factors for AADB estimation.
- Identify and understand a set of common "baseline" methods, discussing the advantages and limitations of each.
- Develop a procedure for testing estimation methods, using a VMU dataset to simulate the collection of one-day TMCs.
- Compare results from the baseline methods in terms of estimation accuracy.

1.3.3 Chapter 4: Adapting the Baseline Expansion Methods for TMCs

Having identified and applied "baseline" expansion methods in Chapter 3, Chapter 4 improves upon those methods by adapting them to better suit TMCs. The objectives for this chapter are to:

- Identify the limitations of using TMCs for AADB estimation, with reference to the results produced in Chapter 3.
- Identify adjustments that could be made to the baseline methods which would address these limitations.
- Propose an updated set of expansion factors and test them on the Pima County dataset; compare these results to the baseline results in terms of estimation accuracy.

1.3.4 Chapter 5: Modelling Factor Groups Using Locational Characteristics

The second to last chapter in this thesis addresses the issue of identifying the correct factor group for a short-term count location (STCL) for which less than 24-hours of count data is available. The research objectives for this chapter are to:

- Collect a set of land-use and roadway-characteristic data for count locations in the Pima County dataset, referencing literature on variables which influence cyclist activity patterns.
- Develop a multivariate linear model relating expansion factors to land use variables.
- Develop a logistic regression model relating nominal factor groups to land use variables.
- Develop an iterative procedure for evaluating these models by predicting the accuracy of factor-group prediction; evaluate the impact of using predicted groups on estimation accuracy.
- Propose an alternative approach to determining STCL factor groups by manually assessing the land use characteristics of each Pima County intersection and relating them to cyclist activity patterns.

Chapter 2
Filtering Continuous Cyclist Turning Movement Counts

Several different sets of data – both "conventional" cyclist-only counts and automated intersection traffic monitoring data – were used to study AADB estimation methods. Data sources are introduced, datasets summarized, and data preparation discussed in this chapter, along with a detailed discussion of the development and adaptation of filtering rules for turning movement count (TMC) data.

2.1 Data Sources

This section will provide an overview of the data used in this study. Several different datasets were used for different purposes. These will be referred to as either the Hamilton, the Milton, or the Pima County datasets, with more information on the source and composition of each found below.

2.1.1 Hamilton, ON Continuous Cyclist Count Data

Conventional cyclist count data was obtained from the City of Hamilton, Ontario, Canada. It was collected by permanent counting equipment installed at six different locations around Hamilton, all on permanent cycling infrastructure. The distinction between "conventional" data and the TMC data which comprises the other two datasets will be expanded on in the rest of this chapter. Refer to Table 1 below for a summary of the Hamilton dataset by location.

Table 1: Summary of Hamilton count locations

ID	Location	AADB	Start Date	End Date	Count Days	Infrastructure Type
1	Bay@Cannon	142	2017-11-21	2019-09-30	678	Bidirectional cycle track
2	Bay@Hunter	177	2017-11-27	2019-09-30	672	Bidirectional cycle track
3	Bay@Stuart	86	2017-12-04	2019-09-30	665	Painted bike lane
4	Cannon@Bay	113	2015-12-11	2019-09-30	1389	Bidirectional cycle track
5	Cannon@West	390	2016-01-19	2019-02-28	1136	Bidirectional cycle track
6	King	270	2016-07-19	2019-09-30	1168	Bidirectional cycle track

2.1.2 Milton, ON Intersection Traffic Monitoring Data

Turning movement count (TMC) data represents the volumes of traffic passing through an intersection over a given time period by approach movement (left, right, or through), often disaggregated by traffic type (light and heavy vehicles, articulated trucks, buses, pedestrians, bicycles, etc.). TMCs are usually collected manually during peak-hours for a single weekday; more information on TMC collection can be found in Section 2.7. However, with the ongoing improvements to video collection and processing technology, it is becoming more common to install video monitoring units (VMUs) at high volume intersections [5]. These collect video which is then processed through image recognition software to produce highly accurate, continuous TMC data [6].

VMU data was collected at eight intersections in Milton, Ontario, Canada. Refer to Table 2 below for a summary of the Milton dataset by location, and to Figure 1 for a map of locations. Raw data contained a timestamped record for every vehicle, bicycle, or pedestrian passing through the intersection; this was aggregated to 15-minute and 24-hour intervals for analysis.

Table 2: Summary of Milton count locations

ID	Name	Start Date	End Date	Count Days	Avg. Daily Motorized	Avg. Daily Bicycles
4	Thompson Road and Childs Drive	2018-06-05	2020-04-16	681	30,769	35.9
5	Thompson Road and Laurier Avenue	2018-06-06	2020-04-16	680	29,912	34.1
9	Ontario Street South and Pine Street	2018-10-12	2020-04-16	552	24,689	12.8
14	Bronte Street North and Main Street West	2018-11-06	2020-04-16	527	23,161	18.8
15	Ontario Street South and Childs Drive	2018-10-12	2020-04-16	552	25,888	10.4
17	Main Street East and Mall Entrance	2018-10-19	2020-04-16	545	21,518	11.3
22	Thompson Road and McCuaig Drive	2018-05-25	2020-04-16	692	30,640	11.7
25	Main Street East and Ontario Street North	2018-10-12	2020-04-16	552	32,259	18.5

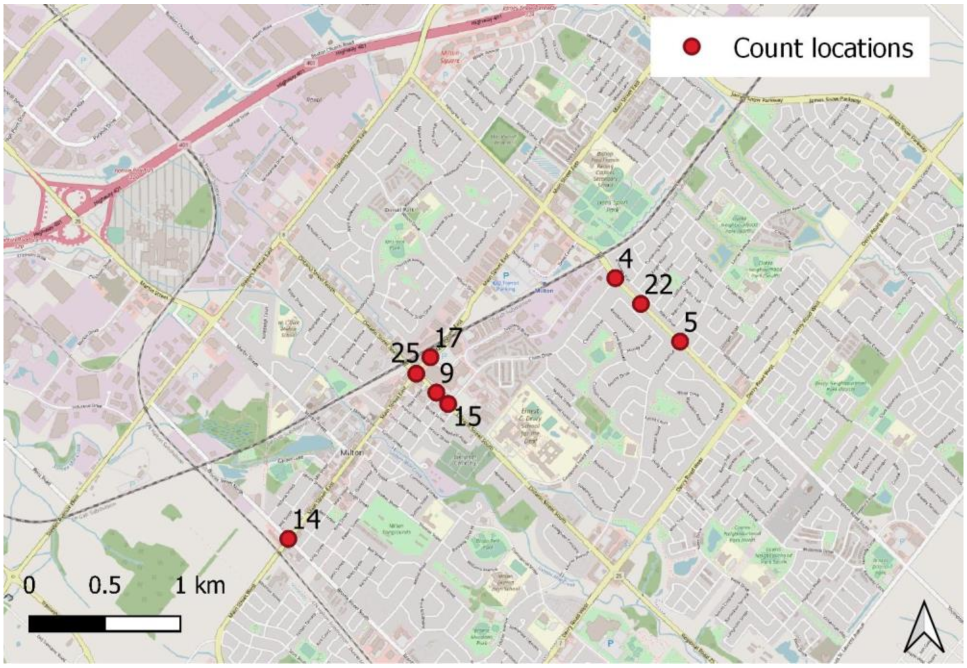


Figure 1: Milton count locations (Map data copyrighted OpenStreetMap contributors [7])

2.1.3 Pima County, AZ Intersection Traffic Monitoring Data

VMU data was also acquired from a number of intersections in Pima County, Arizona, in the same format as the Milton dataset. Refer to Table 3 below for a summary of the Pima County dataset by location, and to Figure 2 for a map of locations.

Table 3: Summary of Pima County count locations

ID	Name	Start Date	End Date	Count Days	Avg. Daily Motorized	Avg. Daily Bicycles
5	Drexel Rd / Palo Verde Rd	2020-01-02	2020-09-10	252	15,877	13.0
6	Craycroft Rd / Sunrise Dr	2019-11-20	2020-09-10	295	25,878	22.3
7	Overton Rd / Shannon Rd	2020-01-05	2020-09-10	249	12,681	8.2
9	Benson Hy / Swan Rd / Valencia Rd	2020-01-02	2020-09-10	252	28,460	7.1
13	La Cholla Bl / Orange Grove Rd	2019-11-18	2020-09-10	297	35,901	8.2
34	La Cholla Bl / Ruthrauff Rd	2019-05-07	2020-09-10	492	30,504	18.8
36	Camino de la Tierra / Ina Rd	2019-03-31	2020-09-10	529	31,221	15.7
38	Sunrise Dr / Swan Rd	2019-12-05	2020-09-10	280	42,778	5.8
39	Cortaro Farms Rd / Magee Rd / Shannon Rd	2020-01-05	2020-09-10	249	21,544	11.4
41	Hardy Rd / La Canada Dr / Overton Rd	2020-01-05	2020-09-08	247	27,038	37.2
42	Romero Rd / Ruthrauff Rd	2019-09-29	2020-09-10	347	20,407	21.7
43	Calle del Marques / Sunrise Dr	2019-12-05	2020-09-10	280	30,837	20.1
49	Sabino Canyon Rd / Sunrise Dr	2019-10-03	2020-09-10	343	5,527	83.3
50	Magee Rd / Shannon Rd / Tuscany Dr	2020-01-05	2020-09-10	249	21,743	8.4
52	1st Av / Christie Dr / Ina Rd	2019-05-12	2020-09-10	487	21,868	9.0
53	La Cholla Bl / Magee Rd	2020-01-05	2020-09-10	249	26,119	9.0
57	Orange Grove Rd / Shannon Rd	2019-11-18	2020-09-10	297	25,584	4.9
59	Hardy Rd / Thornydale Rd	2020-01-05	2020-09-10	249	17,002	22.1
62	Cortaro Farms Rd / Thornydale Rd	2020-01-05	2020-09-10	249	26,041	4.7
69	Orange Grove Rd / Skyline Dr	2019-11-20	2020-09-10	295	32,189	11.2
70	Linda Vista Bl / Thornydale Rd	2018-10-10	2020-09-10	701	19,185	9.7
71	Kolb Rd / Sunrise Dr	2019-10-09	2020-09-10	337	15,407	29.2
72	La Canada Dr / River Rd	2019-12-10	2020-09-10	275	42,050	13.9
73	Ina Rd / La Canada Dr	2019-06-10	2020-09-10	458	41,169	29.1
76	Campo Abierto / Sunrise Dr	2019-12-05	2020-09-10	280	28,911	17.4
77	Camino De La Tierra / Orange Grove Rd	2019-11-17	2020-09-10	298	28,103	10.3
80	Alvernon Wy / Valencia Rd	2020-01-02	2020-09-10	252	30,046	5.2
81	Overton Rd / Thornydale Rd	2018-10-09	2020-09-10	702	18,123	13.9
85	37th St / Golf Links Rd / Palo Verde Rd	2020-01-02	2020-09-10	252	27,309	11.8
96	Ina Rd / Mona Lisa Rd	2019-05-07	2020-09-10	492	30,088	8.1
97	Colossal Cave Rd / Mary Ann Cleveland Wy	2019-11-12	2020-09-10	303	15,078	30.1
100	Pontatoc Rd / Sunrise Dr	2019-12-05	2020-09-10	280	31,362	26.9
101	Ina Rd / Westward Look Dr	2019-11-19	2020-09-10	296	31,733	29.6
105	La Cholla Bl / Overton Rd	2020-01-05	2020-09-10	249	10,182	9.6

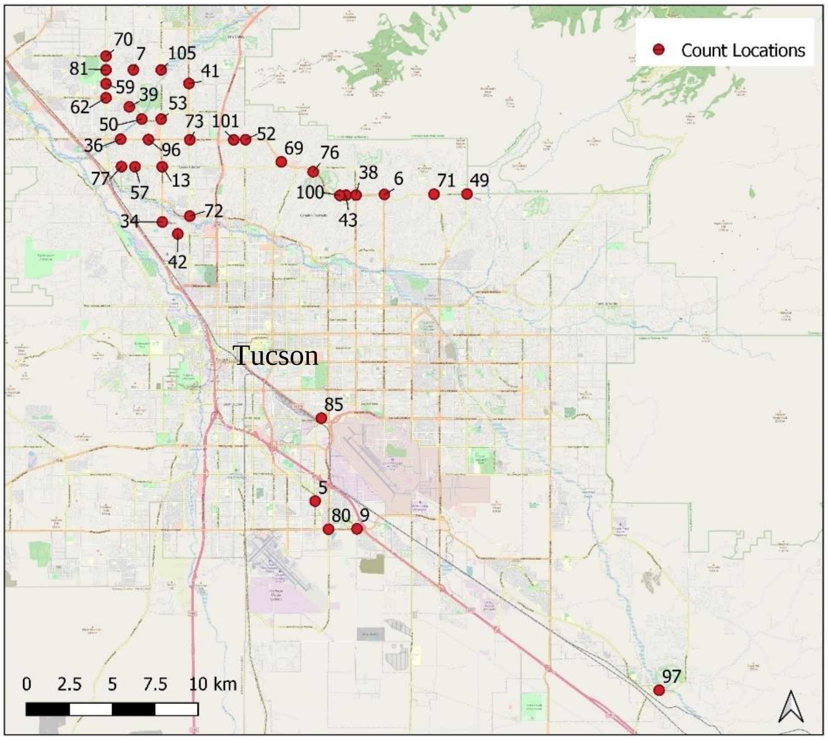


Figure 2: Pima County count locations (Map data copyrighted OpenStreetMap contributors [7])

Although data at several Pima County intersections was available into 2019 and 2018, for consistency it was decided to only use data starting from the last date on which data was available for most locations (this was January 5, 2020, but data was included up to January 1, 2020 where possible). Furthermore, it was assumed that the emergence of the Covid-19 pandemic early in 2020 would substantially impact transportation activity patterns. It was estimated that restrictions relating to the pandemic started to come into effect around March 15, 2020 in Pima County. Figure 3 below very clearly shows motor vehicle traffic volumes across all 34 count locations falls steeply at about this time, while at the same time average cyclist volumes climbed rapidly. For this reason, data for the remainder of 2020 after March 15 was considered unrepresentative and was excluded from the Pima County.

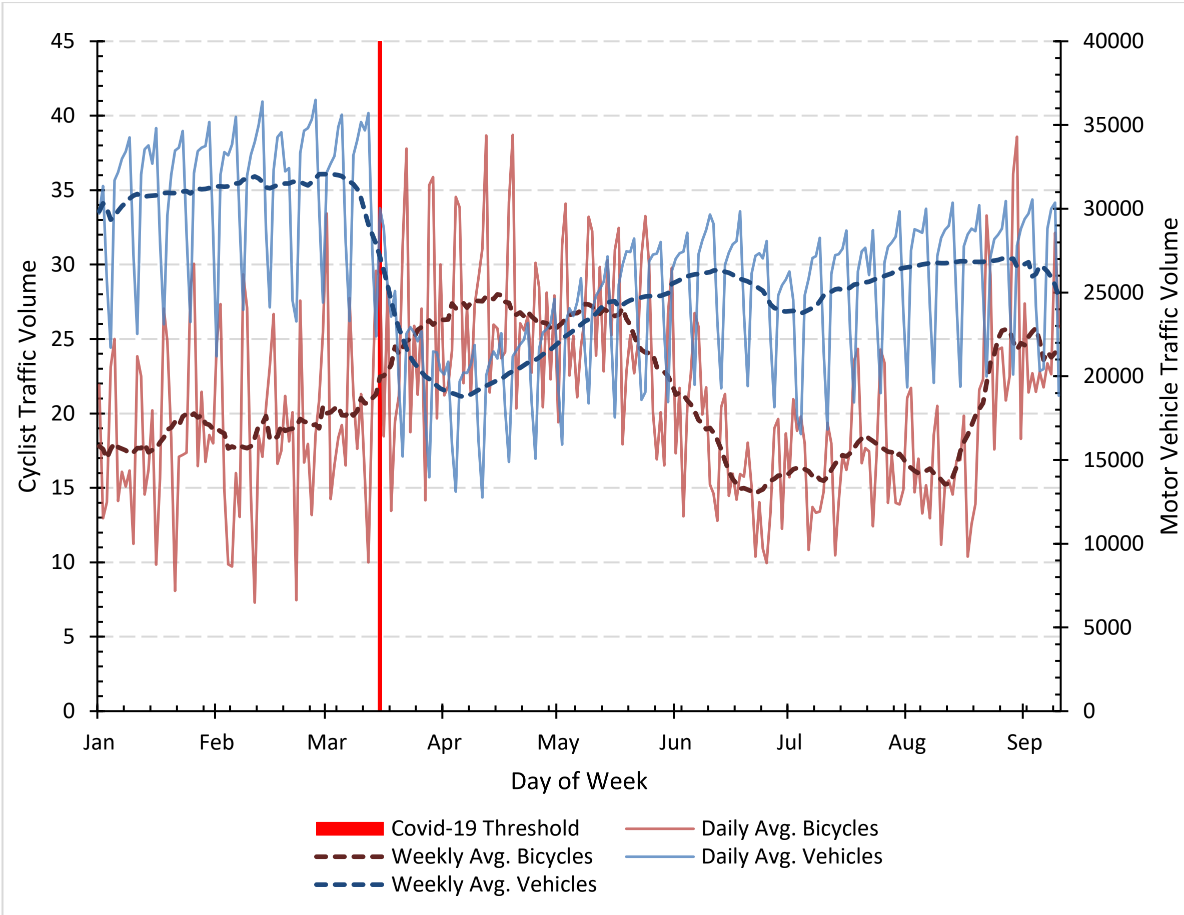


Figure 3: Average daily bicycle and motor vehicle traffic volumes across all 34 Pima County count locations, 01/01/2020 - 10/09/2020

2.2 Filtering Conventional Cyclist Counts

Much of the existing work related to cyclist demand estimation is built on what can be labelled as "conventional" cyclist counts. In general, these counts are bi-directional, often mid-block, and collected continuously by automated infra-red or inductive loop counters installed on high-volume dedicated cycling infrastructure. As a consequence, most common practice for quality checks was developed for conventional counts. The Hamilton dataset was used to develop a cyclist count filtering framework, based on a review of relevant literature on quality checks for active transportation data.

2.2.1 Review of Relevant Literature and Practice

The primary purpose of filtering count data is to identify erroneous or inaccurate data which may adversely affect an analysis. “Erroneous” data in this application specifically refers to inaccurate data measurements caused by counting equipment malfunctions or other external factors. Some possible scenarios which could lead to inaccurate measurements from a conventional bike counter include:

- Counting equipment malfunctions, damage occurring to the equipment, or a loss of power which cause the counter to stop recording or record inconsistently
- Environmental factors, such as snow buildup blocking an infrared sensor or camera causing abnormally low counts; or an impediment on the sidewalk causing pedestrians to walk through a bike lane, resulting in abnormally high counts [8]

Many of these issues are specific to the technology deployed in a location, and furthermore many of the filtering techniques applied across the literature are done so in an ad-hoc fashion to best meet the requirements of any given study or data source. However, several count patterns were identified both from the literature and by inspecting the Hamilton data which probably indicate some form of counting error:

1. “Null”, or missing counts
2. Consecutive zero counts
3. Consecutive, identical non-zero counts
4. Automated count interval outliers (i.e. 15-minute count outliers)
5. 24-hour zero counts
6. 24-hour count outliers

Each of these six count patterns can be identified through a separate sub-process falling under the larger heading of “data filtering”. The goal of these sub-processes is to identify potentially inaccurate or unrepresentative counts for future investigation [9], the removal of which can improve the overall quality of the database. Data are not removed outright from incoming datasets, only flagged if they meet one of the filtering conditions. A researcher can then examine flagged counts in more detail to determine if they should be excluded. A scan of quality control techniques was conducted across several sources and summarized in Table 4.

Table 4: Summary of literature scan for quality control thresholds

Source	"Missing" Counts	Consecutive Zeros	Consecutive Non-zero's	Outliers (interval)	Outliers (24h Count)	24h Zero Counts
Colorado DOT [10]	Targeted visual inspection of temporal plots	Flag series of 8 or longer 1h counts	Flag series of 4 or longer 1h counts	Flag 1h counts outside 2.5(IQR) + Q3 from weekly average for that hour	-	-
Miranda-Moreno et al [11]	Visual inspection of temporal plots	-	-	Visual inspection of temporal plots for "extreme" values	-	-
McNeil et al [12]	"Null" values removed outright	Flag series longer than 99 15-min counts	Flag series of 9 15-min counts, at the longest	Flag 15-min counts >=250 for expected volume <100	-	-
Medury et al [13]	-	Flag periods longer than 24-hours	-	-	"Median Absolute Deviation" method	-
Minnesota DOT [14]	Visual inspection of temporal plots	Undefined "long" runs	-	-	Flag counts outside of 2 Stdv. from mean	Remove daily zeros outright
Nordback et al [9]	-	Flag series of >15 1h counts	Flag series of >6 consecutive records	Flag hours >1500	Flag days >10,000	Flag days with zero counts

Note:"-" in the table indicates that paper did not identify this as something for which filtering was applied

Although the literature provides some guidance on automated cyclist count filtering, there is a lack of consensus on specific thresholds for flagging erroneous data. For this reason, the Hamilton dataset was used to further develop filtering logic for identifying each of the six count patterns listed in the previous section as probably representing a counting error.

2.2.2 Filter 1: Missing Data

Filter 1 (F1) identifies missing data at the disaggregate count interval level (i.e. raw 15-minute counts) by flagging "null" observations. A "null" observation is different from a count of zero which – barring the presence of any other counting errors – would mean that a counter was operational, but that no cyclists passed the counter during a given interval. A "null" (or blank) record in the dataset, however, indicates that the equipment was not recording at all, whether cyclists were present or not (for instance, if the equipment ran out of battery or was damaged in some way). A "null" entry then almost certainly indicates a counting error and should always be flagged.

2.2.3 Filters 2 and 3: Consecutive Identical Counts

There are times when detectors may erroneously report - as what could be a result of an equipment malfunction or environmental interference - sequences of count intervals with identical, consecutive volumes. These volumes can be zero or any positive integer, which appear in a long series of count intervals. Although this issue is commonly encountered, there is no consensus as to what constitutes a "long" series, or what the threshold for the length of a sequence of consecutive identical counts should be for it to be considered suspicious. If we consider cyclist arrivals at a given point to be randomly distributed, then we should assume that sequences of matching counts will occur in a set of count data purely by chance. These 'valid' runs of counts need to be distinguished in some way from the erroneous ones. In other words, we want to establish a threshold *(N)* for the number of consecutive, identical count intervals *(n)* such that when $n \geq N$, those intervals can confidently be flagged as suspect. One strategy for doing so is to examine our count data as being Poisson distributed. Cyclist arrivals at a given detector are, or are close enough to appear to be, randomly distributed and discrete, and so can be examined as a Poisson process where the probability of X cyclists arriving during a time period T can be expressed by Equation (1).

$$P[X] = \frac{(\lambda T)^X e^{-\lambda T}}{X!} \tag{1}$$

Where:

λ	=	mean arrival rate (cyclists/hour)
T	=	time interval duration (minutes)
X	=	number of cyclists arriving during a time interval with duration T

If this is true, then we can also calculate the probability of observing a count of X_i for a 15-minute interval (0.25 hour) i, given an average arrival rate for that count interval (λ). For example, if $\lambda = 12$ cyclists/hour, then the probability of observing exactly 2 cyclists in any given interval i is:

$$P[X = 2] = \frac{(12 \times 0.25)^2 e^{-(12\times 0.25)}}{2!} = 0.224 \tag{2}$$

For this filter, however, we are not just interested in the probability of observing X_i once, but the probability of observing the same value of X in consecutive intervals, or put another way using the numerical example from above, the probability of observing count value $X_i = 2$ given that the value of X in n previous time intervals was also 2. This is expressed as $P[X_i = 2| X_{i-1} = 2| X_{i-2} = 2|\ldots| X_{i-n} = 2]$, and since arrivals are a Poisson process:

$$P[X_i = 2|X_{i-1} = 2| \ldots |X_{i-n} = 2] = P[X_i = 2] \times P[X_{i-1} = 2] \times \ldots P[X_{i-n} = 2] \qquad (3)$$

And the general case is:

$$P[X_i|X_{i-1}| \ldots |X_{i-n}] = P[X]^n = \left(\frac{(\lambda T)^X e^{-\lambda T}}{X!}\right)^n \qquad (4)$$

The implication of this is that it becomes more unlikely to observe the value X in n consecutive time intervals the larger n becomes, which supports the idea that very long runs of X could be arising from some counting error and should be labelled as suspect. Knowing this, we can now try to establish a threshold *(N)* for *n*, which will be a function of how confident we want to be in correctly flagging erroneous runs of consecutive counts.

To begin with we will set a confidence limit, β ($0 < \beta < 1$), representing our willingness to identify valid data as suspicious. A confidence limit of 0.95 for instance would indicate that we wish to identify runs of counts where there is a greater than 95% probability of *not* observing a count value X in n consecutive intervals, and that we are accepting that 5% of the time, correct count volumes will be mistakenly identified by the filter. If the probability of observing n consecutive, identical counts is given by $P[X]^n$ as calculated in Equation (4), then the probability of *not* observing that sequence can be defined as $\beta = 1 - P[X]^n$ and we can rewrite Equation (4) as

$$\beta = 1 - P[X_i|X_{i-1}| \ldots |X_{i-n}] = 1 - P[X]^n = 1 - \left(\frac{(\lambda T)^X e^{-\lambda T}}{X!}\right)^n \qquad (5)$$

Increasing the value of β raises the threshold N, reducing the number of counts which will be flagged, but also decreasing the likelihood of mistakenly declaring valid data as suspect. Instead of calculating an integer value for N, we can use our confidence interval directly as the actual threshold value by filtering interval counts where $(1 - P[X_i])$ is greater than β, or where the probability of observing the count X_i in the n^{th} consecutive interval is lower than we would be confident of observing in an accurate dataset.

Consider again the numerical example from above, $\lambda = 12$ cyclists/hour and $X_i = 2$, evaluated at a β value of 99.9%. The probability of observing the count 2 in n consecutive intervals is calculated by Equation (6) and the evaluation of $P[X_i]$ for n equals one through 6 is seen in Table 5, the third column of which shows that $1-P[2]^n$ surpasses the β threshold at $n = 5$; the probability of *not* observing a count value of X=2 five consecutive times is 99.94%.

$$P[2]^n = \left(\frac{(12 \times 0.25)^2 e^{-12\times 0.25}}{2!}\right)^n \tag{6}$$

Table 5: Evaluation of $P[2]^n$ for n = {1,2,3,4,5,6}

n	$P[2]^n$	**1 -** $P[2]^n$	**(1 -** $P[2]^n$**) ≥ β**
1	22.40%	77.60%	No
2	5.02%	94.98%	No
3	1.12%	98.88%	No
4	0.25%	99.75%	No
5	0.06%	99.94%	Yes
6	0.01%	99.99%	Yes

2.2.3.1 Calculating Arrival Rate (λ)

Consider again the variable N as the value of n at which $(1 - P[X_i])$ surpasses the β threshold. Figure 4 below shows how the value of the N threshold for consecutive values of X=2 responds to the hourly cyclist arrival rate λ, for β values of 98%, 99.8%, and 99.99%.

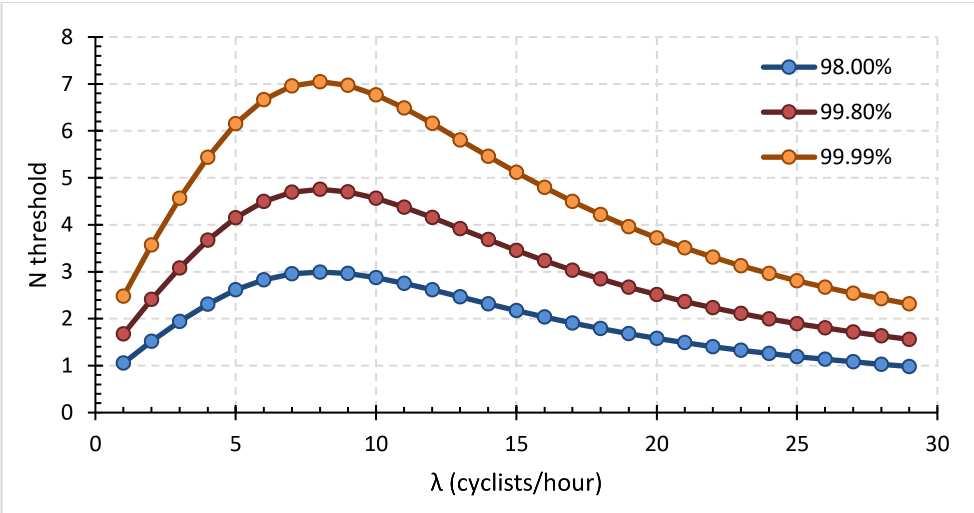


Figure 4: Value of N by hourly cyclist arrival rate (λ) for X = 2, at several levels of confidence

There is clearly some significant variation in *N* according to arrival rate, which is where another difficulty in the Poisson approach arises. Cyclist arrival rates are, of course, not constant from hour to hour, day to day, or month to month. Even within a single day the rate of arrivals at a counting station fluctuates substantially, which is illustrated in Figure 5; on the King Street Cycle Track in Hamilton, the average hourly arrival rate, calculated across multiple years' worth of data for that counting station, has a range of 27 cyclists/hour, following a commuter-peak pattern which is consistent across other locations. We cannot then use just one constant arrival rate to calculate *N* across an entire dataset. Referring again to the example in Figure 4, a λ value of 1 would give us an *N* value of 3 for X=2 at β=99.99%, compared to *N*=7 for λ=8. The higher arrival rate here would result in far fewer sequences of counts being flagged than would the lower. Furthermore, since *N* is also a function of the interval count value, and the functional forms seen in Figure 4 are different for every value of X, a filtering algorithm using a constant arrival rate could not be said to be applying itself consistently across a dataset.

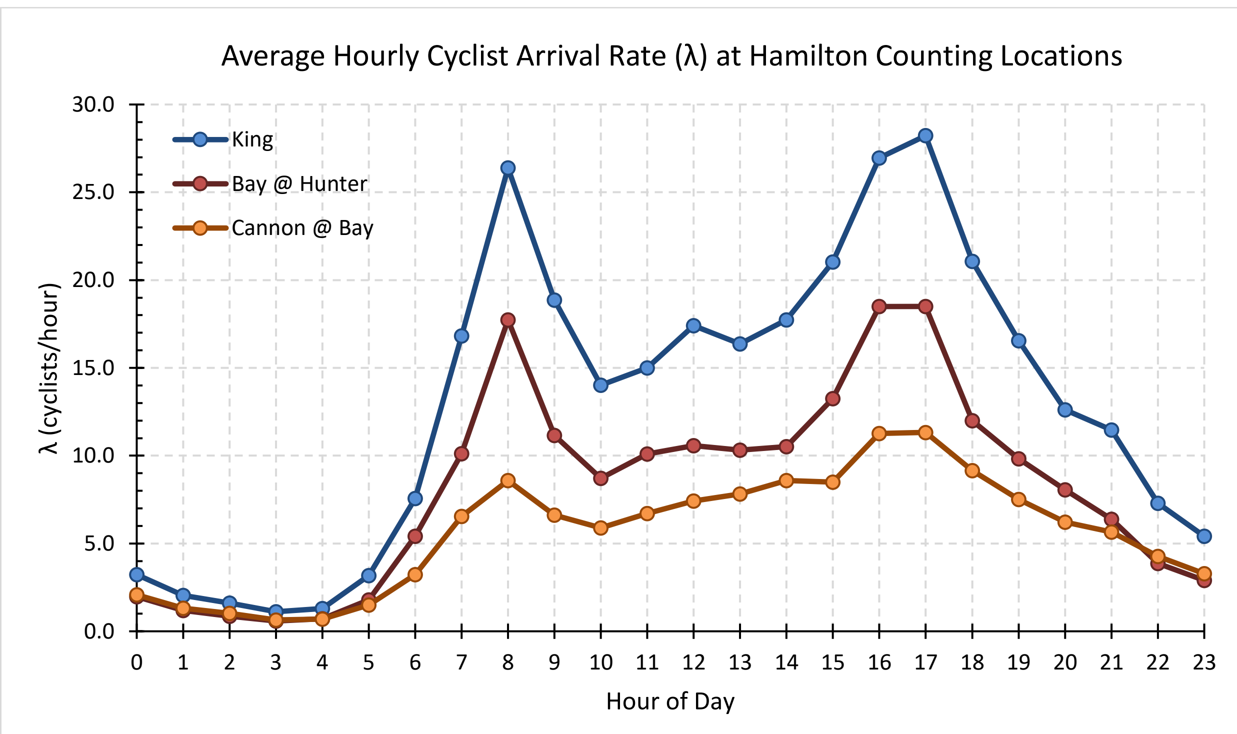


Figure 5: Average hourly cyclist arrival rate (λ) by hour-of-day for three Hamilton counting stations

We can address this issue using a moving window which sums every interval count with the *previous* two and *next* one consecutive values, expressed below in Equation (7):

$$\lambda_i = X_i + X_{i+1} + X_{i-1} + X_{i-2} \tag{7}$$

These four 15-minute intervals are a good representation of the hourly arrival rate at a given 15-minute count interval *i*. Every *i* in the dataset will have an associated λ value calculated in this way, and $P[X_i]$ will be calculated using this value instead of a constant. This sampling pattern was chosen to 'centre' the interval count in an hourly arrival rate, rather than summing the prior 3 or subsequent three counts which would put the interval at either the leading or trailing end of that hour, respectively.

2.2.3.2 Filter 2: Consecutive Zero Counts

Issues with the Poisson method arise when considering runs of consecutive zero counts, or sequences of intervals where no cyclists were counted, which may indicate that the counter has stopped recording counts correctly. As discussed above, this should be distinguished from a "null" or empty data entry, which occurs when the equipment stops recording entirely, and which are all considered counting errors. A zero count may still be a valid entry, however an extremely long sequence of consecutive count intervals with zero counts is usually indicative of a counting error, such as if snow has built up against the sensor so that it is still technically operating but is unable to detect any passing cyclists [8]. Since zero counts are still by far the most common entry in the 15-minute interval datasets we have examined – resulting from the long, lean daily off-peak periods most cyclist infrastructure experiences – this threshold must be appropriately high to avoid incorrectly flagging data entries.

Although it is discussed in the previous section how using a constant λ value is not appropriate for the calculation of *N*, in the case of consecutive zero counts the opposite may be true. As Figure 6 below illustrates, *N* is especially sensitive to the value of λ when λ is small and X = 0. For this reason, most of the relevant studies from the literature use a constant value of *N*.

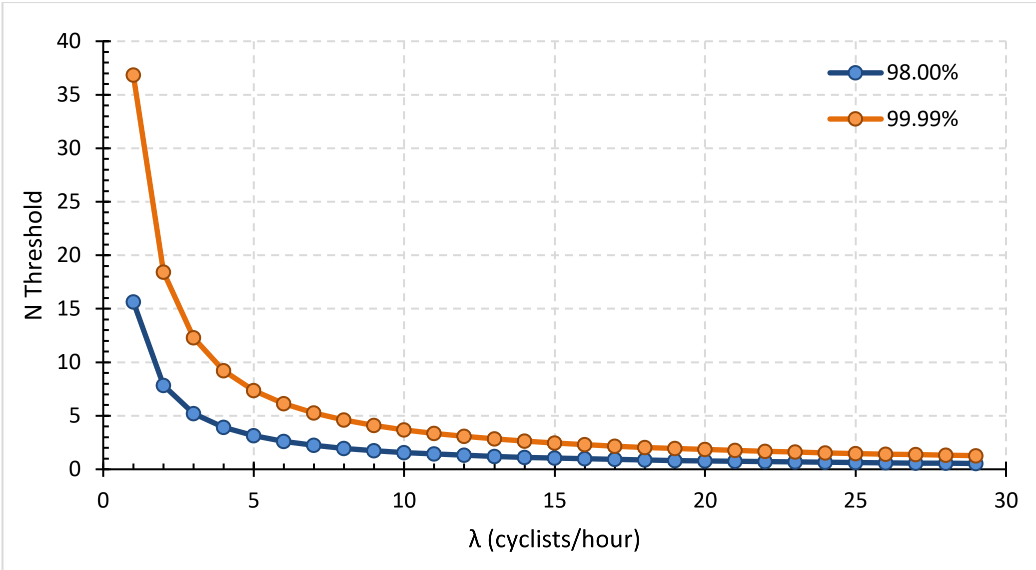


Figure 6: Value of N by hourly cyclist arrival rate (λ) for X = 0, at different levels of confidence

Different researchers have found different thresholds to be appropriate, although there is no real consensus on what the most appropriate value is. While the Colorado Department of Transportation uses a threshold of over 2-days of consecutive zeros [10], the U.S. Federal Highway Administration Travel Monitoring Analysis System (TMAS) filters runs of just 7-hours or longer [12]. Research done at Portland State University placed the threshold somewhere in the middle, finding that runs of over 15-hours were uncommon [9]. It is possible that *N* should be geographically specific, but there is no clear basis for this in the literature, nor does there not appear to be an objective method to calibrate the optimal value for *N* for consecutive zero-counts.

2.2.3.3 Sensitivity Analysis

Unlike for non-zero counts - which will be discussed in the next section, and where counts are filtered using the β confidence limit - there is no way of calibrating our willingness to incorrectly identify valid runs of zero-count entries. Only by setting a constant value for *N* high enough can we be more confident in correctly identifying invalid data, and then we run the risk of failing to identify it at all. For this reason, the filter results should be considered cautiously.

Figure 7 shows the relative frequency of consecutive zero-count runs- by run-length across five Hamilton counting locations, illustrating how the value of N would impact the proportion of sequences being filtered. Instances of individual zero-counts, or zero-counts which are not part of a sequence of consecutive zeros, are excluded from these calculations because: one, they are not the target of this filter; and two, they make up by far the largest proportion of zero-counts, individual interval counts of zero being much more common than observing any other sequence length. The five counting stations included in this graph have very similar run length distributions, and so have been aggregated here for the sake of graphical clarity. From this figure, and from the supporting data, we can make several observations. First, couples of zero-counts, or runs of length 2, are by far the most common. This is to be expected as they are more likely to occur throughout the day, even during on-peak hours, than longer sequences. Second, the relative frequency of run-lengths decreases as N increases, and the rate of that decrease also decreases as N grows. Third, runs of zero-counts between 25 and 60 intervals long (6.25 to 15 hours long) are not common, representing just under 1% of all sequences witnessed, although they are still occasionally observed. Fourth, runs greater than 61 intervals long are very rare, comprising just 0.05% of the observed sequences for these 5 locations. Finally, there does not seem to be objective way to determine the appropriate value of N, but by plotting relative frequency as we have done here it is possible to get an impression of what will be included and excluded from the filter depending on what N is chosen. In this case, it would seem as if 60 would be an appropriate value for N because of the rarity of runs above this length.

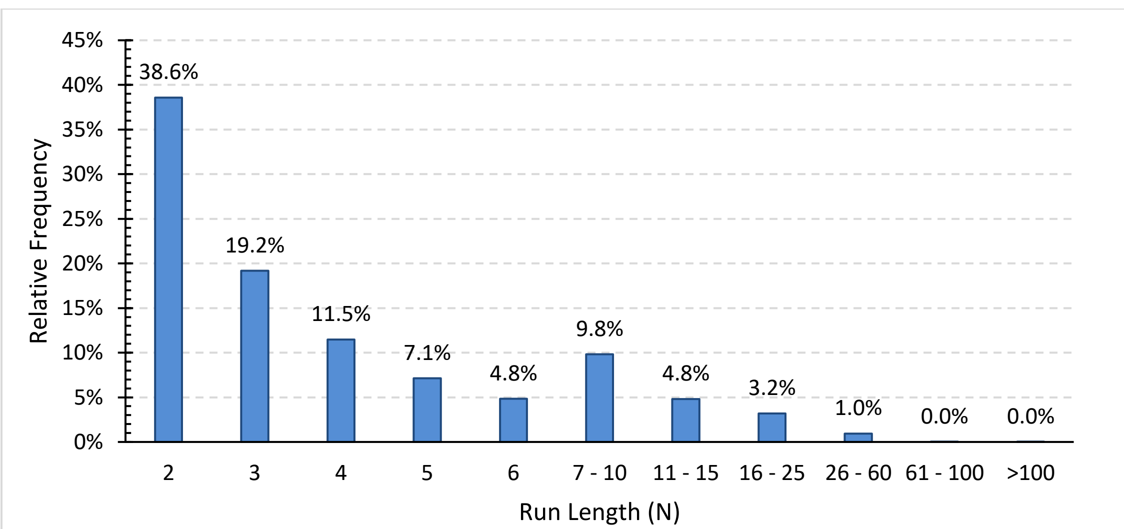


Figure 7: Relative frequency of consecutive zero-runs by run-length, for five Hamilton counting locations

Another way to examine zero-count runs is to plot the relative frequency not of the runs themselves by length, but of intervals by the length of the runs of which they are a part - as is done in Figure 8, which also displays cumulative interval frequency. So, for instance, although runs of length 2 are the most frequently occurring sequences, representing 38% of all sequences, the actual number of intervals which are part of a sequence of two consecutive zero-counts make up only 13% of the total number of intervals which are in a zero-count sequence. This helps give some idea of how many intervals will actually be filtered depending on the value of *N*. From Figure 7 we came to the idea that 60 might be an appropriate value for *N*. Here, we might come to the same conclusion by examining Figure 8. The relative frequency of intervals in a run of between 61 and 100 zero-counts is just 0.24%, far lower than the other bins displayed. 85% of intervals which are in a zero-count sequence would be excluded from this filter. The relative frequency for *N* > 100 is again much higher, but this is to be expected as this bin captures those sequences which run into the thousands of consecutive zero-counts and so are obviously erroneous. An *N* value of 60 then could reasonably be thought of as a threshold where the rate of occurrence of accurate sequences of zero-counts has peaked and where inaccurate sequences begin to appear.

Accordingly, in this study a value of *N* of 60 count intervals has been chosen from the literature [9], representing 15 hours of consecutive intervals, which is consistent with the findings above.

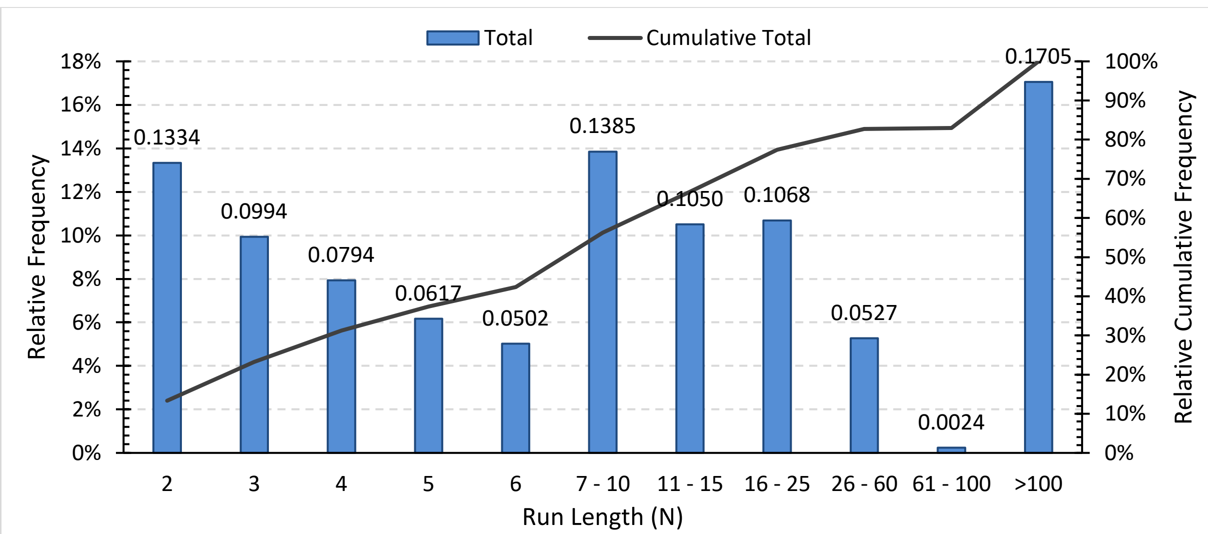


Figure 8: Relative frequency and cumulative relative frequency of interval zero-counts by the length of the run which they are a part of, across five Hamilton counting locations

One Hamilton counting location, Bay at Stuart, was excluded from Figure 7, and displayed instead in Figure 9. The run-length distribution of this location illustrates one of the issues with using this sort of analysis, which is that there are still locations, such as this one, were the constant value which was appropriately applied to other counting stations might still be inappropriate. Bay at Stuart is a very low-volume location, which sees substantially fewer daily cyclists than the other Hamilton locations included in the analysis above. Because of this, long runs of zero-counts are much more likely. We can see from Figure 10 that intervals that are part of runs which are between 61 and 100 intervals long make up a full 5.3% of the sequenced intervals for Bay at Stuart, compared to 0.24% across the other five. Using an *N* threshold of 60, these intervals would all be flagged. It is entirely plausible that the counting equipment was operating correctly when those runs were recorded but that other factors, weather especially being one, may have been responsible for the large number of consecutive runs. For instance, one flagged day in the Bay at Stuart dataset would be January 11, 2019, a day when temperatures in Hamilton were very cold and substantial snow had built up; although the filter clause would be triggered here, it is not inconceivable that there were no cyclists for this entire day even though the literature, and our own analysis on other counting locations, tells us this span of time is unlikely.

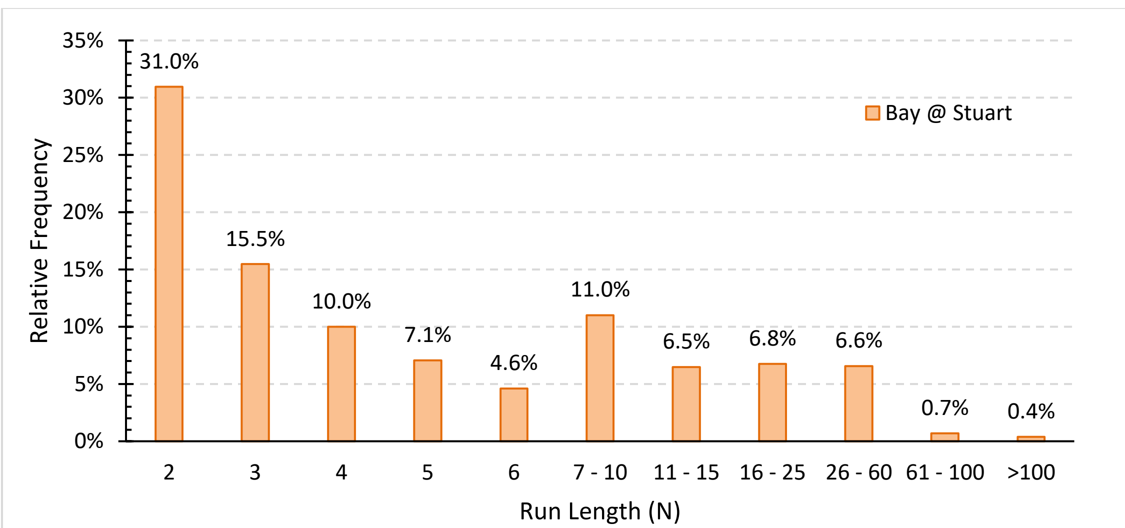


Figure 9: Relative frequency of consecutive zero-runs by run length, for Bay @ Stuart counting location

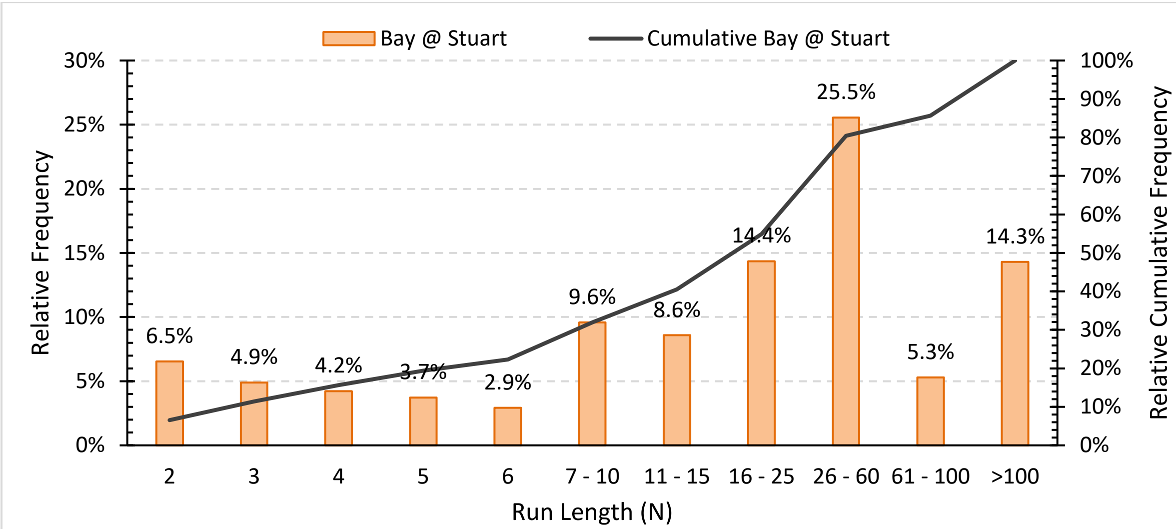


Figure 10: Relative frequency and cumulative relative frequency of interval zero-counts by the length of the run which they are a part of, for the Bay @ Stuart counting location

2.2.3.4 Implementation

To implement Filter 2 (F2), we can use the counting variable *n* to represent the number of consecutive zero-counts in a sequence. If the value of X_i is zero, *n* is set to 1; otherwise it is set to zero. If the value of both X_i and X_{i-1} are zero, n_i is set to $n_i + 1$. For every consecutive interval after this where the count is 0, the value of *n* increases by 1. If a non-zero value appears, n is reset to zero; in this way the value of *n* indicates the number of consecutive count intervals for which the reported count volume = 0. If *n* becomes greater than or equal to the maximum threshold value (*N*) defined in the previous section, then that interval and each of the N-1 previous intervals where the count is zero are flagged. Figure 11 shows a sample implementation of F2 on data from a counting location in Hamilton.

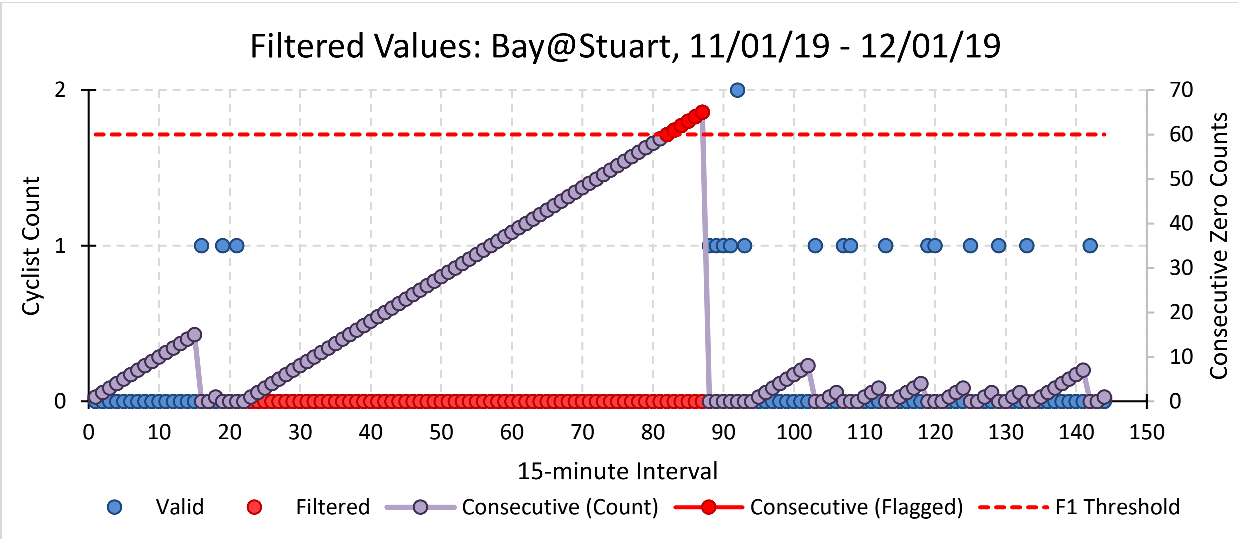


Figure 11: Sample of filtered daily-zero values for N=60, from Hamilton counting location Bay@Stuart

2.2.3.5 Filter 3: Consecutive Non-Zero Counts

Although, as was discussed in the previous section, there is some difficulty in our Poisson approach when the arrival rate λ is equal to zero. When X > 0 so to is λ, so the full Poisson approach can be applied to identify erroneous runs of non-zero counts. This is the basis of Filter 3 (F3).

The algorithm is as follows:

Step 1: Determine the current length of the sequence:

If X_i <> X_{i-1}, then set n = 1; otherwise, n = n+1

Step 2: Determine the arrival rate for the current count interval:

$\lambda_i = \Sigma(X_i , X_{i+1} , X_{i-1} , X_{i-2})$

Step 3: Compute the probability of observing X_i in the current interval:

$$P[X_i = X] = \frac{(\lambda T)^X e^{-\lambda T}}{X!} \tag{8}$$

Step 4: If $n > 1$, then compute the probability of $P[X_i|X_{i-1}| \dots |X_{i-n}]$:

$$P[X_i|X_{i-1}| \dots |X_{i-n}] = P[X_i = X] \times P[X_{i-1}|X_{i-2}| \dots |X_{i-n}] \tag{9}$$

Step 5: If the probability $P[X_i|X_{i-1}| \dots |X_{i-n}]$ is sufficiently small (i.e. $< (1-\beta)$ then declare the count data for intervals *i*, *i*-1, …, *i*-*n* as suspect.

Like with zero-count runs, some calibration of the filter threshold is needed. Although the *N* threshold here is being replaced with the direct β probability threshold, it is useful to plot the relative frequency of non-zero runs by length, as has been done in Figure 12, to get an idea of the frequency with which runs occur. What we can see is that runs of length two are by far the most common, as would be expected under our Poisson process assumption as the probability of observing a run decreases exponentially with its length. While runs of length 3 and 4 are not uncommon, longer runs occur less and less frequently, with no runs longer than 11 observed across the entire Hamilton dataset.

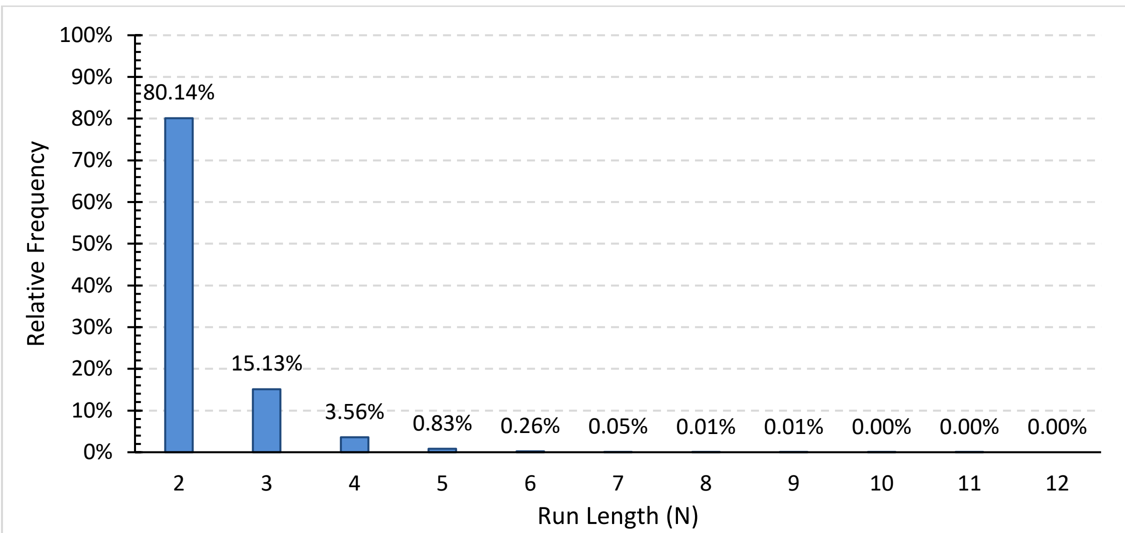


Figure 12: Relative frequency of consecutive non-zero runs by length, across 6 Hamilton counting locations

In testing the filter algorithm, it was noticed that occasionally runs of just two-consecutive non-zero values were filtered. Usually, the cause of this could be attributed to the method used for calculating cyclist arrival rate (λ); for instance, where an interval with a very high count, which brings up the value of λ, was followed by two intervals with much lower count values, then those intervals would be flagged as they are considered unlikely relative to the high arrival rate. Given this, and referring to Figure 12 above, an extra condition was added to Step 5 of the F3 algorithm, preventing it from filtering runs of less than 4 consecutive counts:

Step 5*: *IF:* $P[X_i|X_{i-1}| \dots |X_{i-n}] < (1-\beta)$

AND: *n* > 4, then declare the count data for intervals *i*, *i*-1, …, *i*-*n* as suspect.

One variable which still needs to be set is the confidence limit (β), although once again it is difficult to do so in a purely objective way. Remember that a high β value reduces our risk of mistakenly identifying valid data, but decreases the number of runs which are identified, creating a risk of failing to identify invalid data. A low β increases the probability of mistakenly identifying valid data, but also ensures that all the invalid data will be captured. Since erroneous sequential non-zero counts are relatively rare in our estimation, the best strategy would seem to be setting the threshold on the high-end to avoid filtering too many sequences. With extremely large datasets of counts, sequential counts are bound to happen purely by chance even if the probability of observing them is low. Figure 13 below shows the sensitivity of the number of sequences filtered from the Bay @ Hunter dataset, by β-values, which stresses the importance of appropriately setting the threshold. The difference between 99.85% and 99.90% at this location is the difference between filtering 19 or 8 sequences. At 99.99%, no sequences at all are filtered.

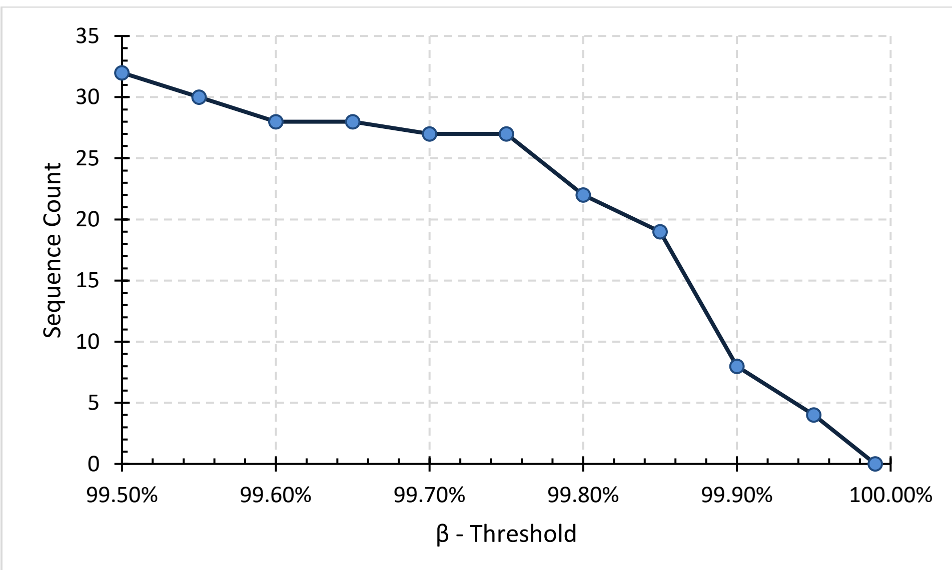


Figure 13: Count of filtered sequences by β-threshold, for Hamilton counting location Bay @ Hunter

For this project, a β-threshold of 99.95% was chosen, high enough to filter only the most improbable consecutive non-zero count sequences.

2.2.4 Filter 4: Hard Cap Value

The likelihood of observing a count in a 15-minute interval which is higher than the expected daily volume at that counting location is very unlikely. For a given day and daily volume, we simply would not expect for all those cyclists to have arrived at the counting station during just one 15-minute interval because of what we know about cyclist behaviour, even if arrival rates do vary greatly over the course of a day. We can observe much more consistent and distributed patterns of arrivals, which makes extreme variations from those patterns suspicious. If such a variation is present in the dataset, it is likely the result of a counting error or equipment malfunction and should be flagged.

As with some previous filters, the conditional value C used here was adapted from the literature. McNeil et al [8] performed an analysis on a large cyclist dataset from Portland, Oregon, to develop cyclist count quality checks, and produced the following table of recommended C values:

Table 6: Recommended hard-cap values, by expected daily volume, adapted from McNeil et al [12]

Expected daily volume	Suspicious	Possibly Suspicious
Unknown	1000	500
< 100	250	100
100 – 500	500	250
> 500	2000	1000

Several of these values were tested on the Hamilton data, but no data points were flagged, even for the lowest thresholds. Clearly such high 15-minute interval counts are also extremely rare, even non-existent, in the Hamilton data. However, the filter was maintained in case the issue arises in future datasets, using a *C* value of 250 (the "suspicious" count value for locations with <100 daily expected riders). Note that, as can be seen in the table above, a hard-cap filter could be made more robust by making it volume dependent, where the hard cap is a function of the expected daily value. However, this raises some complications about the calculation of expected daily values which changes dramatically from season to season and even within the season depending on many external factors. For the purposes of identifying abnormally high 15-minute interval counts here, a non-volume-specific threshold is enough.

2.2.5 Filter 5: 24-hour Zero Counts

Counting equipment is often installed in high-volume locations, to maximize its benefit relative to the cost of installing it. With this in mind, 24-hour counts of zero should be considered suspicious. Although it is possible that no cyclists pass a station in a day, even at a high-volume location because of the seasonal variations inherent to cycling volumes, this filter flags daily zeros for further investigation to determine whether the result was erroneous. For instance, a zero value on Christmas day would be more likely than a zero value during the peak summer cycling season. Most daily zero values will be captured by F2, however the F2 algorithm can be interrupted by a “null” value, in which case the sequence wouldn’t be flagged. So, it is necessary to include this daily filter, and it additionally makes it easier to quickly identify when the zero counts occurred.

2.2.6 Filter 6: Daily Maximum (IQR Cut-off)

While abnormally high-count values are dealt with at the 15-minute interval level by F4, there also needs to be a way of flagging upper outliers in 24-hour counts. These filtered data points don’t necessarily represent a technical failure, as very high daily counts can occur accurately for a number of reasons (for instance, cycling counts were abnormally high in the City of Hamilton datasets on May 28, 2019, “Bike to Work Day”). However, whether these counts are accurate, they should still be flagged and assessed for suitability in future analysis. If an outlying daily count is a result of some one-off event, and not of the permanent, inherent characteristics of that counting location, then it is not suitable for building a model for expanding short term counts.

A relatively common threshold for filtering outliers, not just in the realm of cyclist count quality checks but in other data science applications, is the 2(IQR)+Q3 formula shown by Equation (10) [12]. The issue with using this formula, like with some of the other filters, is that cycling activity varies extremely over the course of the year, in Canada especially, so that the inter quartile range (IQR) of an entire dataset spanning multiple years may not accurately represent any given single day’s count. For this reason, a 27-day moving window is applied to give every daily data feature an associated quartile value based on the previous 13, and the subsequent 13 days. If the daily value exceeds the filtering threshold (β_{F6}) as calculated for that 27-day window, then it is flagged.

$$\beta_{F6,t} = 2\left(Q_{3,t} - Q_{1,t}\right) + Q_{3,t} \tag{10}$$

Where:

$\beta_{F6,t} =$ Filter 6 threshold for day-of-year t

$t =$ The day-of-year, expressed as an integer value between 1 and 365

$Q_{1,t} =$ Quartile 1, calculated across the set of 24-hour counts for days in the range [t-13 .. t+13]

$Q_{3,t} =$ Quartile 3, calculated across the set of 24-hour counts for days in the range [t-13 .. t+13]

Important to note that, as this is the last filtering operation performed, it can be performed on a reduced dataset which does not include any of the previously filtered counts. If a datapoint has been filtered by any of the previous five filters, it is skipped by the moving window when calculating quartiles (e.g., a “null” value will not be one of the 27 values used to calculate the IQR). As well, this technique means that the filter will not be applied to the first 13 and last 13 days of data, as they fall outside of the moving windows range.

Figure 14 below shows the 24-hour counts which were flagged by F6 for a single Hamilton location. Although most of the flagged values are obviously extreme upper outliers, occasionally unexpected values are also flagged; for instance, 08/11/2018 from Bay@Cannon is a relatively low count value of 118, but comes on the downward slope of the count curve and is substantially higher than any of the subsequent 13 daily counts, which is why it was flagged. The filter performed “correctly”, but some consideration might need to be given as to whether this is a value worth eliminating from the dataset.

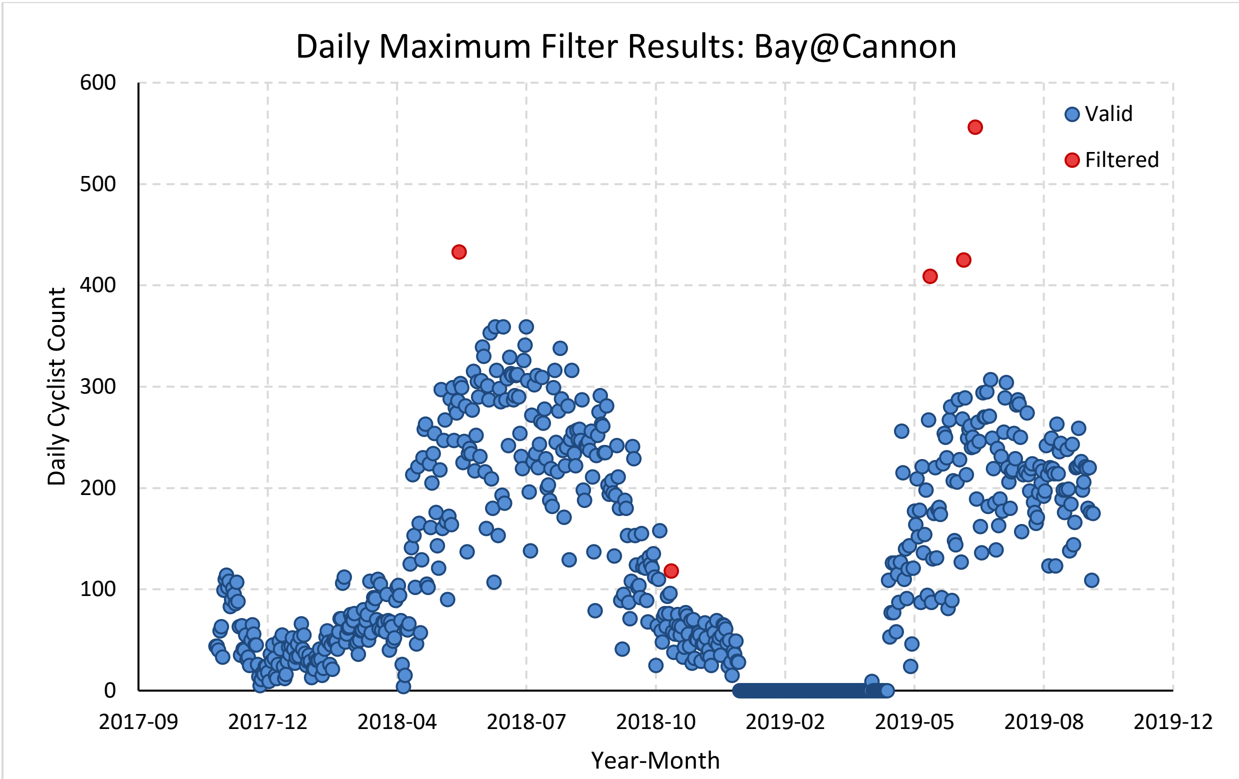


Figure 14: Daily counts flagged by Filter 6 across two years of 24-hour counts from the Bay@Cannon count location

2.2.7 Filter Summary

Table 7 below contains a summary of the filters described in this section. These filters were applied to count data from six of the Hamilton locations; summaries of filtering results are given by Table 8, Table 9, and Table 10.

Table 8 shows the results of Filters 1 through 4, which are applied at to the raw, 15-minute interval counts. A large number of "null" intervals were identified by F1, and an even larger number of intervals which were a part of a run of consecutive zero-counts. Over 10,000 15-minute intervals were flagged by this filter for the Bay@Cannon location, representing over 16% of all data; this is a highly suspicious result, and further investigation found that the road was closed for construction during a long stretch in early 2019, resulting in no cyclists being counted. Very few sequences of consecutive identical non-zero counts were identified, and no counts at all were flagged by F4 for 15-minute outliers.

Table 9 shows filtering results at the daily level; F5 and F6 operate on 24-hour counts, but for Filters 1 through 4 a count is given of days on which at least one 15-minute interval was flagged. A large number of daily zero values were flagged by F5 across all six locations, although examining the results it was noticed that most of these occurred during the winter months and so likely represent valid counts. Very few days were flagged as 24-hour count outliers by F6 (just 26 count days across all 5714 represented in the dataset). The number of count days flagged at each location ranged from as low as 1.85% at Cannon@West, to as high as 18.4% at Bay@Cannon. In total, 89.3% of the count-days represented in the Hamilton dataset were found to be valid according to the six filters described in this section.

Table 7: Summary of filtering process and sub-processes

#	Name	Description	Condition	Threshold
Group One – 15-minute counts				
1	Null Flag	Counts expressed as "null", but not zero	X_i = " "	N = 1
2	Consecutive Zero Counts	Sequences of 60 or more consecutive counts of zero	$X_i = X_{i-1} = 0$	N ≥ 60
3	Consecutive Non-zero Counts	Sequences of non-zero count values where the probability of observing that number of consecutive identical counts is less than the threshold β	$X_i = X_{i-1}$, $X_i > 0$	$P[X]^n < \beta$, B = 0.01%
4	Hard Cap	Upper limit on the number of cyclists observed in a 15-minute period, as a constant value	$X_i \geq 250$	N = 1
Group Two – Daily Counts				
5	Daily Zero Flag	Flag daily intervals with a count value of zero	$X_i = 0$	N = 1
6	Interquartile Range Flag	Upper limit on the number of cyclists observed over one day, based on surrounding days data	$X_i > 2(IQR) + Q3$	N = 1

Table 8: Summary of 15-minute interval flags

Counting Station	**Total 15-min Intervals**	**1 – Null Entry's**	**2 – Zero Counts**	**3 – Non-Zero Counts**	**4 – Hard Caps**
Bay@Cannon	65,184	144 (0.22%)	10778 (16.53%)	4 (0.01%)	0 (0%)
Bay@Hunter	64,608	336 (0.52%)	8835 (13.67%)	3 (0.00%)	0 (0%)
Bay@Stuart	63,936	8 (0.01%)	4858 (7.60%)	9 (0.01%)	0 (0%)
Cannon@Bay	133,440	3175 (2.38%)	7205 (5.40%)	19 (0.01%)	0 (0%)
Cannon@West	109,152	124 (0.11%)	555 (0.51%)	15 (0.01%)	0 (0%)
King West	112,224	857 (0.76%)	9202 (8.20%)	9 (0.01%)	0 (0%)

Table 9: Summary of daily count flags by counting location

Counting Station	**Total Daily Counts**	**1 – Null Entry's**	**2 – Zero Counts**	**3 – Non-Zero Counts**	**4 – Hard Caps**	**5 - Daily Zero**	**6 - Daily Max**	**Flagged Days**	**% Days Flagged**
Bay @ Cannon	679	5	115	2	0	114	5	125	18.41%
Bay @ Hunter	673	7	95	2	0	96	2	104	15.45%
Bay @ Stuart	666	2	89	4	0	43	12	107	16.07%
Cannon @ Bay	1390	39	80	11	0	107	5	134	9.64%
Cannon @ West	1137	4	7	11	0	6	0	21	1.85%
King West	1169	12	98	8	0	103	2	120	10.27%

Table 10: Aggregate filter summary for Hamilton datasets

Filter:	**Days Flagged**
1 - Null	69
2 - Zero Count	484
3 - Non-zero	38
4 - Hard Cap	0
5 – Daily Zero	469
6 - Daily Maximum	26
Total Flagged:	611
Total Days:	5714
% Valid	89.31%

2.3 Adapting Conventional Methods to Turning Movement Counts

As described in the previous section, work was done to develop a robust set of filtering rules for conventional cyclist count data, adapting from the literature accepted filtering logic to go alongside some novel approaches. Importantly though, this work, in consistency to the bulk of the literature on the subject, was done using data from a group of conventional, automated, continuous mid-block counters installed on high-volume dedicated cycling infrastructure in Hamilton, ON. These filtering rules then may not be appropriate for the purpose of filtering turning movement counts (TMCs), or more specifically the video monitoring unit (VMU) data used in this study, which differs from conventional cyclist counts in a number of important ways:

1. Continuous counts of cyclists at intersections are often collected using a real-time vision-based system rather than the infra-red sensors and/or induction loop sensors commonly used to measure conventional active-transportation counts. These technologies are subject to different potential sources of failure.
2. By definition, TMCs are obtained at intersections (most frequently signalized intersections) whereas conventional counts are usually taken "mid-block", or at least, outside of intersections, so that counts are at most bi-directional. TMCs capture the turning-movement specific, cardinal directionality of traffic at intersections.
3. Permanent conventional counters are usually installed at high-volume locations to maximize their effectiveness. VMU data is obtained at intersections, often for the purposes of assessing traffic signal operations, and therefore cyclist volumes may be quite low.
4. Finally, the VMU data contains counts of both motorized vehicles and cyclists, whereas conventional active transportation sensors are most frequently deployed so that they do not provide counts of motorized vehicles.

Accordingly, some adjustments need to be made to the process described in Section 2.2. One year of VMU data – from January 1, 2019 to December 31, 2019 – was obtained for eight intersections in Milton, Ontario, aggregated to 15-minute intervals. This dataset was evaluated under the conventional filtering framework, and the results used to adapt the process for continuous TMC data.

2.3.1 Filter 1: Missing Data

The filtering of conventional cycling count data requires methods to determine when the detector system has malfunctioned and has not recorded count data. Systems may be designed such that when the system malfunctions and no count is made, then a “null” entry is recorded in the database (this was the case of the Hamilton dataset first used to calibrate filtering rules). Conversely, the system can be configured such that when it is malfunctioning and nothing is recorded, the reported count value is zero (this is the case of the Milton VMU data used in this study). The difference between the way these two sets of counts are recorded creates an issue for the application of *Filter 1* on the Milton VMU data, which was designed to look for “null” entries which may no longer exist.

Of course, when a system is operating correctly but no cyclists pass the detector during the count interval, then the system also reports a zero count. In conventional count data, the difference between a valid and invalid zero-count is made somewhat more distinct by the presence of “null” entries. The likelihood of a zero count being valid is highly dependent on the mean volume of cyclists. When the mean volume is small, then the probability that no cyclists pass the detector during the interval is relatively high and we must have many consecutive intervals with a zero count before we can reliably conclude that that the system was malfunctioning.

However, the video-based monitoring system which collected the data used in this analysis has the advantage over conventional cyclist monitoring systems of reporting for each minute not only the number of cyclists, but also the number of motorized vehicles (by type) and pedestrians observed for each turning movement (or pedestrian crossing). These 1-minute data records (termed the “raw” data) were aggregated to provide counts for 15-minute time intervals. We make the assumption that if the system malfunctions (e.g. camera view is obscured, camera power failure, communication failure, etc.), then counts are not available for traffic, cyclists, and pedestrians. Consequently, we use the combined motorized vehicle and cyclist counts (termed traffic counts) across all turning movements as the means to identify when the system is malfunctioning.

A traffic count of zero in a 15-minute interval may represent one of two things: (i) the counting system was operating correctly, but no vehicles or cyclists, passed through the intersection during the aggregation period, or (ii) the detector system was not operating correctly and count of zero has been recorded even though some traffic passed through the intersection, in which case the count should be flagged as invalid.

Examination of the data showed that traffic counts equal to zero for one or two consecutive intervals occurred relatively frequently but when they occurred, they most often occurred during times of the day when traffic volumes are expected to be very low, and so probably did not represent a system malfunction. Consequently, it was determined that if a zero count is reported by the system for three or more consecutive intervals (i.e. for 45 minutes or longer), then it is concluded that the system may have malfunctioned and we label those intervals as suspect. An updated filtering algorithm for Filter 1 (F1) can be seen below:

Step 1 Determine if the count X_i is equal to 0, and adjust count c:

If $X_i = 0$, then $c = 1$; else $c = 0$

Step 2 Check the value of c to determine the length of the sequence:

If $c \geq 3$, then declare count data in intervals

[i, i-1, … , i-(c+1)] to be suspect

2.4 Filters 2-6

The distinction between "null" and zero-counts is discussed above. Conventional sensors for obtaining cyclist counts may exhibit failures by failing to report any count (i.e. "null" count in the database), by reporting a zero count for a series of consecutive intervals, or by reporting the same count value (greater than zero) for a series of consecutive intervals. As such, filtering of data from these conventional sensors typically uses some threshold of reported zero counts over a number of consecutive 15-minute intervals and daily totals of zero to identify system malfunctions.

The application of these filters though is made difficult by some of the distinctions between conventional cyclist counts and TMCs. Normally, filtering approaches would be applied to data obtained from permanent sensors installed on facilities where motorized vehicles are prohibited, as these are often the highest volume locations and the most valuable to collect count data on. Long runs of consecutive zero counts are unlikely at these locations - and daily counts of zero even less so - and can be reliably filtered as erroneous. This is not always the case at intersections where VMUs are installed, which are not selected because of their high cyclist volumes; the least trafficked of the Milton locations included in this study experienced, on average, just over 5 cyclists per day in 2019. At these locations, zero-counts are much more likely, and even full days of daily counts cannot be assumed to have been a result of an equipment malfunction.

Seasonality also plays a large part. Figure 15 shows that daily zero-counts are by far more common in the Milton database during off-season months (December through April) than during the summer peak, when volumes are highest. Evidence shows a strong negative relationship between cycling volumes and winter weather [15], so it might be safer to assume that off-season zero-counts are valid than that they aren't.

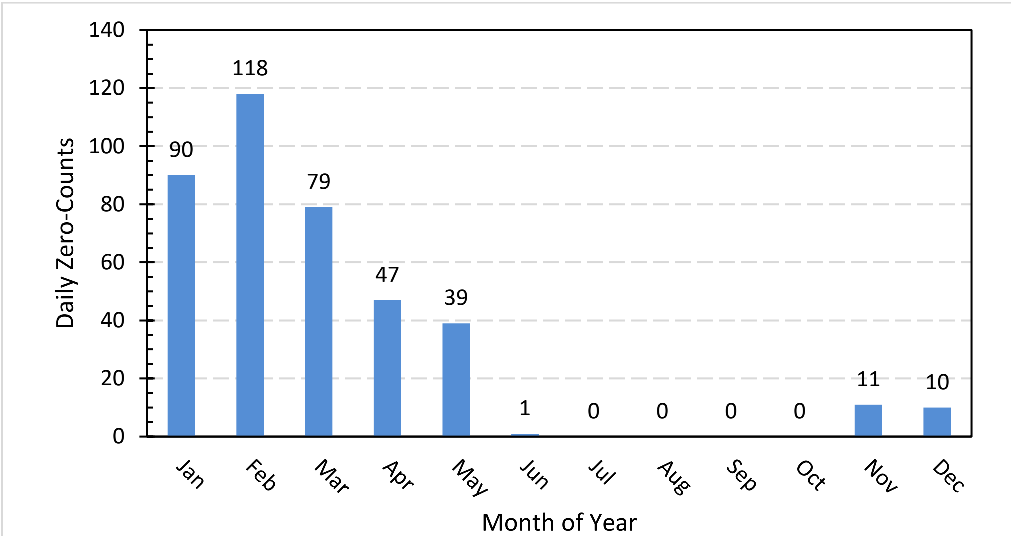


Figure 15: Count of 2019 days with a recorded daily cyclist volume of zero, across eight Milton locations

It can therefore be seen that consecutive and daily zero-counts do not represent a reliable basis for filtering of the Milton VMU data, as it would be for conventional cyclist counts. Unlike conventional count data, however, the Milton dataset provides counts for both automobile and cyclist traffic. Since automobile traffic volumes are much higher than cyclist volumes at all the study locations, missing automobile counts – as is discussed in the previous section – provide a more reliable basis for identifying equipment malfunctions than cyclist counts of zero. For these reasons, results of Filters 2 and 5 are ignored for the purposes of this portion of the project, under the assumption that the erroneous counts which would have been captured by using a zero-cyclist count filter will be identified under the new Filter 1 for zero traffic counts instead.

Filters 3, 4, and 6 were found to be appropriate for VMU data and were carried through as is.

2.5 Primary Filtering Results - Milton

Primary filtering results for 2019 Milton data can be seen in Table 11 and Table 12.

Table 11: Summary of 2019 15-minute filtering results for target Milton count locations

#	Name	Intervals	F1 Flags	F2 Flags	F3 Flags	F4 Flags	Total	%
4	Thompson Road and Childs Drive	35040	4	-	7	0	11	0.0%
5	Thompson Road and Laurier Avenue	35040	426	-	0	0	426	1.2%
9	Ontario Street South and Pine Street	35040	49	-	0	0	49	0.1%
14	Bronte Street North and Main Street West	35040	488	-	0	0	488	1.4%
15	Ontario Street South and Childs Drive	35040	4	-	0	0	4	0.0%
17	Main Street East and Mall Entrance	35040	8	-	0	0	8	0.0%
22	Thompson Road and McCuaig Drive	35040	1224	-	0	0	1224	3.5%
25	Main Street East and Ontario Street North	35040	429	-	0	0	429	1.2%
Total		280320	2632	-	7	0	2639	0.9%

Table 12: Summary of 2019 24-hour filtering results for target Milton count locations

#	Days	F1 Days	F2 Days	F3 Days	F4 Days	F5 Days	F6 Days	Total	%
4	365	1	-	1	0	-	0	2	0.5%
5	365	22	-	0	0	-	0	22	6.0%
9	365	2	-	0	0	-	4	6	1.6%
14	365	7	-	0	0	-	5	12	3.3%
15	365	1	-	0	0	-	15	15	4.1%
17	365	2	-	0	0	-	9	10	2.7%
22	365	17	-	0	0	-	8	25	6.8%
25	365	101	-	0	0	-	2	103	28.2%
Total	2920	153	-	1	0	-	43	195	6.7%

Relatively few days were flagged by F1, with the exception of Location 25 where at least one "null" 15-minute interval was observed on a large number of days. Subsequent examination found that all of these "null" intervals occurred during off-peak hours, when traffic is expected to be the lowest. For this reason, these counts were not considered suspicious. F1 results for other locations however still raised some concern. Take location 14 for instance, at Bronte Street and Main Street. At this location, 488 15-minute intervals were flagged under F1, but these came from only 7 days. For four of the seven flagged days, total traffic counts for all 96 15-minute intervals were zero, which we can say with certainty represents an inaccurate count. It is not reasonable to think that 24 hours passed without a single vehicle of any kind passing through the intersection, even at a low volume location.

Only one sequence, which contained only seven consecutive identical values, was flagged under F3, this being at Location 4. No F4 flags were thrown for any of the eight target locations. The results for F2 and F5 are not included here, acknowledging the discussion of the issues they present in the previous section.

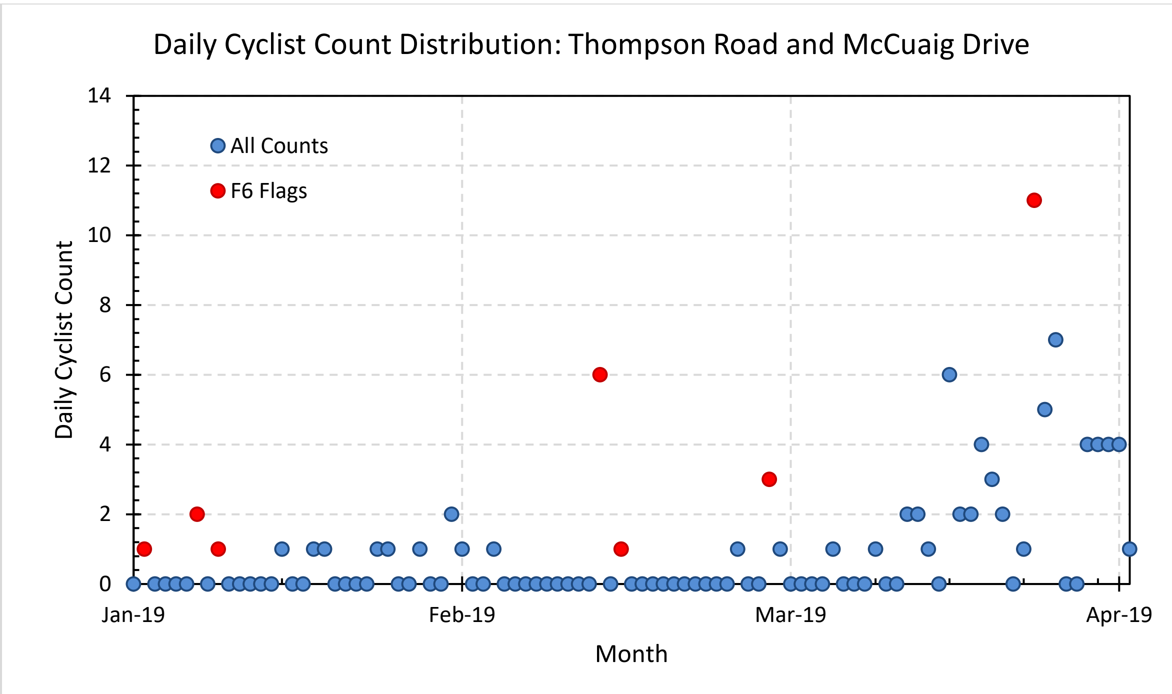


Figure 16: F6 (Daily Upper Outliers) flags for Milton count location Thompson Road and McCuaig Drive, Jan-2019 to April-2019

Quite a few days are identified by filter F6 for having count values outside of the 2(IQR) range. Keep in mind that these days do not necessarily represent "incorrect" counts, but this filter too should not be ignored as the counts flagged by it are still problematic. Take the early-2019 count distribution for a relatively low volume count location shown in Figure 16. On February 19, 2019 a daily count value of 6 was flagged by F6. While this may not seem like a particularly high-count value, it represents a significant outlier compared to the rest of the month, during which only 10 total cyclists were counted. This count on this day is not representative of the average expected volume at this location and should be identified as such. For this reason, a minimum count threshold of 15 was implemented, so F6 would only be triggered when a 24-hour count was greater than this value.

2.5.1 Secondary Data Filtering - Milton

The previous sections presented approaches used to identify suspect counts at the 15-minute count interval level. We have termed those filters the *primary* data filtering. It is also necessary to establish the maximum acceptable number of suspect 15-minute intervals in a day before the daily count is also labelled as suspect. This process, described in this section, is termed *secondary* filtering.

Because we are using the combined traffic count and not just the cyclist count, the mean volume is usually much larger than zero and thus the probability of having zero vehicles (motorized plus cyclists) in a 15-minute is relatively low. However, we would expect to observe some number of valid zero -counts in a day, especially late at night at low-volume locations when 15 minutes may plausibly pass without any traffic at an intersection. Figure 17 shows a frequency histogram of total time within a day for which the reported traffic count for an individual site is zero. The graph shows the results summed across all eight locations for the calendar year 2019 (2,920 total days).

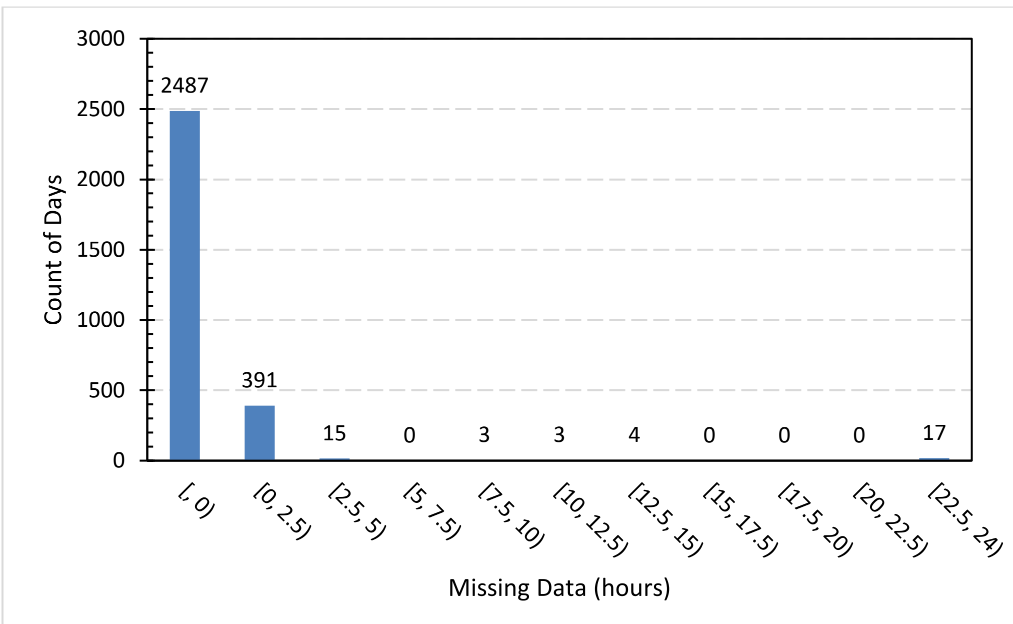


Figure 17: Frequency of 2019 days across 8 counting locations by hours of missing data

While the vast majority of days did not have any intervals with zero traffic counts, there were 391 days that experienced between one and ten 15-minute intervals (0.25 – 2.5 hours) during which no traffic was counted. A very small number of days (25) experienced between 2.5 and 15 hours of zero-counts, and no days were observed with between 15 and 22.5 hours of zero-counts.

These results might indicate that 5-hours might be a reasonable threshold for identifying erroneous daily counts from 15-minute zero counts. However, it is plausible that this threshold might result in daily counts being flagged which are correct, such as at low volume locations or on holidays (and in fact, some investigation found that most of the days with more than 5 and less than 15 hours of 15-minute zero counts were either December 25th or 26th). In order to be more certain that valid count days were not filtered, a more conservative threshold of 20 was used, as the days with greater than this amount of zero-counts could be said to be invalid with much more certainty.

2.6 Filtering Results – Pima County

The filtering methods described in the rest of this section, developed using intersection traffic monitoring data from Milton, ON, were applied to the Pima County dataset. Results are presented in Table 13, which shows the number of count days flagged by each filter.

Across data for all 34 locations from 01/01/2019 – 03/15/219, just 57 days, or 2.3% of the total number of count days, were flagged by Filters 1, 3, 4, and 6 (remember that filters 2 and 5 are not relevant for VMU data). The highest number of days flagged at any location was six, at location 76, with almost 60% of locations having one or fewer flagged days. Furthermore, only 13 of the flagged days were TMC days (Tuesday, Wednesday, or Thursday). Overall, these results indicate a high enough standard of data quality to proceed with the rest of the study.

Table 13: Summary of filtering results for the Pima County TMC dataset

Location	**Count Days**	**F1**	**F3**	**F4**	**F6**	**Total**	**%**
5	74	1	0	0	1	2	2.7%
6	75	0	1	0	1	2	2.7%
7	71	1	0	0	0	1	1.4%
9	74	1	0	0	1	2	2.7%
13	75	0	0	0	1	1	1.3%
34	75	0	0	0	1	1	1.3%
36	75	0	0	0	0	0	0.0%
38	75	0	0	0	0	0	0.0%
39	71	1	0	0	0	1	1.4%
41	71	1	0	0	0	1	1.4%
42	75	0	0	0	0	0	0.0%
43	75	0	0	0	1	1	1.3%
49	75	0	0	0	3	3	4.0%
50	71	3	0	0	0	3	4.2%
52	75	0	0	0	3	3	4.0%
53	71	1	0	0	2	3	4.2%
57	75	0	0	0	1	1	1.3%
59	71	1	0	0	0	1	1.4%
62	71	1	0	0	0	1	1.4%
69	75	3	0	0	1	4	5.3%
70	75	1	0	0	0	1	1.3%
71	75	0	0	0	0	0	0.0%
72	75	2	0	0	3	5	6.7%
73	75	2	0	0	0	2	2.7%
76	75	4	0	0	2	6	8.0%
77	75	0	0	0	1	1	1.3%
80	74	1	0	0	2	3	4.1%
81	75	1	0	0	0	1	1.3%
85	74	1	0	0	0	1	1.4%
96	75	0	0	0	1	1	1.3%
97	75	0	0	0	0	0	0.0%
100	75	0	0	0	3	3	4.0%
101	75	0	0	0	0	0	0.0%
105	71	1	0	0	1	2	2.8%
Total	2514	27	1	0	29	57	2.3%
%	-	1.1%	0.0%	0.0%	1.2%	2.3%	

2.7 Turning Movement Count Collection

An essential element of this study is to simulate the collection of one-day, 8-hour turning movement counts (TMCs), which is necessary to be able to assess the viability of using these counts as the basis of AADB estimates in actual practice. TMCs indicate directional volumes of traffic passing through an intersection during a count collection period, specific to right, left, and through movements for every intersection approach. It is very common for North American transportation agencies to manually collect turning movement counts as part of either infrequent traffic studies or regular traffic monitoring programs [16] [17]. Pima County, for example, requires morning and evening peak period TMCs to be obtained for every cross-street intersection as part of the traffic impact studies for most large developments, extrapolated to a maximum of three years [18].

Although automated methods are becoming more and more widespread, manual traffic counting remains the most common and cost-effective way of collecting TMCs [5]. This means that extra consideration needs to be given to real-world collection procedures, to best simulate them in this study. Although the specifics of TMC collection may vary by jurisdiction, general guidelines are to collect morning, mid-day, and afternoon peak period counts on a weekday during the "on-season". For Milton, it was assumed that TMCs would be made:

- During the Spring (April – June) and Fall (September – November) months
- On a weekday (Tuesday, Wednesday, or Thursday)
- During the two-hour AM peak (7:00-9:00); three-hour midday peak (11:00-2:00); and the three-hour PM peak (15:00-18:00)

The same assumptions were made for Pima County, with the exception that the "on-season" for short-term count collection was assumed to be the Winter months (January, February, March, April, May, as well as September and October) to avoid the extremely hot summer and the holidays in December.

2.8 Adverse Weather Conditions

Consideration needs to be given to inclement weather conditions, not because of their effect on cyclist activity patterns – although this topic is very widely studied – but because it is unlikely in practice that TMCs be collected on "poor weather" days, and so it makes sense to try and exclude them from the dataset for this study. Conditions which would prevent a transportation agency from collecting TMCs likely vary substantially from jurisdiction to jurisdiction.

In the absence of a jurisdictional survey, or some local knowledge which would help formulate better guidelines for establishing “inclement weather” conditions, a brief analysis of historical weather data can be used. Weather data for Milton was obtained from Environment Canada [19] and for Pima County from the U.S. National Oceanic and Atmospheric Administration [20]. Unfortunately, data from 2019 for Milton itself was not available, so data from Oakville, ON – which is approximately 20km south of Milton – was used instead, with the assumption that its proximity would make it a suitable approximate.

Thresholds were set for total daily rainfall, total daily snowfall, and mean daily temperature, based on the frequency of these events in the historical climate data, and on an estimation of what conditions could reasonably interfere with a short-term count program. Days for which the rainfall or snowfall exceeded the threshold, or for which the mean temperature was less than the threshold were designated as having inclement weather and were removed from consideration for the selection of TMC days.

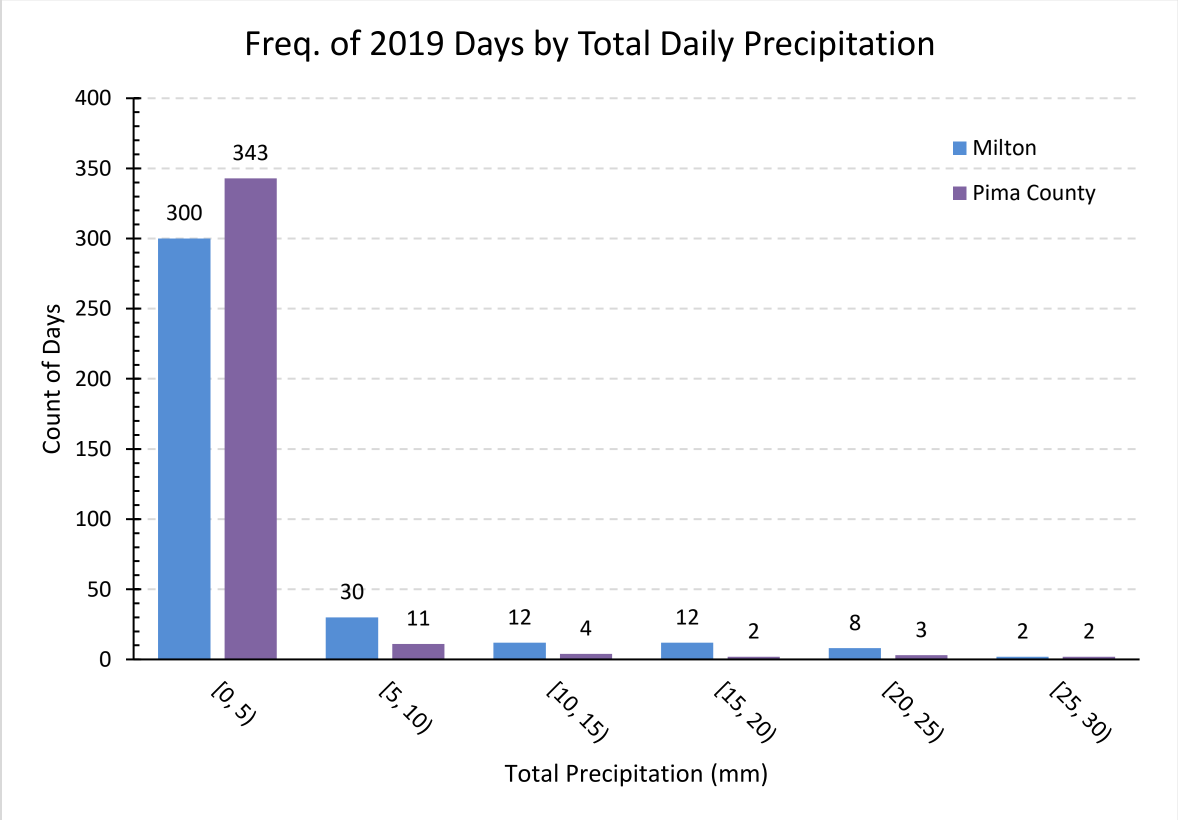


Figure 18: Frequency of 2019 days by total precipitation recorded, for Milton and Pima County

Thresholds for inclement weather were chosen somewhat subjectively, based on plots of the frequency of weather events. Figure 18 for instance shows the frequency of rainfall events recorded in Milton and Pima County for 2019. Days that received less than 10mm of rain were relatively common, while rainfall greater than this amount were relatively rare, so 10mm was chosen as the threshold for rainfall which classifies a day as "inclement". Surprisingly, given the radically different climates of the two urban areas examined here, it seemed appropriate to use the same weather thresholds; as can be seen from Figure 18, relative rainfall patterns are similar, although Milton is obviously somewhat more wet. A similar process was used to select a threshold for total daily snowfall $\geq$ 1cm, and a mean daily temperature threshold $\leq$ -5^{o} C.

2.9 Holidays

Consideration was also given to statutory holidays, both because of their impact on cyclist activity patterns, and because of their impact on TMC collection. It is unlikely in practice that one-day TMCs would be collected on a statutory holiday, so they were excluded both from the calculation of expansion factors and from the selection of one-day STCs for AADB estimation. Also considered were extended school holidays, such as spring break, which could substantially impact cyclist activity patterns. These then were also excluded from the analysis.

Datasets of holiday dates were created using information from the relevant jurisdictional school boards (the Halton District School Board calendar for Milton; and the Pima County Schools Superintendent's Office calendar for Pima County).

2.10 Conclusions

This chapter outlined the compounding process of building data filtering rules for continuous turning movement counts (TMC), or video monitoring unit (VMU) data. Although it will not be used in the remainder of this study, conventional cyclist count data from Hamilton, Ontario was useful for establishing baseline filtering logic, referring extensively to the literature on the subject. Having established filters using conventional count data, however, it was determined that they may not have been appropriate for filtering VMU data, because of the inherent differences between TMCs and conventional cyclist counts, and the technological differences in how they are collected. A VMU dataset from Milton, Ontario was used to adapt baseline filters; the adapted filtering logic was applied to a second VMU dataset from Pima County, Arizona. With the help of some researcher judgement, the filters were able to identify some obviously erroneous observations, although for both datasets the number of filtered days was very small. The outcome of this filtering process is two “clean” VMU datasets, which can be more confidently used for the AADB estimation process which will be described in subsequent chapters.

Chapter 3
Baseline AADB Estimation Methods

This chapter provides an overview of common-practice, factor-based methods for expanding short-term counts (STCs) to AADB, which are in general adaptations of equivalent methods for motor-vehicle AADT estimation. Applying these methods provides a baseline for testing the feasibility of using 8-hour turning movement counts (TMCs) for AADB estimation as part of ongoing traffic monitoring programs.

3.1 Expansion Methods

Over the past decade there has been a growing body of literature on the subject of expanding short term cyclist counts to AADB estimates, where once the field may have predominantly focused on motor vehicle traffic. From this we can identify a set of conventional methods, which are commonly applied throughout the literature but also which appear to be commonly applied in practice.

3.1.1 Annual Average Daily Bicycles

To test the estimation accuracy of expanding short-duration counts, it is necessary to first calculate "truth". As is the case for motor vehicle traffic, there are at least two general methods for calculating annual average daily bicyclists (AADB) for an intersection, road segment, or piece of active transportation infrastructure. The first is the "simple average" method [*Equation (11)*], which is the sum of 24-hour bicycle count volumes (24hBVs) over the number of days for which they are available.

$$AADB_j = \frac{1}{n}\sum_{t=1}^{365} V_{j,t} \tag{11}$$

Where

$AADB_j =$ True AADB at count location *j* for a given year

$n =$ Number of days on which cyclist counts were collected at location *j* in the year of interest

$V_{j,k} =$ 24hBV at individual count location *j* on day-of-year *t*

This method requires a full 365 days of valid 24-hour cyclist counts to be entirely accurate [21]. As is evident from the data used in this study, a full year of reliable data is rarely available, even where permanent counting equipment is installed. For this reason, an alternative, adapted AASHTO formulation [*Equation (12)*] is used in this thesis wherever it is necessary to calculate AADB. This method is more flexible, as it takes the average of counts across days-of-week within months-of-year to give monthly average daily bicyclists (MADB) values for every month in which data is available, which are then averaged to AADB. It too has data requirements, that at least one valid day of data be available for every day-of-week-of-month to be accurate, but this is less demanding than for the simple average method.

$$AADB_j = \frac{1}{n_j}\sum_{m=1}^{12}\left[\frac{1}{n_{jm}}\sum_{d=1}^{7}\left(\frac{1}{n_{jmd}}\sum_{i=1}^{n_{jmd}} V_{jmdi}\right)\right] \tag{12}$$

Where

$AADB_j =$	True AADB at count location *j* for a given year
$n_j =$	Number of months-of-year *m* for which there is at least one valid 24hBV observation
$n_{jm} =$	Number of days-of-week *d* in month-of-year *m* for which there is at least one valid 24hBV observation
$n_{jmd} =$	Number of valid 24hBV observations for day-of-week *d* in month-of-year *m* at location *j*
$V_{jmdi} =$	24hBV at location *j* on the i^{th} occurrence of day-of-week *d* in month *m*

3.1.2 Expanding 8-hour Cyclist Counts

Although a substantial amount of research has attempted to deal with STCs shorter than 24-hours, it is generally common in the literature to expand daily (24-hour) counts to estimates of AADB. This makes sense where the data is available, as expanding STCs shorter than 24-hours is conclusively less accurate than expanding a full daily count [22]. However, in consistency with real-world practice for TMC collection, it is more realistic to assume that only 8-hours of count data will be available for any given day.

There are two options for dealing with 8-hour counts in a factor-based expansion method. The first is to express the difference between 8-hour and 24-hour counts explicitly through the inclusion of a

"K-factor", which factors the 8-hour count to an estimate of 24-hour volume, which can then be expanded to AADB using traditional methods [23]; or, factors can be calculated using 8-hour counts in place of 24-hour counts so that the difference between the two is represented implicitly in the traditional expansion factors. El Esawey [23] found that including an explicit k-factor did not substantially improve estimation accuracy compared to "direct" AADB estimation from a 1-hour count [23], so for the purposes of testing the traditional methods presented in Section 3.1.3, factors will be calculated using 8-hour instead of 24-hour counts. However, a more detailed exploration of this topic will appear in Chapter 4.

3.1.3 Factoring Methods for Non-motorized Traffic

As outlined in the U.S. Federal Highway Administration's Traffic Monitoring Guide (TMG), factors representing the ratio of one number to another – the most relevant to this study being the ratios of average day-of-week or month-of-year traffic volumes to AADT, which are used to adjust short-duration traffic counts - are common to a variety of transportation engineering analysis [21]. Although general factoring procedures are not as commonly established for non-motorized traffic [24], several recurrent factoring methods can still be identified in the literature:

1. Day-of-week & Month-of-year (DOW/MOY)
2. Day-of-week-of-month (DOWOM)
3. Day-of-year (DOY)

3.1.3.1 Day-of-week & Month-of-year (Traditional)

Frequently termed the "traditional" method, the day-of-week and month-of-year (DOW/MOY) factoring approach – which may be the most common, especially in practice – is an adaptation of a corresponding method for estimating motor vehicle annual average daily traffic (AADT) from the U.S. Department of Transportation's Traffic Monitoring Guide [21]. This method involves the separate calculation of seven "day-of-week" and 12 "month-of-year" factors, for 19 total, reflecting the average weekly and monthly variation from the AADB of counts across the calibration dataset. Nordback et al [25] achieved factor-group specific mean absolute percent errors (MAPE) of between 20% and 46% when applying this method to multi-day short term counts (STCs). They found 7-days to be the most cost-efficient STC duration, as beyond this the reductions in error related to count duration began to level out. Another study by Hankey et al [24] found that the "traditional" method resulted in much higher errors than other disaggregate methods, to be discussed below, but especially when expanding from counts shorter than 7-days. The high estimation error associated with this method is usually attributed to its failure to account for short-term variability related to weather, special events, etcetera, to which cyclist counts are particularly sensitive [26] [27].

Formulas for day-of-week (DOW) and month-of-year (MOY) factors, and for calculating AADB estimates are shown in *Equations (13)*, *(14)*, and *(15)* below. Note that DOW factors are calculated from 8-hour bicycle volumes, and MOY from 24-hour bicycle volumes. Remember from Section 3.1.2 that the difference between 8 and 24-hour bicycle volume is being expressed implicitly in the methods discussed here. However, if both the DOW and MOY factors were calculated from 8-hour counts, this difference would be "double counted". So, in this case, the DOW factor can be thought of serving the dual role of expanding 8-hour counts to an estimate of 24-hour volume, and also adjusting those 24-hour estimates according to day-of-week patterns. Then, the MOY factor further adjusts the 24-hour estimate according to month-of-year patterns. It should not matter whether the DOW or MOY factor in this method is calculated from 8-hour counts, only that just one of them is. Tests were performed for both options, and an insignificant difference in estimation accuracy observed.

$$DOW_{jd} = \frac{\frac{1}{n_{jd}}\sum_{m=1}^{12}\left(\frac{1}{n_{jmd}}\sum_{i=1}^{n_{jmd}} 8hBV_{jidm}\right)}{AADB_j} \tag{13}$$

Where

$DOW_{jd} =$ Day-of-week factor for location *j* and day-of-week *d*

$n_{jd} =$ Number of months-of-year *m* for which there is at least one valid 8hBV observation for day-of-week *d* at location *j*

$n_{jmd} =$ Number of 8hBV observations collected on a day-of-week *d* in month *m* at location *j*

$8hBV_{jidm} =$ 8-hour bicycle volume at location *j* on the i^{th} occurrence of day-of-week *d* in month *m*

$AADB_j =$ Annual Average Daily Bicyclists for location *j*

$$MOY_{jm} = \frac{\frac{1}{n_{jm}}\sum_{d=1}^{7}\left(\frac{1}{n_{jmd}}\sum_{i=1}^{n_{jmd}} 24hBV_{jidm}\right)}{AADB_j} \tag{14}$$

Where

$MOY_{jm} =$ Month-of-year factor for location *j* and month *m*

$n_{jm} =$ Number of days-of-week *d* for which there is at least one valid 24hBV observation for month-of-year *m* at location *j*

$n_{jmd} =$ Number of 24hBV observations collected on a day-of-week *d* in month *m* at location *j*

$24hBV_{jidm} =$ 24-hour bicycle volume at location *j* on the i^{th} occurrence of day-of-week *d* in month *m*

$AADB_j =$ Annual Average Daily Bicyclists for location *j*

$$\widehat{AADB}_{sf^m f^d} = 8hBV_{sdm} \times \frac{1}{DOW_{f^d d}} \times \frac{1}{MOY_{f^m m}} \tag{15}$$

Where

$\widehat{AADB}_{sf} =$ AADB estimate for STCL *s*, which belongs to factor group *f*

$8hBV_{sdm} =$ 8-hour bicycle volume at STCL *s* on day-of-week *d* in month *m*

$DOW_{f^d d} =$ DOW-factor for factor group f^d on day-of-week *d*

$MOY_{f^m m} =$ MOY-factor for factor group f^m for month-of-year m

3.1.3.2 Day-of-week-of-month (DOWOM)

Day-of-week-of-month (DOWOM) factoring is a more disaggregate method than the traditional one, with 84 total factors being calculated across every day-of-week in every month-of-year. While they expected that this would better capture seasonal variation, Nordback et al [28] saw similar results using the DOWOM method to the traditional method (20% MAPE for 7-day STCs and 34% for 1-day STCs). Notably, they found significant variation in estimation error by count location, producing a 17% MAPE for 1-day DOWOM factoring at the lowest error location but 79% at the highest. Consistent with the results discussed above for the traditional method, error was significantly reduced when using 7-day STCs compared to 1-day [28]. Other studies found similar results in terms of similar MAPE compared to the traditional method, and also found significant reductions in MAPE as the STC duration increased [26]. There is also some evidence that error can be reduced using a "correcting function". Figliozzi et al. proposed a regression-calibrated model to adjust for factors like holidays, rain, and other significant variables and found that it reduced MAPE by as much as a 4% for 1-day DOWOM factoring [29]. Formulas for day-of-week-of-month (DOWOM) factors, and for calculating AADB estimates are shown in Equations *(16)* and *(17)*.

$$DOWOM_{jmd} = \frac{\frac{1}{n_{jmd}} \sum_{i=1}^{n_{jmd}} 8hBV_{jmdi}}{AADB_j} \tag{16}$$

Where

$DOWOM_{jmd} =$ Day-of-week-of-month factor for DOW *d*, month *m*, at location *j*

$n_{jmd} =$ Number of 8hBV observations collected on a day-of-week *d* in month *m* at location *j*

$8hBV_{jidm} =$ 8-hour bicycle volume at location *j* on the i^{th} occurrence of day-of-week *d* in month *m*

$AADB_j =$ Annual Average Daily Bicyclists for location *j*

$$\widehat{AADB}_{sf} = 8hBV_{sdm} \times \frac{1}{DOWOM_{fmd}} \qquad (17)$$

Where

$\widehat{AADB}_{sf} =$ AADB estimate for STCL *s*, which belongs to factor group *f*

$8hBV_{sdm} =$ 8-hour bicycle volume at STCL *s* on day-of-week *d* in month *m*

$DOWOM_{fmd} =$ DOWOM-factor for factor group *f* on day-of-week *d* in month *m*

3.1.3.3 Day-of-year (DOY)

The day-of-year (DOY), or “disaggregated” method, is generally found to be the most accurate – when a suitable amount of data is available – of the baseline expansion methods, as it in theory best accounts for seasonal and short-term variation [26] [30]. Budowski [31] achieved a 10% MAPE for seasonal average daily traffic estimation (SADT) from 14-day STCs using this method, which makes use of individual factors for every day-of-year included (up to 365 individual factors). El Esawey [30] found a MAPE of 17.5% for DOY factoring, compared to 24.5% for DOWOM. Nosal et al [26] achieved a 14% MAPE across all counting locations for 1-day DOY expansion, a very low error value considering the length of the STC and the high cyclist volumes of the locations being studied. Extended treatments, such as the separation of weekday and weekend counts [27], or of wet and dry day counts [30] may reduce estimation error. Formulas for day-of-year (DOY) factors, and for calculating AADB estimates are shown in equations *(18)* and *(19)*.

$$DOY_{jt} = \frac{8hBV_{jt}}{AADB_j} \qquad (18)$$

Where

$DOY_{jt} =$ Day-of-year factor for location *j* on day-of-year *t*

$8hBV_{jt} =$ 8-hour bicycle volume at location *j* on day-of-year *t*

$AADB_j =$ Annual Average Daily Bicyclists for location *j*

$$\widehat{AADB}_{sf} = 8hBV_{st} \times \frac{1}{DOY_{ft}} \tag{19}$$

Where

$\widehat{AADB}_{sf} =$ AADB estimate for STCL s, which belongs to factor group f

$8hBV_{st} =$ 8-hour TWT bicycle volume at STCL s on day-of-year t

$DOY_{ft} =$ DOY-factor for factor group f on day-of-year t

3.2 Factor Grouping

A barrier to accurately estimating AADB from short-duration cyclist counts is the substantial variation between count locations not only in terms of actual cyclist volumes, but also in terms of relative, temporal activity patterns. Factor-based approaches to AADB estimation normalize average cyclist counts to a ratio of AADB. This makes the absolute value of a cyclist count relatively less important than its value relative to the AADB of the location where it was collected, when calculating a set of factors.

There is significant evidence that grouping count locations with similar cyclist traffic patterns can reduce AADB estimation error [22] [23] [28]. Factor grouping of permanent count locations (PCLs) used to calculate expansion factors should minimize variation within groups and maximize variation between groups [32]. Short term count locations (STCLs) should be matched with the factor group that most closely matches their temporal cyclist patterns to generate the most accurate estimates of AADB. Inaccurate factor grouping, or incorrectly matching an STCL to a factor group, is likely to result in even higher estimation error than if factor grouping was not performed. Consider Figure 19 below, which shows the variability of average DOW factor values for the Pima County locations (each blue line represents a location). Lines showing the minimum and maximum values for every day-of-week illustrate the range of factors across all 34 locations. For example, the maximum 'Monday' factor value is ~1.2, meaning that average Monday cyclist volume at that location is 120% of AADB; the minimum 'Monday' value is 0.6, indicating that Monday volume at another location is just 60% of AADB. It is evident that using the factor calculated from the first location to estimate AADB at the second will result in a large overestimation. By grouping the second location with other, more similar locations, a factor can be calculated which more closely matches the "true" ratio of Monday volume to AADB, resulting in more accurate estimates.

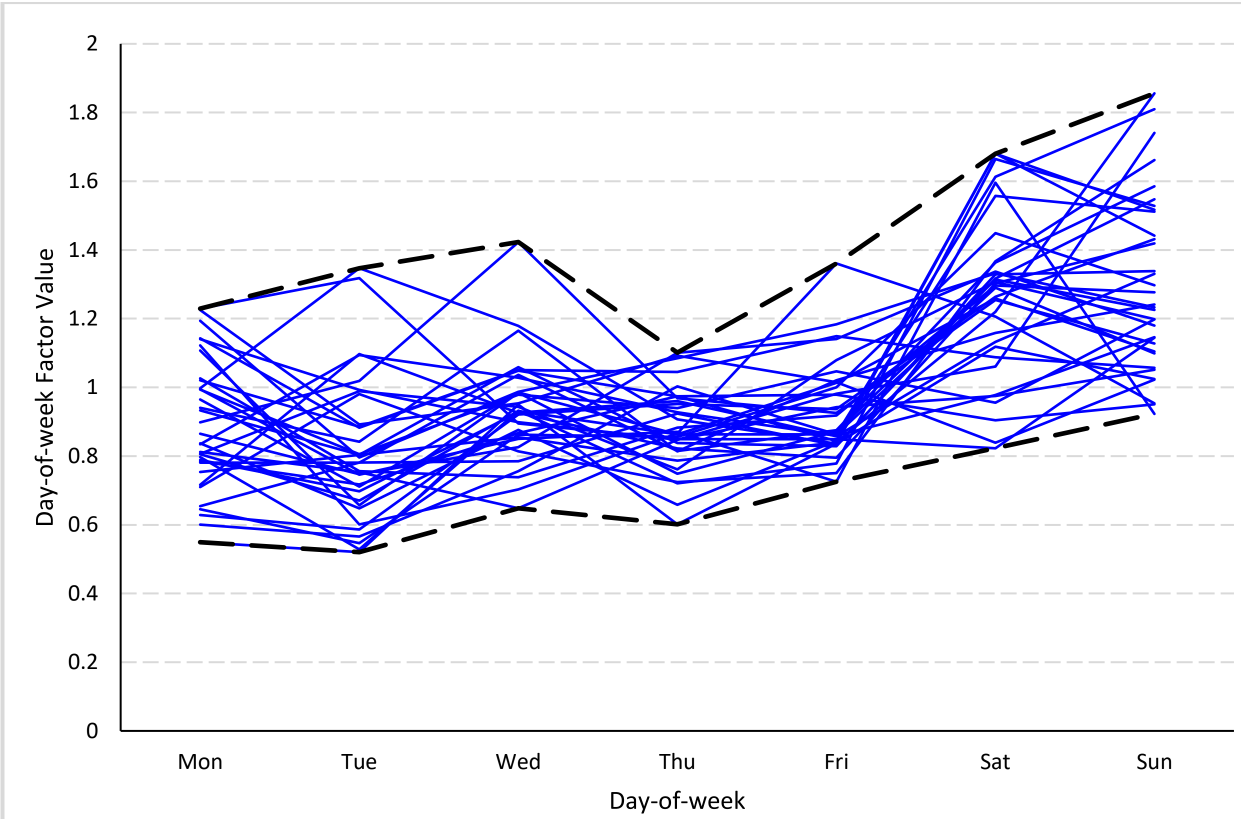


Figure 19: Day-of-week Factor Variability among Pima County Count Locations

Griswold [32] identifies two distinct approaches to establishing factors groups. The 'land use classification' approach relies on the assumption that land uses surrounding a count location will have some influence on cyclist activity. This means that locations can be grouped according to their physical attributes, and that locations within groups should have similar cyclist activity patterns. 'Empirical' factor grouping uses statistical methods to group locations based on observed count values. This approach should do a better job of minimizing variation within groups, and so should usually result in more accurate volume estimates, meaning that an empirical approach is more appropriate wherever continuous count data is available [32]. One of the challenges in the AADB estimation process is that empirical factor grouping is not possible when only 8-hours of count data is available, as is the case when expanding TMCs. Thus, it becomes important to be able to link locational characteristics to empirical factor groups; this is the subject of Chapter 5. There are a few alternative approaches to empirical factor grouping, some of which are covered below.

3.2.1 Visual Analysis

The most common method of factor grouping in practice, according to the FWHA Traffic Monitoring Guide, is visual analysis and charting of existing data [21]. It is possible to identify groups by plotting real cyclist volumes; Nordback [33] similarly divided locations into "commute" and "non-commute" groups based on a visual analysis of hour-of-day traffic plots, with "commute" locations assumed to have two distinct rush-hour peaks and non-commute assumed to have just one [25]. However, it may be easier to normalize cyclist volumes to ratio values, so that locations with different absolute ridership levels can be compared. Turner [10] plotted average hour-of-day, day-of-week, and month-of-year volumes as a percentage of AADB, using a visual inspection of these profiles to establish commuter, recreational, and mixed-purpose groups [10]. Broad, purpose-based classification scheme (i.e. commuter versus recreational, especially) is very common, as these patterns seem to often be the most easily identified in continuous cyclist data.

This method may be appropriate where only data is only available for a few locations, but it can become inconvenient as more locations are added. Furthermore, a judgement-based visual analysis might not be expected to capture all the variation present in some data; there may be pattern groups present which do not become evident without more detailed analysis. Establishing groups using day-of-year traffic profiles, for instance, might be difficult, as day-of-year plots contain a lot of noise which would make comparison difficult.

3.2.2 Temporal Indices

An extension to normalizing counts is to use temporal indices as aggregated, standardized expressions of cyclist activity patterns. The most commonly applied of these are the "Weekend-Weekday Index" (WWI) and an "AM-Midday Index" (AMI) first proposed by Miranda-Moreno et al. [11] shown in Equations *(20)* and *(21)* respectively, although additional indices such as a "Peak-Non Peak Index" (PPI) have also been suggested [27]. AMI should be calculated using only weekday counts, as the commuter patterns it is meant to capture might not appear on weekends. Definitions of which hours to use for calculating AMI vary study to study and may have to be determined on a contextual basis using researcher judgement.

$$WWI = \frac{\bar{v}_{we}}{\bar{v}_{wd}} \quad (20)$$

Where

$WWI =$ Weekend-Weekday index for a given count location

$\bar{v}_{we} =$ Average 24-hour weekend cyclist volume

$\bar{v}_{wd} =$ Average 24-hour weekday cyclist volume

$$AMI = \frac{\sum_{h=7}^{9} \bar{v}_h}{\sum_{h=11}^{13} \bar{v}_h} \quad (21)$$

Where

$AMI =$ AM-Midday index for a given count location

$\bar{v}_h =$ Average weekday hourly volume for hour *h*

To illustrate the application of temporal indexes for factor grouping, WWI was calculated for the eight Milton locations, and used to establish empirical factor groups. Common practice is to divide locations along a WWI threshold of 1, with values above 1 indicating a recreational activity pattern, and values below 1 indicating a utilitarian pattern. The resulting factor groups are summarized in Table 14.

Table 14: Milton count locations and WWI factor groups

Location	AADB	WWI	Group
4	45.69	0.79	1
5	43.09	0.86	1
9	9.21	0.91	1
14	15.40	1.18	2
15	11.36	0.94	1
17	5.57	1.05	2
22	16.71	0.68	1
25	13.55	1.22	2

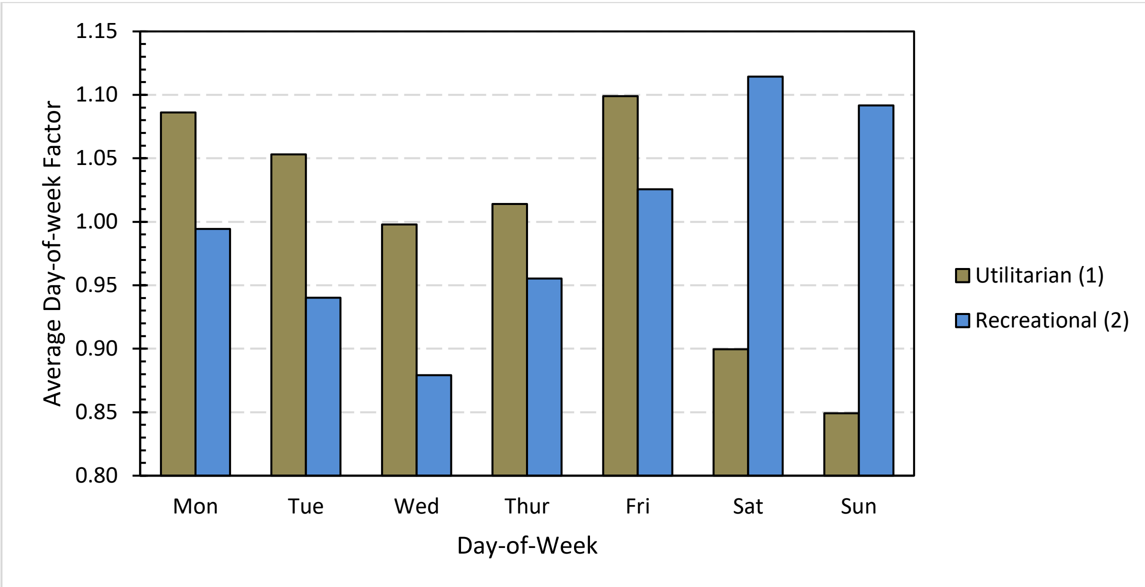


Figure 20: Average daily day-of-week volume as a ratio of AADB, by factor group

Day-of-week factors were calculated across both factor groups and plotted in Figure 20. The plot shows a distinct difference between the cyclist activity patterns across the two factor groups established using WWI, with relative volume for the "recreational" group (2) clearly peaking on Saturday and Sunday, and relative volume for the "utilitarian" group (1) being much higher on weekdays. However, although these indices are intuitive, useful metrics for establishing empirical factor groups, both AMI and WWI are of limited application for TMC data. TMCs are usually collected on Tuesdays, Wednesdays, or Thursdays, and never on weekends, meaning for a short-term count location (STCL) it would be impossible to calculate WWI. Although it would be possible to calculate AMI, there is little evidence that this value corresponds well to day-of-week and month-of-year variation in cyclist volumes, and so is not a suitable basis for factor grouping as part of AADB estimation.

3.2.3 Geographic Grouping

Grouping count locations on the basis of geographic proximity is common practice, based on the assumption that some temporal variation can be explained by both the large-scale location of a PCL (i.e. the city or metropolitan region) and the small-scale location (i.e. the location within a city or metropolitan region). In some sense, nearly every study includes implicit geographic factor grouping; it is rare that factors calculated from PCLs in one city or metropolitan region be applied to estimate

AADB at STCLs in another. This reveals a limitation in the factor-based estimation approach, which is that it may not be spatially transferrable, and in fact "…no models have been developed to…estimate missing volumes for some locations using data from other locations" [23]. This being said, some previous research has made successful use of data from across multiple jurisdictions. Nosal [26] used data from both Montreal and Ottawa to test the accuracy of AADB estimation methods with weather-based adjustment factors, finding a MAPE of just 18% across all locations for 1-day counts. However, these two cities are very similar in terms of both weather and, subjectively, "cycling culture". In cases such as this, it may be possible to group together metropolitan areas.

3.2.4 Statistical Clustering Methods

Where data is available for a large number of locations, and so manual classification becomes onerous, it may be appropriate to use statistical clustering to create empirical factor groups. In principle, this might be considered to be the "best" approach, as the result should be an "optimal" grouping schema, although how this is defined may vary. Commonly used is a *k*-means clustering algorithm using Euclidean distance as a dissimilarity measure, which aims to minimize variation within clusters and maximize the variation between groups. This algorithm has been similarly applied to temporal indices, such as WWI and AMI [27], and normalized count data [31] [25] [13] (i.e. day-of-week, month-of-year, and day-of-season data).

To illustrate this method, the 34 Pima County locations were clustered using their normalized day-of-week cyclist traffic profiles. In this case, each location was associated with a vector of seven day-of-week factor values, and the algorithm minimized the sum-of-squared distance between the members of the groups and the group means across all seven values. The results of this process are shown in Figure 3, which shows two very evident pattern groups as identified by the k-means clustering algorithm: group "1", which has relatively stable volume across all seven days; and group "2", which sees an increase in weekend volume.

For the purposes of evaluating the AADB estimation methods discussed in this thesis, a factor group cannot have just a single member. Because of this, after the algorithm was run, the resulting factor groups were checked for if any had less than two members. If any did, the "lonely" member was removed and the remaining locations re-clustered. Finally, the "lonely" member was re-assigned to the group with the nearest mean.

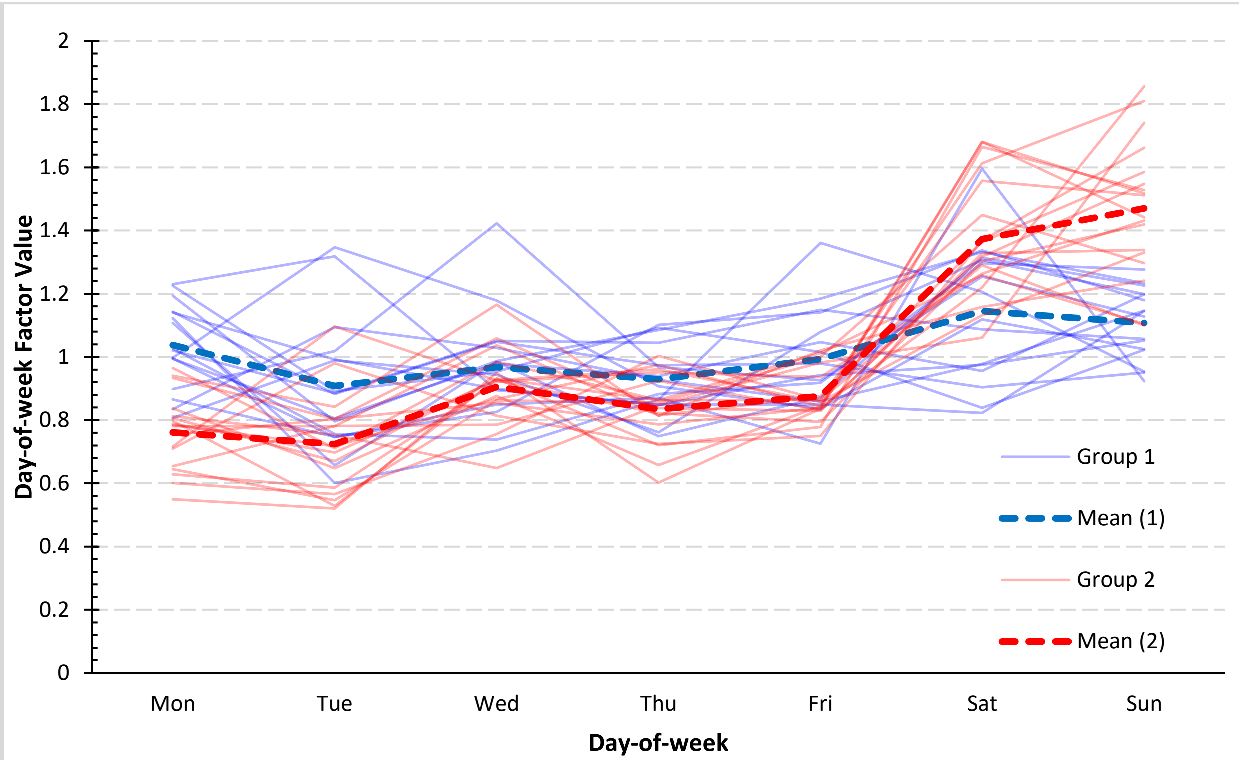


Figure 21: Day-of-week cyclist traffic profiles for the Pima County locations, grouped using the k*-means clustering method*

Ultimately, even where statistical methods are applied, some element of researcher judgement and local knowledge is almost always present in the factor grouping process. For instance, Budowski [31] used statistical clustering to "...provide guidance for grouping stations with similar characteristics together, but the final decision on how groups should be formed [was] based on the analyst's judgement". Using local knowledge, they identified two groups within their sample of count locations from Winnipeg, Manitoba: a "Winnipeg" group with more traditional seasonal traffic patterns, and a "Post-Secondary" group which saw large spikes in volume corresponding with event days at nearby stadiums.

3.2.5 Computing Intra-Group Factors

After empirical factor groups have been established, factors are calculated as an average across each group, using Equation (*22*). This equation is the same for all the factoring methods discussed in Section 3.1.3, although for the Traditional method, two sets of factors (DOW and MOY) need to be calculated for each group.

$$F_f = \frac{1}{n_f} \sum_{j=1}^{N} \alpha_j \times F_j \qquad (22)$$

Where

$F_f =$ Factor value for factor group *f*

$F_j =$ Factor value for count location *j*

$\alpha_j =$ 1 when count location *j* is in factor group *f*; otherwise, 0

$N =$ Total number of individual permanent count locations

$n_f =$ Number of individual permanent count locations in factor group *f*

3.3 Error Metrics

Several measures of AADB estimation accuracy appear commonly throughout the literature. Applying these common measures will make it easier to compare results to past findings and make conclusions about the viability of using TMCs in place of conventional cyclist counts. The most commonly used is absolute percent error (APE) (*Equation* (*23*)), which is the absolute difference between estimated and actual AADB relative to the actual AADB. This can be averaged to mean absolute percent error (MAPE).

$$APE_{j,k} = \frac{\left|AADB_{j,t}^{*} - AADB_j\right|}{AADB_j} \qquad (23)$$

Where

$AADB_k^{*} =$ Estimated AADB for location *j* on day-of-year *k*

$AADB_j =$ Observed AADB for location *j*

$APE_{j,k} =$ Absolute percent error of AADB estimation for location *j* on day-of-year *k*

Although it is useful to standardize error measures for the sake of comparison, it should be noted that APE tends to be positively biased for low-volume locations. Say an AADB estimate of 2 was produced for a location with a true AADB of 1; the APE for this estimate would be 100%, even though it was within only one cyclist of the correct value. Now compare this to a second estimate of 51, produced for a location with a true AADB of 50. The APE for this estimate is 2%, much lower than the APE for the previous estimate, even though in terms of actual absolute error both were equally accurate. This means that when MAPE is calculated as a summary metric across all study locations, it will be weighted towards the low volume locations where average percent errors tend to be highest.

To somewhat avoid this problem, a second summary metric is proposed, a "volume-weighted mean absolute percent error" (VWMAPE), shown in *Equation (24)*. This takes the sum of mean absolute error across all locations and divides it by the sum of AADB values, which should provide a less biased expression of average error.

$$VWMAPE = \frac{\sum_{j=1}^{N} MAE_j}{\sum_{j=1}^{N} AADB_j} \qquad (24)$$

Where

$VWMAPE =$ Volume weighted mean absolute percent error across all count locations

$MAE_j =$ Mean absolute estimation error for location *j*

$AADB_j =$ True AADB for location *j*

$N =$ Total number of individual count locations

3.4 Estimating AADB from TMCs Using Baseline Methods

In actual practice, factors are calculated from data collected at permanent count locations (PCLs), and then applied to short-term counts (STCs) which are collected at short-term count locations (STCLs) for which AADB is not known, meaning that there is no way of evaluating estimation accuracy. To test the baseline methods discussed in this chapter, the Milton and Pima County TMC datasets will be used to both calibrate sets of factors, and to simulate the collection of STCs, allowing the methods to be evaluated based on the accuracy of the estimates they produce. This workflow is described by Figure 22. First, sets of expansion factors are calculated for every location in the filtered input dataset (1), as well as true AADB (2). Then, locations are empirically grouped, based on their factor-vectors established in the previous step (3). Note that for the Pima County results, a *k*-means clustering approach is taken; for the Milton results, locations are grouped based on their WWI values. One by one, locations are removed from the dataset, and average factors are calculated across the other locations in their group (4). Repeating this process will provide every location with an associated set of factors calculated as if it were a STCL. Finally, AADB estimates can be generated for every day of data at every location in the input dataset (5), treating each day as if it were an STC. These estimates can be compared to the true AADB calculated in the first step (6).

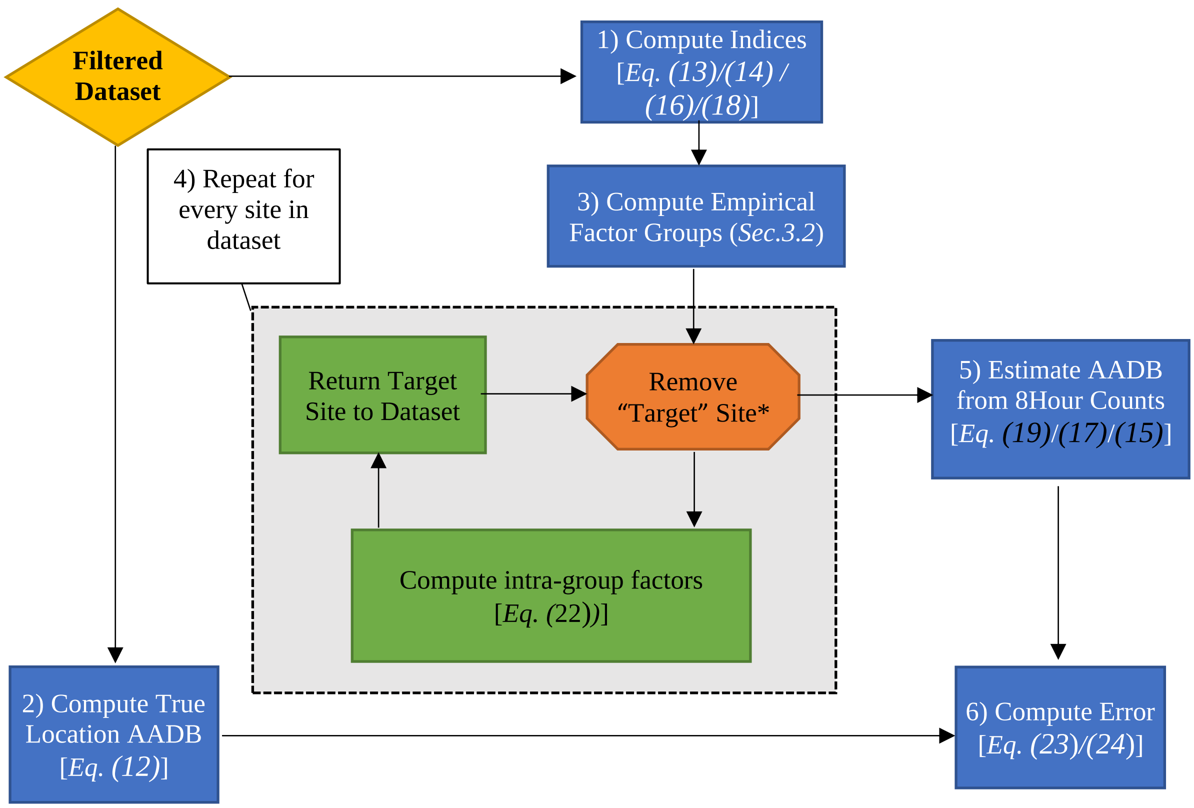


Figure 22: AADB Estimation Workflow (adapted from Nordback et al [28])

The DOW/MOY, DOWOM, and DOY expansion methods were applied to the Milton and Pima County TMC datasets. AADB estimation results are presented and discussed below.

3.4.1 Milton, ON

A summary of the baseline results for Methods 1 – 3 can be found below, using 8-hour TMCs, for Tuesday-Wednesday-Thursdays of spring and summer months of 2019. Both "grouped" and "ungrouped" trials were conducted. For grouped trials, the WWI factor grouping logic outlined in Section 3.2.2 was followed, so that results were collected separately for a "utilitarian" group (1) and a "recreational" group (2). For the ungrouped trial, the STCLs were treated as if they were all part of a single factor group.

Remember that in order to best simulate real-world counting conditions, only counts which met the following criteria are included in the calculation of estimation error:

- Counts greater than 0
- Counts which fall into the Spring (April – June) or Fall (September – November) seasons
- Counts on mid-week workdays (Tuesday, Wednesday, or Thursday)
- Counts from days which meet the weather criteria of experiencing less than 10mm of rainfall, less than 1cm of snowfall and having an average temperature greater than -5o C

A summary of estimation results is shown below in Table 15. Of the three baseline methods discussed in this chapter, the disaggregate DOY method produced the most accurate estimates from the Milton TMC data in the *ungrouped* trial, although the DOWOM method was by far more accurate in the *grouped* trial. The 'traditional' DOW/MOY method produced both the highest error, and somewhat surprisingly, the highest spread of estimation results; standard deviation of MAPE for the *ungrouped* trial was 71.3%, indicating that estimation accuracy varied greatly day to day. In principle, the DOY approach should be most susceptible to short-term variation, resulting in less consistent estimates. Aggregating counts to DOW and MOY averages is supposed to eliminate some of that variation.

Table 15: Summary of Baseline AADB Estimation Results for Milton TMC Data

	Method	*DOW/MOY*		*DOWOM*		*DOY*	
	Stat	**Avg.**	**SD.**	**Avg.**	**SD.**	**Avg.**	**SD.**
Ungrouped	APE	62.3%	71.3%	47.8%	51.9%	46.9%	55.2%
	AE	13.2	20.3	9.9	15.1	9.3	13.6
Grouped	APE	69.5%	89.8%	54.8%	66.2%	64.3%	138.7%
	AE	15.6	28.6	11.7	21.3	12.9	28.1
Diff.	APE	+7.1%	+18.5%	+7.0%	+14.3%	+17.4%	+83.5%
	AE	+2.4	+8.3	+1.8	+6.2	+3.6	+14.5

Although we would expect the *grouped* results to be on average more accurate, it would seem that this is not the case, with MAPE being higher in the *grouped* trial than in the *ungrouped* trial for all three baseline methods. The spread of estimation results also increased substantially under the *grouped* scenario, so estimates were both less accurate and less consistent. It may be that the positive effect of calibrating expansion factors from groups of locations with similar activity patterns is offset by the effect of having fewer locations per factor group, which is known to increase estimation error [28]. There have been several attempts to establish minimum factor group sizes in the literature, with Nordback et al. [28] recommending a minimum of eight locations per group where day-of-week-of-month factors are used. Although this recommendation is met in the *ungrouped* trial, the *grouped* trial sees the factor group sizes fall to five (Group 1) and three (Group 2), which may have had an adverse impact on results.

This line of thinking is further supported by examining the *grouped* results by location, shown in Table 16. Estimation error across Group 1 is close to the same as results from the *ungrouped* trial, but error for Group 2 is extremely high; MAPE for the DOY method across just Group 2 locations was 92%, with a standard deviation of 203%. These results would not be acceptable in practice. For any given one-day TMC at location 14 for instance, the resulting AADB estimate might be anywhere between 0 and ~70. In fact, the highest DOY estimate produced for location 14, on 3/4/2019, was 295. Evidently, more locations may be needed to calibrate a set of factors for accurate AADB estimation, and estimation months may need to be further limited to avoid the high-variation shoulder season.

Table 16: Milton AADB Estimation Results, by Location

		Method:	*DOW / MOY*			*DOWOM*			*DOY*		
Loc	**AADB**	**Group**	**Avg Est**	**MAPE**	**SD of APE**	**Avg Est**	**MAPE**	**SD of APE**	**Avg Est**	**MAPE**	**SD of APE**
4	45.5	1	63.1	54.9%	49.5%	51.1	36.4%	37.8%	53.0	35.1%	43.3%
5	42.9	1	83.4	108.0%	138.1%	67.1	78.5%	105.5%	67.8	77.1%	108.1%
9	9.2	1	12.4	54.5%	67.8%	9.6	40.6%	43.0%	10.0	39.6%	42.1%
14	15.1	2	24.7	99.9%	152.2%	21.1	80.8%	105.2%	29.3	138.9%	332.7%
15	11.2	1	11.8	45.4%	35.6%	10.0	41.3%	29.5%	9.6	35.3%	24.7%
17	5.4	2	6.4	58.1%	66.7%	5.9	48.7%	43.5%	6.4	60.9%	84.4%
22	16.5	1	22.3	65.4%	44.7%	17.7	44.9%	31.4%	17.6	44.2%	33.9%
25	13.4	2	17.9	62.0%	47.4%	16.5	62.5%	51.6%	18.1	76.2%	76.3%
Average:		1	38.6	65.6%	76.6%	31.1	48.4%	57.0%	31.6	46.2%	58.5%
		2	16.3	73.3%	99.7%	14.5	64.0%	72.2%	18.0	92.0%	203.0%

Because a full year of continuous data is available for the Milton locations, it is possible to plot how estimation error changes across TMC months (April – June, September – November). MAPE across all eight locations, by month, for the "ungrouped" results is shown in Figure 23. From this we can see that, although error in the shoulder months (April, May, and November) is unacceptable relative to some of the results found in the literature – with error for the DOW/MOY method in April being on average over 100% - AADB estimates for the "on-season" months were generally much more accurate. This is consistent with a common recommendation from the literature, that STCs should be collected during the months when count variation is the lowest in order to minimize estimation error [25] [28], which in Milton should be in June through September.

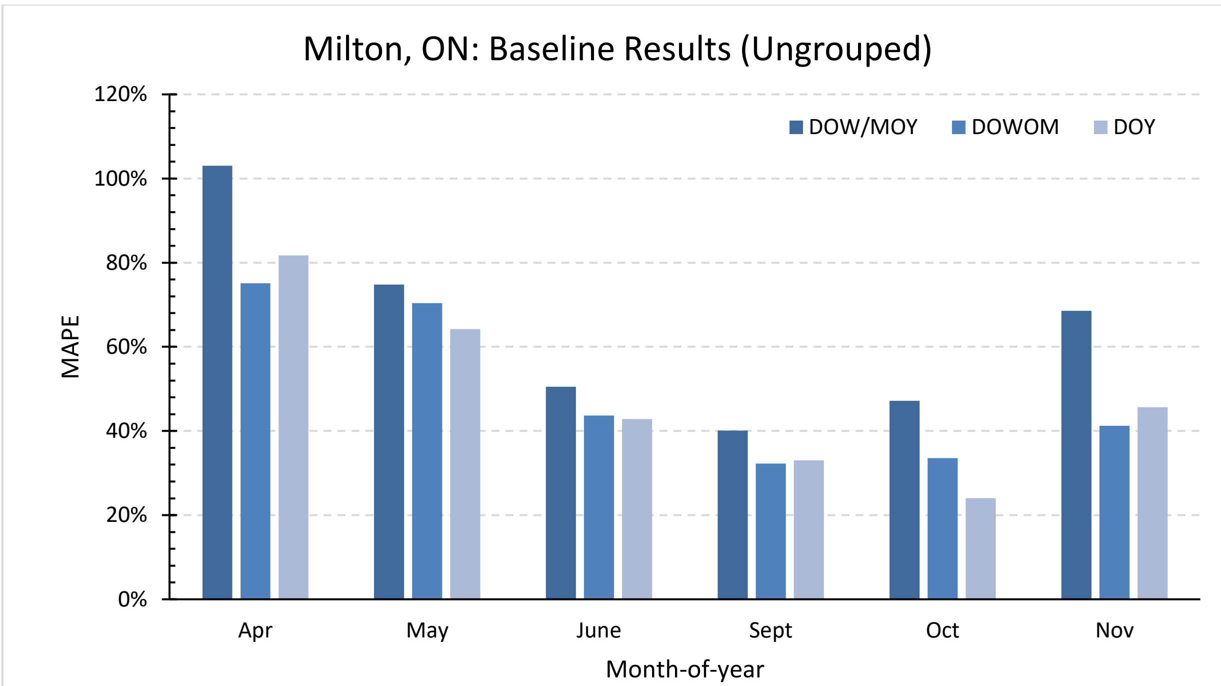


Figure 23: MAPE across Milton Locations for "Ungrouped" AADB estimates, by month

3.4.2 Pima County, AZ

Baseline methods were used to generate AADB estimates for 34 Pima County count locations (the selection of which is discussed in detail in Chapter 2). As was also discussed in Chapter 2, the Pima County dataset was limited by the emergence of the Covid-19 pandemic in early 2020. Under the assumption that the pandemic drastically impacted prevailing travel patterns, only data from between January 1st and March 15th was used in this analysis. Additionally, as was done for the Milton results in order to best simulate real-world TMC collection, target count days were limited to Tuesdays, Wednesdays, and Thursdays, which did not meet any of the "inclement weather" conditions, which were not holidays, and on which the 24-hour cyclist count at a given location was greater than zero.

The result was 30 (or occasionally fewer) STC days – and thus 30 AADB estimates – for each of the 34 count locations. Average estimation results across all locations are shown below in Table 17.

Table 17: Summary of baseline AADB estimation results for Pima County

Method	Groups	MAPE	SD of APE	Weighted MAPE	ANOVA (p)
DOW / MOY (Traditional)	1	44.4%	37.8%	38.8%	0.819
	2	43.4%	34.8%	39.4%	
	3	42.2%	33.7%	39.8%	
	4	41.8%	33.6%	39.5%	
DOWOM	1	43.9%	37.5%	38.4%	0.983
	2	42.8%	35.0%	38.9%	
	3	42.3%	38.6%	38.6%	
	4	41.8%	38.7%	38.3%	
DOY (Disaggregate)	1	41.7%	39.2%	35.4%	0.998
	2	41.8%	46.1%	35.2%	
	3	41.3%	45.5%	35.8%	
	4	41.2%	41.3%	36.0%	

Empirical factor groups were established using K-means clustering on factor values, so that separate grouping schema were established for each of the three expansion methods. When the number of groups is "1", this indicates that all locations were placed in a single factor group, essentially meaning that they are ungrouped.

The disaggregate DOY method produced the lowest average error, both in terms of real MAPE and weighted MAPE, however it also had the highest average standard deviation of estimation results (39.2% for 1 factor group, 41.3% for 4 factor groups), indicating that it may be the least reliable method. The lowest average error for the other two methods was produced by the Traditional method using a 4-group factor grouping schema. This method also had the lowest average spread of estimation results (SD of APE = 33.6%), indicating that it may be the most reliably accurate of the baseline methods.

An analysis of variance (ANOVA) was used to test the association between the number of factor groups and average estimation error, across the three methods tested. In each case, ANOVA results fail to show that there is any significant difference between the mean of estimation error for any of the four grouping schemas. Manually inspecting MAPE values, it seems as though increasing the number of factor groups at least slightly improved estimation results for all three cases, but this conclusion is not statistically supported. This indicates that, for the Pima County sample, performing factor grouping for AADB estimation using traditional expansion methods does not produce substantially different results than acting as if all locations are part of a single factor group.

Weighted MAPE for all three methods either stays consistent or slightly increases, indicating that any improvement in estimation results is occurring in low volume, high error locations. As well, the standard deviation of APE seems to increase with the number of factor groups for the DOWOM and DOY methods, meaning that while estimates on average may become more accurate, the spread of estimates also increases around that average.

To further examine the effect of factor grouping on AADB estimation using the Traditional method, location-specific results are provided in Table 18, for both the 1-group and 4-group trials. As mentioned before, low volume locations tend to have higher APE, and this is evident from the disaggregate results; MAPE for 1-group results across locations with AADB<10 is 60.9%, and across locations with AADB≥10 is 38.5%. Furthermore, nearly all the improvement in MAPE observed by increasing the number of factor groups to 4 occurs for just a handful of low volume locations; locations 62, 57, 28, 80, 9, 85, 38, and 105 all have AADB less than 13 and saw at least a 10% improvement in MAPE from the 1-group to 4-group scenarios. Nearly all other locations actually saw an increase in MAPE. Furthermore, nearly all the improvement occurred for locations in DOW-group #3; the average difference in MAPE for these locations was -18%, while for groups 0, 1, and 2 the average difference was -2.1%, +3.3%, and -1.5% respectively. This implies that using more disaggregate factor grouping did have a substantial, positive effect on estimation accuracy, but only for select locations.

3.4.3 Conclusions

Separate AADB estimates were generated using continuous TMC data from Milton, ON, and Pima County, AZ, using three different factor-based methods, which we have called the “baseline” methods. Although overall estimation error for Milton was very high (even relative to what is to be expected from the literature when expanding from low-volume counts less than 24-hours in length [34]), further investigation found that error was substantially lower in the “on-season” summer and fall months (June, September, October) than in the “shoulder” months (April, May, November) when day-to-day cyclist activity is most variable. Although in practice it may not be possible to be selective about when a TMC is collected, it would be best to try and estimate AADB using counts from on-season months. The number of Milton locations for which data was available (eight) made it difficult to evaluate the efficacy of baseline factor-grouping methods. Because factor grouping is such an important part of the AADB estimation process, subsequent chapters will rely on the Pima County dataset instead, for which data was available at 34 locations.

AADB estimation results for Pima County fell within a very wide range, with location-specific MAPE ranging from 22% at the lowest to 103% at the highest for an un-grouped scenario using the Traditional method, although estimation error was inflated at low-volume locations. Statistical analysis was not able to show a significant effect of empirical factor grouping, but a noticeable decline in estimation error was observed for an increased number of factor-groups.

For both datasets the most accurate of the baseline methods was the disaggregate DOY method, although it suffered somewhat from a high spread of estimates. The Traditional DOW/MOY method, for Pima County, was not substantially less accurate than the other two methods. Furthermore, we expect that it should be possible to model empirical factor groups using land use characteristics for this method, while it may be more difficult for the other two, as generalized relationships between temporal activity patterns and land uses should become more apparent at a more aggregate level. For this reason, the next chapter will attempt to make improvements upon the Traditional method which address several of the problems related to expanding 8-hour TMCs.

Table 18: AADB Estimation Results for Pima County Locations (Trad. Method, 1-group vs. 4-groups)

		Factor Group		1-Group				4-Group				
Loc	*AADB*	*DOW*	*MOY*	*Avg. Est.*	*MAE*	*SD of MAE*	*MAPE*	*Avg. Est.*	*MAE*	*SD of MAE*	*MAPE*	*MAPE Diff.*
7	5.6	0	0	6.2	3.0	2.3	53%	6.4	3.0	2.3	55%	+1%
96	10.8	0	0	13.5	5.5	3.8	51%	14.0	5.7	4.3	53%	+2%
77	14.8	0	0	15.2	4.3	3.6	29%	15.7	4.4	3.5	30%	+1%
41	34.9	0	0	29.8	13.8	9.5	40%	30.3	13.8	9.5	40%	+0%
5	17.6	0	1	12.2	8.8	7.4	50%	13.1	8.5	7.7	48%	-2%
62	2.5	0	2	4.6	2.6	2.3	103%	4.0	1.7	1.9	68%	-35%
53	8.8	0	3	10.0	2.9	2.3	33%	10.9	3.6	2.6	41%	+8%
42	19.6	0	3	17.3	4.5	3.3	23%	18.4	4.5	2.9	23%	-0%
36	22.8	0	3	22.6	6.3	4.6	27%	24.4	7.1	5.1	31%	+4%
34	25.0	0	3	22.1	7.8	4.1	31%	23.5	7.8	4.2	31%	+0%
69	11.9	1	0	11.9	5.1	3.3	43%	12.9	5.4	3.7	45%	+2%
52	14.3	1	0	12.7	6.9	4.6	48%	13.8	7.2	5.4	50%	+2%
43	22.5	1	0	23.7	8.6	6.5	38%	25.7	9.4	7.0	42%	+4%
72	18.5	1	1	19.3	6.2	4.5	34%	22.6	8.1	5.0	44%	+10%
97	24.0	1	1	16.8	10.5	6.5	44%	19.0	9.9	6.2	41%	-2%
100	24.2	1	1	24.6	8.6	6.5	35%	29.0	10.9	9.1	45%	+10%
81	13.2	1	3	10.8	4.7	3.4	35%	12.2	4.9	3.7	37%	+2%
76	15.7	1	3	15.0	5.0	3.2	31%	17.2	5.3	4.3	34%	+2%
101	22.0	1	3	16.1	7.8	5.6	35%	17.9	7.8	5.6	35%	-0%
73	31.6	1	3	26.1	12.1	8.3	38%	29.4	12.5	9.7	40%	+1%
71	35.5	1	3	36.3	13.4	10.7	38%	41.6	15.4	14.2	43%	+6%
50	9.8	2	0	11.3	4.1	3.9	42%	10.8	3.7	3.4	37%	-5%
13	13.2	2	0	16.2	6.0	5.0	45%	15.8	5.5	4.5	42%	-3%
70	11.9	2	1	12.7	4.7	3.6	39%	13.0	4.7	3.1	39%	-0%
59	20.7	2	1	20.1	4.6	3.5	22%	20.8	5.9	4.0	28%	+6%
49	74.5	2	1	68.2	24.9	19.4	33%	69.9	27.9	16.2	37%	+4%
57	5.5	2	2	6.5	4.1	4.2	75%	5.7	3.3	2.5	60%	-15%
39	8.7	2	3	7.8	4.8	4.0	55%	7.7	4.8	4.3	54%	-1%
6	20.2	2	3	22.3	7.4	5.2	37%	22.8	7.8	6.2	39%	+2%
80	7.7	3	0	12.7	6.2	4.2	80%	9.9	4.1	2.5	53%	-28%
9	10.2	3	0	13.7	5.5	4.5	54%	10.3	3.4	3.0	34%	-20%
85	12.9	3	0	15.8	7.5	5.4	58%	11.7	5.6	3.8	43%	-15%
38	5.4	3	1	7.2	2.7	2.7	49%	5.7	1.9	1.5	35%	-15%
105	4.8	3	3	5.9	2.7	2.1	56%	4.5	2.1	1.4	43%	-12%

Chapter 4
Refining the Traditional AADB Estimation Method

This chapter will draw upon the results and conclusions from Chapter 3 to develop an update to traditional factor-based AADB estimation approaches – which have typically been developed using "conventional" cyclist counts – to make them more suitable to one-day, 8-hour turning movement count (TMC) data. Issues posed by using TMCs for AADB estimation will be defined in a 4-part problem statement. Statistical analysis will be used to TMC-appropriate factor aggregations, for Tuesday-Wednesday-Thursday (TWT) day-of-week factors, and for a newly introduced 8-hour/24-hour "K"-factor. Formulas will be defined for the proposed method, and the method will be evaluated using the Pima County dataset.

4.1 Problem Statement

The 8-hour TMC expansion process can be expressed in four distinct steps:

1. Select a factoring method and calibrate a set of factors for every permanent count location (PCL)
2. Group PCLs with similar temporal cyclist volume patterns into 'factor-groups', using the factors calculated in step (1).
3. Identify which factor-group a short-term count location (STCL) belongs to.
4. Apply factors to an 8-hour count from the STCL to generate an AADB estimate for that location.

The difficulty in proposing a method for expanding cyclist TMCs is that the method must work with only the minimum amount of data which might be collected as part of a TMC study, which for any given location is just the eight "peak" hours of a single weekday. Working backwards through the four steps further elucidates this problem.

Starting at step (4) with the collection of short-term counts (STCs), which in this study are 8-hour TMCs: In general, the longer the duration of an STC, the higher the accuracy of the AADB estimate produced from that STC [26] [28]. Nosal [26]found that mean absolute percent error (MAPE) could be reduced by one-third by increasing the length of a STC from one to five days [26]; Nordback [35]found furthermore that MAPE for 8-hour counts was four percent higher than for 24-hour counts, and another 16% higher for 2-hour weekday counts [35]. Cyclist traffic volume is subject to extremely high day-to-day variation, which contributes to high estimation error. Collecting a larger sample (i.e. a STC collected over a longer time period) can help to mitigate this effect, but this is not necessarily possible with TMCs.

Next, for step (3), it becomes easier to correctly identify the factor group to which a STCL belongs the longer the duration of an STC. With seven contiguous days of data, the day-of-week profile can be matched to that of the appropriate factor group. Even with just two full days of data, one weekday and one weekend, a weekend-weekday index (WWI) can be calculated, which forms a reliable basis for minimizing intra-group variation [11]. However, being limited to a single day of data poses a challenge. Similar to WWI, a 'morning-midday index' (AMI) can be calculated as the ratio of morning to midday traffic, and this can be used to group locations with less than 24-hours of counts [35] [11]. However, AMI values may not consistently reflect day-of-week or month-of-year traffic patterns, which are both more important to AADB estimation under any of the traditional factoring approaches. If this is the case, then an AMI-based factor-grouping approach will not be adequate. Some other, more robust approach may be needed.

Third, in respect to step (2), a grouping schema should be chosen which is most conducive to step (3). We would assume that it will be more difficult to identify STCL group membership in a more disaggregate grouping schema, as a more complex temporal pattern will be more difficult to model using only a single day of count data. Although a more disaggregate schema may improve estimation accuracy by capturing more of the random variation in cyclist counts, we are limited in our approach to those schemas which we are able to model from an 8-hour count.

Finally, the factoring approach chosen for step (1) is, by extension, limited in the same way as the factor grouping approach in step (2). Intuitively, it makes sense that the factor types used to establish groups should also be used to generate AADB estimates; so if Day-of-week-of-month (DOWOM) factors are to be used, for instance, then DOWOM factors should be used to establish empirical factor groups (i.e. factor groups established from PCLs using statistical methods, such as K-means clustering), so that intra-group variation between locational DOWOM profiles is minimized. Traffic patterns for a count location at one temporal scale may not have any relationship to traffic patterns for the same location at a different temporal scale. So, a group of locations with very similar day-of-week traffic profiles may have very different month-of-year patterns. Day-of-week (DOW) factors calculated across this group will produce more accurate estimates than month-of-year (MOY) factors, as variation has been minimized at the DOW level but may still remain high at the MOY level. For this reason, the factoring approach chosen for step (1) should match the factor grouping approach chosen for step (2), and so is subject to the same constraints from all previous steps.

Given these points, some further points might be made with regard to adapting traditional factor-based approaches to AADB estimation for TMCs. First, the factoring method chosen should be as aggregate as possible while still providing accurate estimates, to best enable the correct identification of empirical factor group membership for STCLs for which there is only 8-hours of count data. This would indicate that the most appropriate method for turning movement counts is the “Traditional” Day-of-week/Month-of-year (DOW/MOY) approach. Under this approach, day-of-week and month-of-year variations are expressed separately, meaning that they should be easier to model using locational land-use characteristics. Second, and further along these lines, an additional factor should be added to express the difference between average 8-hour and 24-hour counts, which we will call the “K” factor. One would expect the locational characteristics which are correlated with average K-factor values to be different than those correlated with day-of-week traffic profiles. Therefore, for the purposes of modelling STCL factor-group membership, it makes sense to separate K and DOW factors. Finally, in order to minimize the intra-group variation between locations in terms of all three of their K, DOW, and MOY factor profiles, empirical factor grouping should be performed at three different levels, one for each factor type.

4.2 K-factor for 24-hour Count Estimation

Following the recommendations outlined in Section 4.1, this section will explore the viability of using a “K”-factor - calculated as the ratio of average daily 8-hour to average daily 24-hour cyclist count volume - to compute estimates of 24-hour volume from TMCs. The primary purpose of introducing a K-factor here is to separate peak-hour temporal patterns from day-of-week and month-of-year patterns, to enable more accurate factor-group modelling. However, it may have additional benefits in terms of estimation accuracy. 8-hour TMCs tend to be lower volume than conventional cyclist counts, which is detrimental to estimation accuracy. Conventional cyclist counts are generally taken on permanent cyclist infrastructure, where volumes might be expected to be both higher [36] and more consistent.

The existing literature recommends that STCs of less than 24-hours be avoided [28] [33] [37], as estimation accuracy is less reliable without a full day of data. Working from TMCs means that this recommendation cannot always be met. To help address this challenge, a "K-factor" can be calculated as the ratio of 8-hour to 24-hour cyclist volume, as proposed by El Esawey [23]. This value acts as an additional expansion factor which brings 8-hour counts to an estimate of 24-hour volume; the 24-hour count can then be expanded using traditional factors to an estimate of AADB. The formula for a one-day, location-specific K-factor is shown in Equation (*25*).

$$K_j = \frac{8hBV_j}{24hBV_j} \tag{25}$$

Where

$K_j =$ K-factor for location *j* for a given day

$24hBV_j =$ 24-hour bicycle volume at location *j* for a given day

$8hBV_j =$ 8-hour bicycle volume at location *j* for a given day

4.2.1 Nested ANOVA of K-factor Values

K-factors were calculated for every available count day at the 34 Pima County locations, using the formula in Equation (*25*). A nested analysis of variance (ANOVA) was conducted on the dataset to determine the proportion of variance in K-factor values explained by several different variables. El Esawey [23] found that 62% of variability in K-factor values could be "…attributed to the variation from weekends to weekdays". However, this difference is not relevant to a study of TMCs, which are collected on Tuesday, Wednesday, or Thursdays (TWT) by convention. Accordingly, the analysis was conducted using only valid TWT counts, from between January 1, 2020 and March 15, 2020.

The ANOVA model equation is shown by *Equation* (*26*).

$$K_{jmd} = \mu + \alpha_j + \beta_{jm} + \gamma_{jmd} + \epsilon_{jmdk} \tag{26}$$

Where

$K_{jmd} =$ K-factor value for location *j* on a day-of-week *d* in month *m*

$\mu =$ Overall mean *K*-value

$\alpha_j =$ Effect for j^{th} count location (j=1, 2, 3, …, 34)

$\beta_{jm} =$ Effect for m^{th} month (m=Jan, Feb, Mar) at j^{th} count location

$\gamma_{jmd} =$ Effect for d^{th} day-of-week (d=Tue, Wed, Thu) of m^{th} month at j^{th} count location

$\epsilon_{jmdk} =$ Unexplained variance

Table 19: Summary of nested ANOVA results for K-factor values

Variance Component	DF	Sum Sq.	Mean Sq.	F	Pr(>F)	η^2	ω^2
location	33	5.530	0.168	4.653	1.96E-15*	13.1%	0.103
location : month	68	3.641	0.054	1.487	0.009*	8.6%	0.029
location : month : DOW	204	8.048	0.039	1.095	0.202	19.1%	0.019
within DOWs (residuals)	705	24.996	0.035	-	-	59.2%	-

**Significant at a 95% level of confidence / ^Negative ω^2 value*

Results in Table 19 indicate that the mean of K-factor values are *not* significantly different between Tuesdays, Wednesdays and Thursdays within a given month at a given location. The month-of-year at a given location was found to be a significant variable at a 5% confidence level, however the proportion of variance explained (as expressed by η^2) by the month-of-year was found to be just 8.6%, a much smaller proportion than either location or day-of-week.

While exploring factor data by location, it was noted that the relative day-to-day variation of K-factor values tended to be very high at low volume locations. To avoid the influence of these locations, a second set of ANOVA results were obtained from the 25 locations where the average 8-hour TWT volume was greater than ten. These results are shown in Table 20.

Table 20: Summary of nested ANOVA results for K-factor values (AA8hB ≥ 10)

Variance Component	DF	Sum Sq.	Mean Sq.	F	Pr(>F)	η^2	ω^2
location	24	2.876	0.120	4.543	2.02E-11*	13.3%	0.104
location : month	50	1.313	0.026	0.996	0.485	6.1%	0.000
location : month : DOW	150	3.779	0.025	0.955	0.627	17.5%	0^
within DOWs (residuals)	517	13.636	0.026	-	-	63.1%	-

**Significant at a 95% level of confidence / ^Negative ω^2 value*

The second set of results shown in Table 20 indicate that, excluding low volume locations, neither month-of-year nor day-of-week have significant effects on the ratio of 8-hour to 24-hour volume at the Pima County locations. Only ‘location’ was found to be a significant variable, as would be expected given the obvious differences in temporal count patterns between count locations.

Eta-squared (η^2) is included to provide an estimate of the proportion of variance explained by each variable. For high-volume locations, approximately 13% of the variation in the ratio of 8-hour to 24-hour cyclist volume can be attributed to location, just 6.1% to MOY, and 17.5% to DOW, leaving 63.1% of the variation unexplained. This would suggest that DOW is a more powerful predictor than location. However, η^2 is known to be positively biased [38], so a less biased expression of effect size, omega-squared (ω^2) [38], was also computed. This shows the effect size of month and day-of-week to be essentially zero, while only location has any statistical effect ($\omega^2 = 0.104$).

Given these model outcomes, we can be comfortable calculating a single K-factor for every location, rather than calculating DOW or MOY specific K-factors, since the mean ratio of 8-hour to 24-hour volume across DOWs or MOYs has not been shown to be significantly different. This may not always be the case, as it is expected that relative temporal variation differs substantially by geographic region. For instance, cyclist traffic may be consistent between January, February, and March in Pima County, where the weather for these months is very stable. But the cyclist traffic patterns may change radically between TMC months in Milton, ON – where short-term weather patterns are much more variable – in which case separate MOY-K-factors may be needed. This analysis should be repeated when establishing factoring models in a new geographic region, but existing research showing that monthly variation in K-factors is limited [23] makes us more confident in aggregating the factor here.

Only a single formula then is needed for location-specific K-factors, which is shown in Equation (*27*). Only TWT volumes need to be considered by the formula, as these are the only days captured by TMCs.

$$K_j = \frac{\frac{1}{n_j}\sum_{m=1}^{12}\left[\frac{1}{n_{jm}}\sum_{d=1}^{3}\left(\frac{1}{n_{jmd}}\sum_{i=1}^{n_{jmd}} 8hBV_{jmdi}\right)\right]}{\frac{1}{n_j}\sum_{m=1}^{12}\left[\frac{1}{n_{jm}}\sum_{d=1}^{3}\left(\frac{1}{n_{jmd}}\sum_{i=1}^{n_{jmd}} 24hBV_{jmdi}\right)\right]} \tag{27}$$

Where

$K_j =$	K-factor for location *j*
$24hBV_{jdmi} =$	24-hour bicycle volume at location *j* on the *i*th occurrence of day-of-week *d* in month *m*
$8hBV_{jdmi} =$	8-hour bicycle volume at location *j* on the *i*th occurrence of day-of-week *d* in month *m*
$n_{jdm} =$	Number of valid 8hBV observations for day-of-week *d* in month-of-year *m* at location *j*
$n_{jm} =$	Number of TMC day-of-weeks (1, 2, or 3) for which there is at least one valid 8hBV observation in month *m*
$n_j =$	Number of months for which data is available for a given year (in the current data set, data are available for Jan, Feb, and March and therefore $n_j = 3$)

4.2.2 K-factor Validation

A subset of count locations from the Pima County dataset was used to test whether K-factors could be used to generate accurate 24-hour count estimates from 8-hour counts. For every subject count location, an average K-factor was calculated across all other count locations in the subset, simulating a scenario where all count locations belonged to a single factor group. The average factor group K-factor was then applied to the subject count location 8-hour counts from 01/01/2020 to 15/03/2020 to produce 24-hour count estimates, which could be compared to the actual 24-hour counts for that location for those same days. This process was repeated so that each site was treated in turn as the subject count location. Resulting predicted 24-hour counts are shown by Figure 24 (each datapoint corresponds to one day of data at a given location to which a K-factor was applied).

Predicted 24-hour counts were relatively accurate, as illustrated by an obtained R^2 of ~0.88 between actual and estimated 24-hour counts. The mean absolute percent error (MAPE) across all estimates at all subject count locations was 24.1%, with a standard deviation of absolute percent errors (APEs) of 18.1%. Even without grouping count locations with similar daily count profiles, K-factors calibrated using data from PCLs were able to produce estimated 24-hour counts from one-day 8-hour counts that were within ~24% of the actual 24-hour count for that day, on average. Although the spread of results was relatively high, this indicates that in the aggregate K-factors could be a viable component of the AADB estimation process.

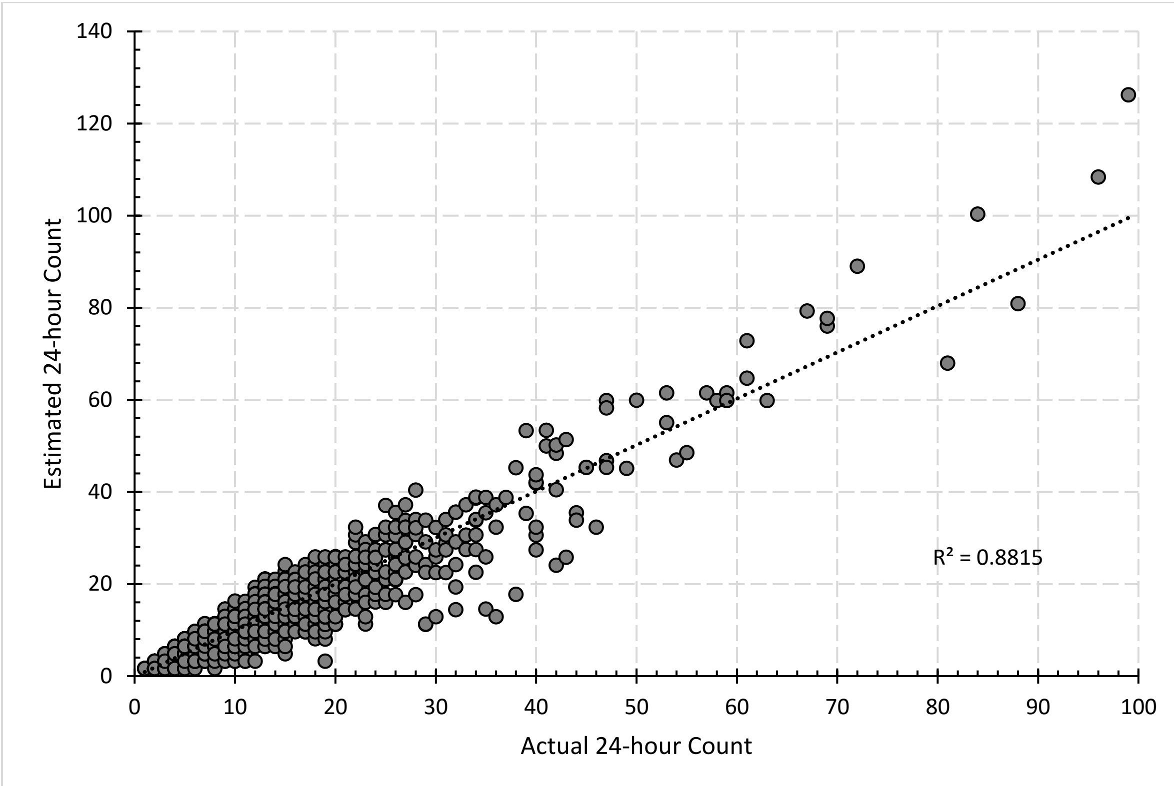


Figure 24: Actual and predicted 24-hour cyclist count values for Pima County dataset

4.3 Adapting the "Traditional" Method

The "DOW/MOY" method, often referred to as the "traditional" method [31] addresses several of the issues with expanding 8-hour TMCs identified in Section 4.1. Most importantly, its relative simplicity in terms of the number of factors per count location – 19 factors, for seven DOWs and twelve MOYs, compared to 84 factors for the day-of-week-of-month method, or 365 for the day-of-year method – should make correctly identifying STCL factor group membership more reliable.

However, the traditional method was developed for conventional cyclist count data, which differs from 8-hour TMC data in several important ways and consequently some adjustments to the method may be needed to adapt it for use in the TMC context.

4.3.1 Refined DOW and MOY Factors

As was done with K-factor's, a nested ANOVA was conducted to assess the proportion of variance in daily traffic patterns that was attributable to different temporal factors. A "daily factor" was calculated for every day of data at every location in the subject dataset. This value is the ratio of the 24-hour count on a given day to the AADB of the location at which that count was collected.

Again, since weekend traffic is not relevant to TMCs, the analysis was limited to TWT counts from January, February, and March. Results for the full set of Pima County locations are summarized in Table 21. As was done for K-factors, a second ANOVA was conducted on a subset of high-volume locations to avoid the disproportionate variation associated with low-volume locations.

Table 21: Summary of nested ANOVA results for Daily Factor values

Variance Component	DF	Sum Sq.	Mean Sq.	F	Pr(>F)	η^2	ω^2
location	33	11.86	0.359	2.296	6.49E-05*	7.1%	0.040
location : month	68	16.99	0.250	1.596	0.002*	10.2%	0.038
location : month : DOW	204	29.22	0.143	0.915	0.777	17.5%	0^
within DOWs (residuals)	694	108.63	0.157	-	-	65.2%	-

**Significant at a 95% level of confidence / ^Negative ω^2 value*

Table 22: Summary of nested ANOVA results for Daily Factor values (AA8hB ≥10)

Variance Component	DF	Sum Sq.	Mean Sq.	F	Pr(>F)	η^2	ω^2
location	24	8.05	0.335	2.650	4.25E-05*	8.0%	0.050
location : month	50	6.09	0.122	0.962	0.550	6.0%	0^
location : month : DOW	150	21.37	0.142	1.126	0.174	21.2%	0.024
within DOWs (residuals)	517	65.42	0.127	-	-	64.8%	-

**Significant at a 95% level of confidence / ^Negative ω^2 value*

The ANOVA conducted on the full set of 34 locations (Table 21) found MOY to be a significant variable, accounting for a larger proportion of variance than either location or DOW. However, model outcomes for the high-volume subset of locations (Table 22) were similar to those for K-factors, with MOY and DOW explaining very small proportions of variance in the dependent variable. In addition to the difference between mean of daily factors being insignificant between MOYs and DOWs within those MOYs, the mean of daily factors was also not significantly different between locations.

Given the lack of evidence to indicate that there is a statistically significant effect size of the DOW variable (p-values are much larger than 5%, and therefore there is not enough evidence to confidently reject the null hypothesis that there is a different mean K-factor value on each TMC day-of-week) it is appropriate to combine Tuesday, Wednesday, and Thursday traffic into an aggregate count, as was done with K-factors. A more aggregate approach is desirable in that it should be possible to model a single TWT factor value for STCLs more reliably than it would be to model a disaggregate TWT profile. Accordingly, a combined "Tuesday-Wednesday-Thursday" (TWT) factor is proposed, representing the ratio of average 24-hour TWT volume to AADB (Equation (*28*]).

$$TWT_j = \frac{\frac{1}{n_j}\sum_{m=1}^{12}\left[\frac{1}{n_{jm}}\sum_{d=1}^{3}\left(\frac{1}{n_{jmd}}\sum_{i=1}^{n_{jmd}} 24hBV_{jmdi}\right)\right]}{AADB_j} \tag{28}$$

Where

TWT_j	=	Tuesday-Wednesday-Thursday combined factor for location *j*
$AADB_j$	=	Annual average daily bicycle volume at count location *j*
$24hBV_{jdmi}$	=	24-hour bicycle volume at location *j* on the *i*th occurrence of day-of-week *d* in month *m*, where *d* is 1=Tuesday, 2=Wednesday, or 3=Thursday
n_{jdm}	=	Number of valid 8hBV observations for day-of-week *d* in month-of-year *m* at location *j*
n_{jm}	=	Number of TMC day-of-weeks (1, 2, or 3) for which there is at least one valid 8hBV observation
n_j	=	Number of TMC months-of-year *m* for which there is at least one valid 8hBV observation

The results in Table 22 also show that there is insufficient evidence to indicate that MOY is a significant variable in daily factor values for high-volume locations. However, this does not mean that the MOY factor should be aggregated as the DOW factor was. Data for the daily factor ANOVA model was limited to January, February, and the first half of March, a period across which the weather is consistent (and, subjectively, comfortable) in Pima County where the data was collected. Although relative TWT volume may not vary significantly from January to March in Pima County, this does not mean that there will be no difference between STC-months in other locations. As a result, we hypothesize that MOY would have been shown to be statistically significant if data for the entire year had been available. Given the limitations of the available data set, this could not be confirmed, however, for this reason, we have decided not to remove or aggregate the MOY factor.

Thus, the calculation of MOY factors remains unchanged from the traditional method and can be seen in Equation (*29*). Note that while the TWT and K factors are calculated using only Tuesday-Wednesday-Thursday counts, the MOY factor makes use of all available data. Calculating the MOY factor with just TWT counts would result in the relative difference between average day-of-week volumes and AADB being "double-counted", increasing estimation error.

$$MOY_{jm} = \frac{\frac{1}{n_{jm}}\sum_{d=1}^{7}\left(\frac{1}{n_{jmd}}\sum_{i=1}^{n_{jmd}} 24hBV_{jdmi}\right)}{AADB_j} \tag{29}$$

Where

$MOY_{jm} =$ Month-of-year factor for month *m* location *j*

$AADB_j =$ Annual average daily bicycles for location *j*

$24hBV_{jdmi} =$ 24-hour bicycle volume at location *j* on the i^{th} occurrence of day-of-week *d* in month *m*

$n_{jdm} =$ Number of valid 8hBV observations for day-of-week *d* in month-of-year *m* at location *j*

$n_{jm} =$ Number of days-of-week in month *m* for which there is at least one valid 8hBV observation at location *j*

4.3.2 Improved Factor Grouping Methods

In past research, short term count locations are generally assigned to a single factor group (i.e. group of permanent count locations) based on one or some combination of traffic pattern characteristics ([11], [31], [39], [33]). For example, Nordback et al [28] use a combination of WWI values, visual inspection of temporal plots, and local expertise to divide count locations into three distinct groups: recreational, mixed, and utilitarian. Although several input metrics were applied, the output was a single vector containing a single factor group ID for each count location.

However, this approach may not be sufficient when using TMCs to estimate AADB. The goal of factor grouping is to minimize temporal variation among locations in a group; if locations are grouped according to their day-of-week traffic profiles, then variation is minimized with respect to day-of-week patterns. But there is no guarantee that a location's day-of-week profile has any relationship to its month-of-year profile; within that same group, variation may be very high with respect to month-of-year patterns. The result of using this single grouping schema will be accurate monthly average daily bicycle (MADB) estimates and inaccurate AADB estimates. Therefore, a "triple-grouping" approach is proposed to factor grouping for the refined traditional method. Under this approach, three separate grouping schemas are established, one for each factor type: A K-group (f^{k}), a TWT-group (f^{d}), and a MOY-group (f^{m}).

Factor values can then be calculated across each group, using Equations (*30*), (*31*), and (*32*):

$$K_{f^k} = \frac{1}{n_{f^k}} \sum_{j=1}^{N} \alpha_j \times K_j \quad (30)$$

Where

$K_{f^k} =$ K-factor for factor group f^k

$K_j =$ K-factor for count location *j*

$\alpha_j =$ 1 when count location *j* is in factor group f^k; otherwise, 0

$N =$ Total number of individual permanent count locations

$n_{f^k} =$ Number of individual permanent count locations in factor group f^k

$$TWT_{f^d} = \frac{1}{n_{f^d}} \sum_{j=1}^{N} \alpha_j \times TWT_j \quad (31)$$

Where

$TWT_{f^d} =$ Tuesday-Wednesday-Thursday factor for factor group f^d

$TWT_j =$ Tuesday-Wednesday-Thursday factor for count location *j*

$\alpha_j =$ 1 when count location *j* is in factor group f^d; otherwise, 0

$N =$ Total number of individual permanent count locations

$n_{f^d} =$ Number of individual permanent count locations in factor group f^d

$$MOY_{f^m m} = \frac{1}{n_{f^m}} \sum_{j=1}^{N} \alpha_j \times MOY_{jm} \quad (32)$$

Where

$MOY_{f^m m} =$ Month-of-year factor for month *m* and factor group f^m

$MOY_{jm} =$ Month-of-year factor for month *m* and count location *j*

$\alpha_j =$ 1 when *j* is in factor group f^m; otherwise, 0

$N =$ Total number of individual permanent count locations

$n_{f^m} =$ Number of individual permanent count locations in factor group f^m

The empirical factor grouping methodology, making use of a K-means clustering algorithm, is discussed in more detail in Chapter 3.

4.3.3 Final Proposed Method

The summation of the previous sub-sections is an updated method for AADB estimation from 8-hour TMCs which addresses the problems associated with this type of cyclist count data as identified in Section 4.1. The calculation of factors is performed using data from PCLs, which have up until now been denoted by the subscript *j*. AADB estimates are meant to be generated at STCLs where it cannot be calculated directly; these locations will be denoted by the subscript *s*. The method is expressed by Equation (*33*):

$$\widehat{AADB}_{sf^kf^df^m} = 8hBV_{sdm} \times \frac{1}{K_{f^k}} \times \frac{1}{TWT_{f^d}} \times \frac{1}{MOY_{mf^m}} \quad (33)$$

Where

$\widehat{AADB}_{sf^kf^df^m}$ = AADB estimate for STCL *s* in month *m*, which is a member of factor groups f^k, f^d, and f^m

$8hBV_{sdm}$ = 8-hour TWT bicycle volume at STCL *s* on day-of-week *d* in month *m*, where *d* is one of Tuesday, Wednesday, or Thursday

K_{f^k} = K-factor for factor group f^k

TWT_{f^d} = TWT-factor for factor group f^d

MOY_{f^m} = MOY-factor for factor group f^m

4.4 Results

The proposed method for 8-hour TMC expansion was tested using the Pima County dataset. The full dataset was filtered and refined, as was outlined in Section 4.4.1. Only data from January 1, 2020 to March 15, 2020 was used under the assumption that the COVID-19 pandemic would significantly alter traffic patterns in the second half of March and beyond. The analysis workflow is shown in Figure 25. The numbers in parentheses in some of the boxes indicate the equations used to carry out the calculations for that step in the process. The analysis workflow was repeated for 4 different factor grouping schemes, namely schemes that consisted of dividing the sites into only a single group, 2 groups, 3 groups, and 4 groups.

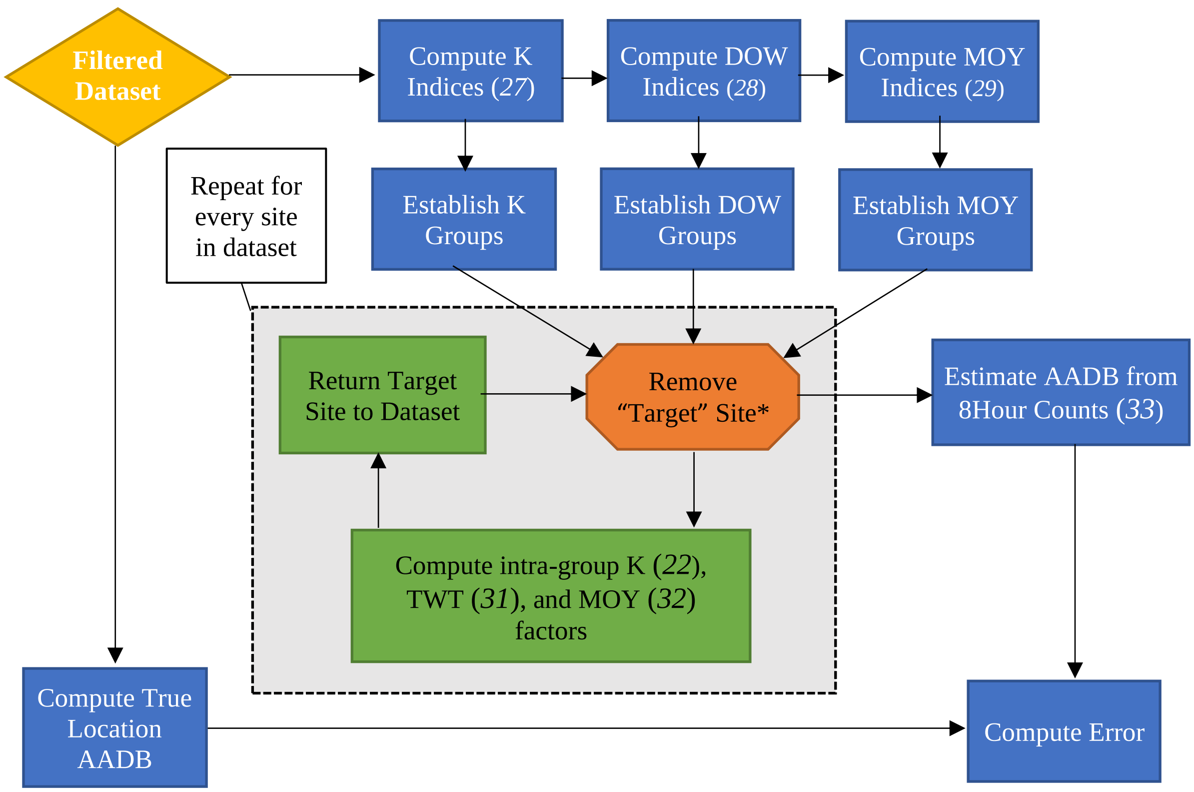


Figure 25: Analysis workflow for testing improved AADB estimation method, adapted from Nordback et al [28]

Table 23: Summary of AADB estimation results for the "Proposed" TMC expansion method

Method	Groups	MAPE	SD of APE	Weighted MAPE	p-value (T<=t)
Proposed Method	1	44.5%	37.9%	38.8%	-
	2	42.9%	37.6%	38.5%	0.14
	3	**40.9%**	**33.6%**	**37.9%**	0.046*
	4	40.6%	34.0%	38.5%	0.039*

**Significant at a 95% level of confidence*

A summary of estimation results using the "proposed" method is shown in Table 23. Estimation accuracy is quantified in terms of the mean absolute percent error (MAPE). There is a noticeable improvement in MAPE for 2, 3, and 4-group scenarios compared to placing all locations in a single factor group (44.5% MAPE for 1-group, 40.6% MAPE for 4-groups). Paired two-sample T-tests were used to compare successive grouping scenarios to the baseline of a single factor group. Under the assumption that increasing the number of groups should decrease estimation error, p-values for a onetail T-test is provided. Although a significant difference was not found for the 2-group scenario, both the 3 and 4-group scenarios were found to produce statistically significantly lower average error across all locations than the 1-group scenario, at a 5% level of significance. This would indicate that factor grouping using empirical clustering at least slightly improved estimation accuracy, as would be expected to be the case, although an analysis of variance across all four scenarios did not find a significant effect for the number of factor groups ($F\{3\}=0.627, p=0.599$).

Results for the proposed method were comparable or better than results for the three baseline methods (as described in Chapter 3), although means testing using paired two-sample T-tests was unable to show a significant difference between the proposed and any of methods for the same number of factor groups ($t\{33\} = -0.359, p = 0.722$ for 2-group proposed method versus 2-group DOW/MOY method).

A summary of estimation results by location for the 3-group scenario is shown in Table 24. The standard deviation of absolute percent error (APE) across all locations was 33.6%, indicating a wide spread of estimates, as was the case in previous trials. This is further illustrated by Figure 26, which shows the spread of AADB estimates around their mean, by location. Note that true AADB is indicated by a red dot. In most cases, the mean of AADB estimates was close to truth; however, the spread of estimates was very large (the range of estimates for Location 49, for instance, was 16 to 162, even though MAPE was only 33%). As was the case for the baseline results, MAPE tends to be higher at locations with low AADB, as does the relative spread of estimates (MAPE for locations with AADB less than 10 is 50.1%; for AADB greater than 10, 37.6%). Intuitively it makes sense that mean estimation error at a location is directly related to count variability. The more spread-out counts are around the AADB, the more spread out AADB estimates will be. Future work could be done to investigate whether count-variability can be explained by site characteristics, so that potentially high-error sites could be identified and a type of confidence interval for estimates established.

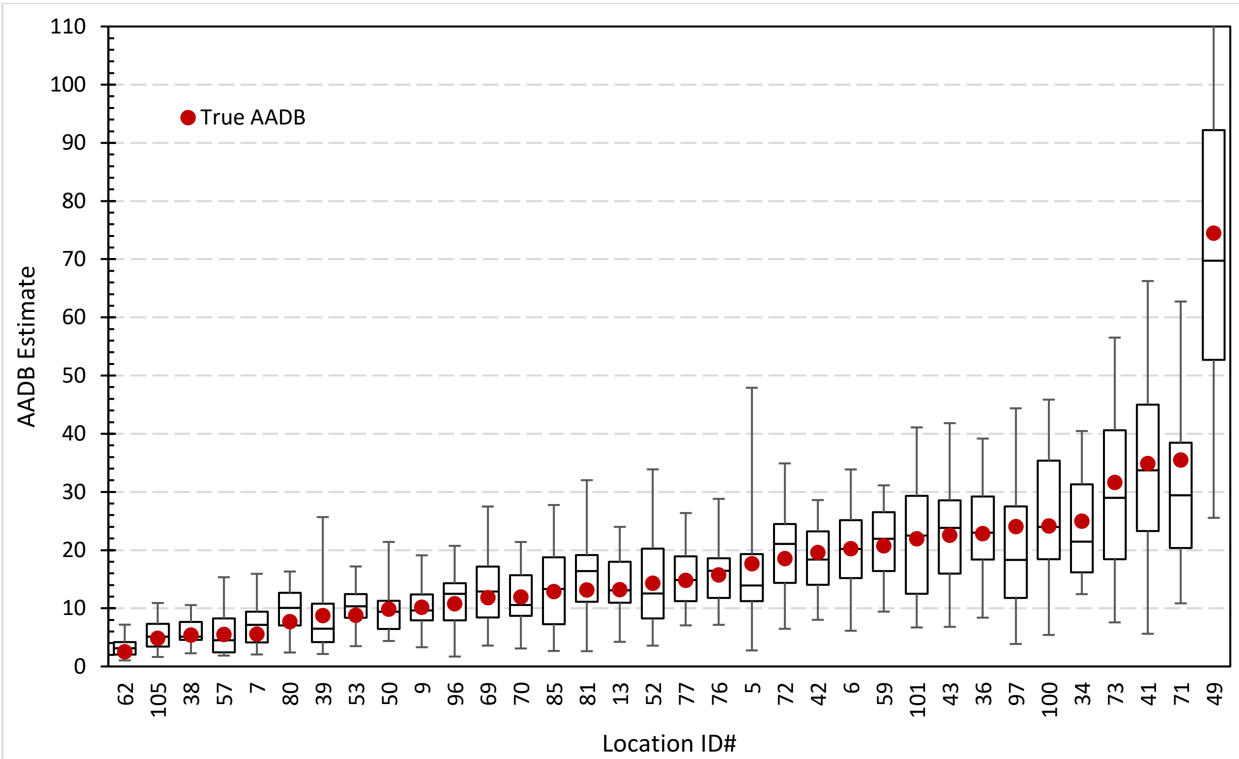


Figure 26: 95% Confidence Interval Plot for AADB Estimation Results by Location (3-group Scenario)

Table 24: Summary of AADB estimation results by location for the 3-group scenario

Location	True AADB	K group	TWT group	MOY group	Avg. Estimate	MAE	SD of AE	MAPE	SD of APE
39	8.7	0	0	2	9.2	0.5	5.3	5.0	60.8%
81	13.2	0	1	0	16.1	2.9	6.8	5.4	51.6%
5	17.6	0	1	0	17.5	-0.1	8.0	12.1	45.5%
101	22.0	0	1	2	22.1	0.1	8.1	6.3	36.8%
42	19.6	0	2	0	18.4	-1.2	4.8	3.2	24.4%
34	25.0	0	2	0	23.6	-1.4	8.1	4.3	32.5%
77	14.8	0	2	2	15.7	0.9	4.4	3.6	29.6%
53	8.8	1	0	0	10.4	1.6	3.3	2.4	37.9%
76	15.7	1	0	0	15.8	0.0	4.8	3.6	30.5%
72	18.5	1	0	0	20.1	1.6	6.8	4.3	36.6%
59	20.7	1	0	0	21.0	0.3	5.7	3.3	27.4%
36	22.8	1	0	0	23.6	0.8	6.8	5.5	29.6%
100	24.2	1	0	0	25.6	1.4	9.0	6.9	37.2%
57	5.5	1	0	1	5.8	0.3	3.4	2.6	62.2%
43	22.5	1	0	2	23.4	0.8	8.2	6.2	36.4%
97	24.0	1	1	0	20.0	-4.0	10.0	6.7	41.6%
7	5.6	1	1	2	7.2	1.7	3.6	2.8	64.8%
69	11.9	1	1	2	13.8	2.0	5.9	4.2	50.0%
52	14.3	1	1	2	14.7	0.4	7.3	5.9	51.2%
73	31.6	1	1	2	30.0	-1.7	12.3	8.9	38.8%
41	34.9	1	1	2	34.4	-0.5	15.1	10.4	43.4%
6	20.2	1	2	0	19.9	-0.3	6.4	4.3	31.8%
96	10.8	1	2	2	11.5	0.7	4.4	3.0	40.7%
85	12.9	1	2	2	13.5	0.6	6.2	4.2	48.1%
105	4.8	2	0	0	5.4	0.6	2.5	1.6	51.5%
70	11.9	2	0	0	11.5	-0.4	4.4	2.8	36.8%
62	2.5	2	0	1	3.3	0.7	1.3	1.3	52.8%
50	9.8	2	0	2	9.8	0.0	3.6	3.0	36.1%
13	13.2	2	0	2	14.1	0.9	4.9	3.6	37.2%
71	35.5	2	0	2	31.4	-4.1	12.3	8.6	34.8%
49	74.5	2	1	0	72.1	-2.4	25.3	18.1	34.0%
38	5.4	2	2	0	5.7	0.3	1.9	1.6	34.8%
80	7.7	2	2	2	9.7	1.9	3.9	2.4	50.0%
9	10.2	2	2	2	10.3	0.1	3.4	2.9	33.0%
							All:	40.9%	33.6%
							VMAPE:	37.9%	-

4.4.1 Permanent Count Location Selection

Extensive filtering has already been done to identify and remove from the dataset erroneous counts. This process is detailed extensively in Chapter 2. However, even after removing obviously erroneous counts, it was noticed that the day-to-day variation in factor values was much higher at some locations that others. It was thought that the wide spread of count values relative to AADB at these locations might be having an undue influence on factor calculation. Accordingly, daily factors were calculated for every Tuesday, Wednesday, and Thursday in the Pima County dataset as the daily count for that day divided by AADB for the location at which that count was collected. A coefficient of variation (CV) was then calculated for every location as the standard deviation of these daily factor values over their mean. Those in the top half of daily factor CVs were flagged as "high variation" count locations. Figure 27 illustrates the spread of daily factor values at Pima County locations, with the "high variation" locations indicated by a red dot.

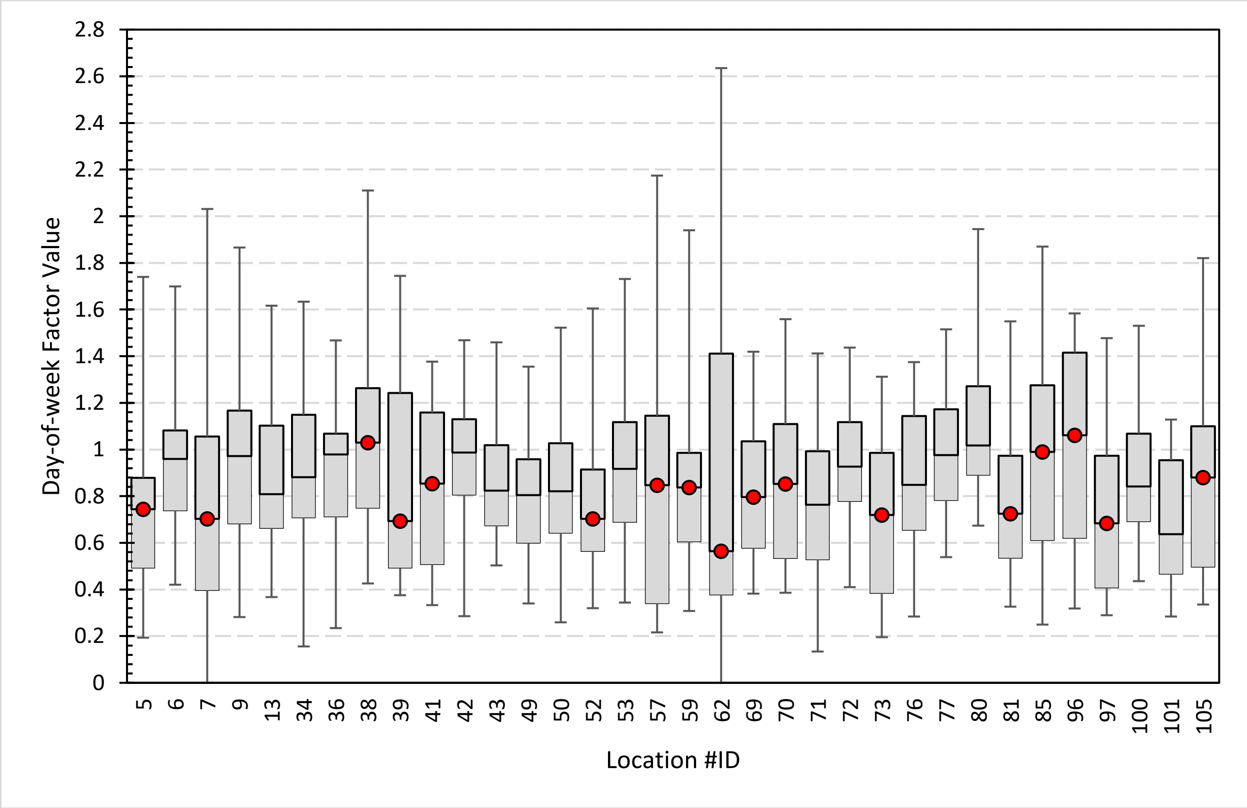


Figure 27: Plots of Tuesday, Wednesday, and Thursday Daily Factor values by Pima County count location. Boxes show the inter-quartile range, and whiskers the 95% confidence interval. "High Variation" locations are indicated by a red dot at the median.

Accordingly, it was hypothesized that estimation error might be reduced by splitting the dataset into two groups, a “training” set comprised of locations with low relative variation in their counts; and a “test” set comprised of the remaining locations. Factors would be calculated using data from the training set, and applied to both the training and test set, as was done in the previous section.

Several different methods of selecting locations for the training set were tested. Methods were evaluated using the proposed method shown in Section 4.3.3, under a 1-group factor grouping scenario. These results are shown in Table 25 below. As can be seen, in no case did using a reduced “training” set result in significantly different estimation results than the baseline, and in only one case was overall estimation error observed to be reduced (when the training set was chosen as the 50^{th} percentile of locations in terms of their CVs in daily factor values). This would imply that removing locations with high variation in this case made no difference to estimation results. Although average estimates did change slightly on a location-by-location basis, overall accuracy was similar. Even comparing within the model and test groups, average error was about the same in both cases.

Without significant differences in MAPE, there was no metric for objectively determining the “correct” set of locations for model calibration, so in this instance there is no reason not to use data from all available count locations to calibrate the factor model. It may still be beneficial in some circumstances to remove data from a calibration dataset – say, if one has access to a very large pool of data or if some locations are obvious outliers in terms of their temporal patterns – but in this case it did not have any significant effect.

Table 25: Pima County estimation results for reduced training set

Training Set	**Set Size**	**MAPE**	**SD of APE**	**Weighted MAPE**	**p-value (T<=t)**
Baseline	34	44.5%	37.9%	38.8%	-
AA8hB > 3	26	46.5%	40.7%	40.2%	0.631
AA8hB > 8	17	49.5%	44.9%	42.5%	0.257
50^{th} percentile CV for K-factors	17	47.8%	42.4%	41.1%	0.448
50^{th} percentile CV for Daily factors	17	43.0%	35.6%	37.9%	0.398

**Significant at a 95% level of confidence*

Chapter 5
Modelling Factor Groups Using Locational Characteristics

This chapter will address the challenge presented by matching short-term count locations (STCLs) with appropriate "factor groups" for the purposes of expanding 8-hour turning movement counts (TMCs) to estimates of annual average daily bicycles (AADB). The estimation methods outlined in previous chapters require that permanent count locations (PCLs) be grouped according to the similarity of their bicycle traffic patterns. A set of expansion factors can then be calculated across each group and applied to one-day TMCs from STCLSs to generate AADB estimates. Various literature on this subject states that AADB estimation error can be reduced by grouping STCLs with the correct factor group, as this will minimize the difference between the "true" traffic patterns present at the STCL and the traffic patterns expressed by the factor group. However, with less than 24-hours of count data, it is difficult (or impossible) to do so empirically, since the multi-day traffic profiles used to group the PCLs are unavailable. Thus, several alternative approaches have been proposed, which will be explored here: first, a linear regression approach, which attempts to directly model average factor values as a function of several locational attributes; second, a logit regression approach which models the probability of a location belonging to a given group; and finally, a generalized approach which uses a quasi-subjective evaluation framework to construct context-specific guidelines for assigning STCLs to a factor group.

5.1 Multivariate Linear Regression Approach

Little explicit emphasis has been made in the literature on the relationships between temporal cyclist count patterns and spatial count location characteristics. However, attempts at modelling direct cyclist demand are common, and provide a good starting point for identifying potentially significant variables. Figliozzi et al [29], for instance, calibrated a "correcting function" that used a combination of temporal and weather variables to improve factor-based AADB estimates by 15% on average. Griswold [40] calibrated separate log-linear models for weekend and weekday 2-hour afternoon cyclist volumes at intersections. They found the positive influence of commercial retail and post-secondary schools to be greater on weekdays than on weekends; and the influence of bicycle facilities to be greater on weekends [40]. A number of other similar sources were reviewed and summarized in Table 26.

Table 26: Summary of past research on direct demand cyclist count modeling

Variable	Expected Relationship	Sources
Socio-Economic Variables		
Age - % of Residents over the age of 65 or under the age of 5	(-)	[24]
Educational Attainment - % College Educated	(+)	[24] [41]
Household Income	(+)	[24] [41]
Land Use Variables		
Population Density	(+)	[24]
Number of commercial properties	(+)	[40]
Distance to post-secondary institutions	(-)	[40]
Distance from the CBD	(-)	[41]
"Land use mix"	(+)	[41] [42]
Employment	(+)	[42] [43]
Presence of schools	(+)	[42]
Presence of a metro station	(+)	[42]
Presence of an area of commercial use	(+)	[42]
Presence of institutional land uses	(+)	[44]
Bus stops	(+)	[44]
Roadway Characteristic Variables		
Presence of bicycle facilities	(+)	[40] [41]
Average slope of surrounding terrain	(-)	[40]
"Node Connectivity" of intersection	(+)	[40]
Bicycle Level of Service	(+)	[43]
Bicycle-trail access	(+)	[43]
Roadway leads to a bridge	(+)	[43]
Number of lanes	(+)	[43]
Speed limit	(-)	[43]
Bike-lane width	(+)	[43]
Weather		
Precipitation	(-)	[29] [45]
Temperature	(+)	[29] [45]
Humidity	(-)	[45]
Temporal Variables		
Holidays	(-)	[29]
Lagging variables (DOY-1)	N/A	[29] [45]
Morning period	(-)	[43]
Hour-of-day, Day-of-week, Month-of-year	N/A	[45]

5.1.1 Data Collection

Data for this study was derived from the Pima County Miovision data, Pima County Open Geospatial Data, and U.S. Census Bureau statistics. Data was primarily processed in the QGIS open-source geographic information system application, using a combination of geoprocessing tools to link spatially referenced data to the set of study locations.

One difficult decision to make when collecting data in a cyclist demand model is choosing the scale at which to measure variables. Roadway characteristics are inherent to an intersection, but land use characteristics are usually defined in terms of some buffer zone surrounding an intersection. Including variables at multiple scales can improve model development [44]. Buffer distances vary from study to study. Medury used different search distances for different land use classes, ranging from 500 to 1200 meters, for a study of mixed pedestrian-cyclist traffic [13]; Tabeshian and Kattan performed a multi-scaled analysis to develop a model for direct cyclist demand at intersections which included institutional land uses within 0.5 miles, low density residential within 0.1 miles, commercial density within 0.1 miles, and bus stops within 0.25 miles [44]; Mukoko cites National Household Travel Survey data which says that about 60% of all bicycling trips in the United States are one mile or less in establishing a one-mile "ideal buffer width" [46]; Strauss used a 400 meter buffer for employment and schools, and an 800 meter buffer for metro stations and land use mix, but just a 50 meter search area for commercial land uses [42]. Common across all these studies seems to be the assumed positive relationship between the "gravity" of a land use and its area of influence (i.e. the ideal buffer distance for a university is probably wider than the ideal buffer for a convenience store).

The land use patterns of Pima County further complicate this choice. A substantial number of the sample count locations do not have any land uses abutting them, meaning that too small a buffer will not capture the dominant surrounding land uses, which likely still influence traffic patterns. After trialing several alternatives in QGIS, 450-meter and 900-meter buffer distances seemed to adequately capture surrounding land-uses in the Pima County context. Both distances were tested for most variables, with the better correlated between the two options carried through to the final analysis.

Variables considered are summarized in Table 27.

Table 27: Summary of locational land-use and roadway characteristics

Variable Name	**Short Form**	**Description**	**Source**
Annual average Daily Traffic	*aadt*	Sum of average daily intersection volume across all approaches.	Derived from intersection traffic monitoring data.
Weekend-Weekday Index	*wwi*	Ratio of intersection average weekend to weekday motorized vehicle traffic.	Derived from intersection traffic monitoring data.
Speed Limit	*sl*	Average speed limit across all intersection approaches.	Pima County Open Data (*Speed Limits*, 2021)
Traffic Lanes	*lanes*	Sum of *through lanes* (no turning lanes) across the widest of the North-South approaches, and the widest of the West-East approaches.	Pima County Open Data (*Number of Lanes*, 2021)
Bike Lane	*b_lane*	"1" if a bike-lane is present on any of the approaches, otherwise "0".	Pima County Open Data (*Bicycle Routes,* 2021)
Proportion of Heavy Vehicles	*%hv*	Proportion of AADT which is "heavy" traffic (trucks, semi-trailers, buses).	Derived from intersection traffic monitoring data.
Presence of a Safer Alternative	*posa*	"1" if an off-road cycling route (such as a multi-use path) is present within 450 meters of an intersection; otherwise "0".	Pima County Open Data (*Bicycle Routes*, 2021)
Population	*pop*	Total population of the census division containing an intersection, divided by 10,000.	U.S. Census Bureau (2019) *American Community Survey: Age and Sex*
Population Density	*per_km*	Population per square kilometer of the census division containing an intersection.	U.S. Census Bureau (2019) *American Community Survey: Age and Sex*
Age	*age*	Median age of the census division containing an intersection.	U.S. Census Bureau (2019) *American Community Survey: Age and Sex*
Income	*inc*	Median income of census division containing an intersection, divided by 10,000.	U.S. Census Bureau (2019) *American Community Survey: Income*
Parcels	*count_all*	Count of all land use parcel centroids within 900 meters of an intersection, divided by 1000.	Pima County Open Data (*Parcel Centroids*, 2021)

Variable Name	Short Form	Description	Source
Parcels – Residential	*p_r*	Proportion of parcel centroids within 900 meters of an intersection associated with a "residential" land use.	Pima County Open Data (*Parcel Centroids*, 2021)
Parcels - Commercial	*count_c*	Count of parcel centroids within 900 meters of an intersection associated with a "commercial" land use, divided by 100.	Pima County Open Data (*Parcel Centroids*, 2021)
Parcels - Industrial	*industrial*	"1" if any parcel centroids within 900 meters of an intersection are associated with an industrial land use; otherwise, "0".	Pima County Open Data (*Parcel Centroids*, 2021)
Business Licenses	*licenses*	Count of business licenses geo-located within 900 meters of an intersection, divided by 100.	Pima County Open Data (*Business Licenses*, 2021)
Urban Boundary	*ub*	Distance in kilometers of an intersection to the urban boundary, as delineated by the regional zoning code.	Pima County Open Data (*Zoning – All Jurisdictions*, 2021)
Points of Interest	*poi*	"1" if any civic point of interest is within 450 meters of an intersection (e.g. hospital, library, shopping center, etc.); otherwise "0".	Pima County Open Data (*Points of Interest*, 2021)
Schools	*school*	"1" if any middle school, high school, or post-secondary school is within 900 meters of an intersection; otherwise "0".	Pima County Open Data (*Schools*, 2021)
Parks	*park*	"1" if any park or recreational area larger than 1 km^2 is within 900 meters of an intersection; otherwise "0".	Pima County Open Data (*Parks and Recreation*, 2021)
Trailheads	*trailhead*	"1" if any access point to an off-road path is within 900 meters of an intersection; otherwise "0".	Pima County Open Data (*Trailheads and Trails*, 2021).

5.1.2 Model Development

It is mostly acknowledged that "…OLS is not the best approach for traffic counts," since they are not normally distributed and are non-negative integers [41]. However, this is not an issue here, as factor values are being modelled in place of actual cyclist counts. Average K and TWT factor values were tested using the Shapiro-Wilk normality test, which found in both cases the factors to be normally distributed (see Table 28). This conclusion is also supported by visually inspecting plots of factor values; the histogram of K-factors in Figure 28 shows a distinctly normal distribution. For this reason, OLS regression was deemed to be appropriate.

Table 28: Shapiro-Wilk tests for count-location average factor values

Factor Value	**Mean**	**SD**	**W-Statistic**	**DF**	**Sig.***
K	0.619	0.068	0.974	33	0.594
TWT	0.865	0.101	0.971	33	0.497

** Conclude that sample is not Normally distributed if Sig value ≤ 0.05*

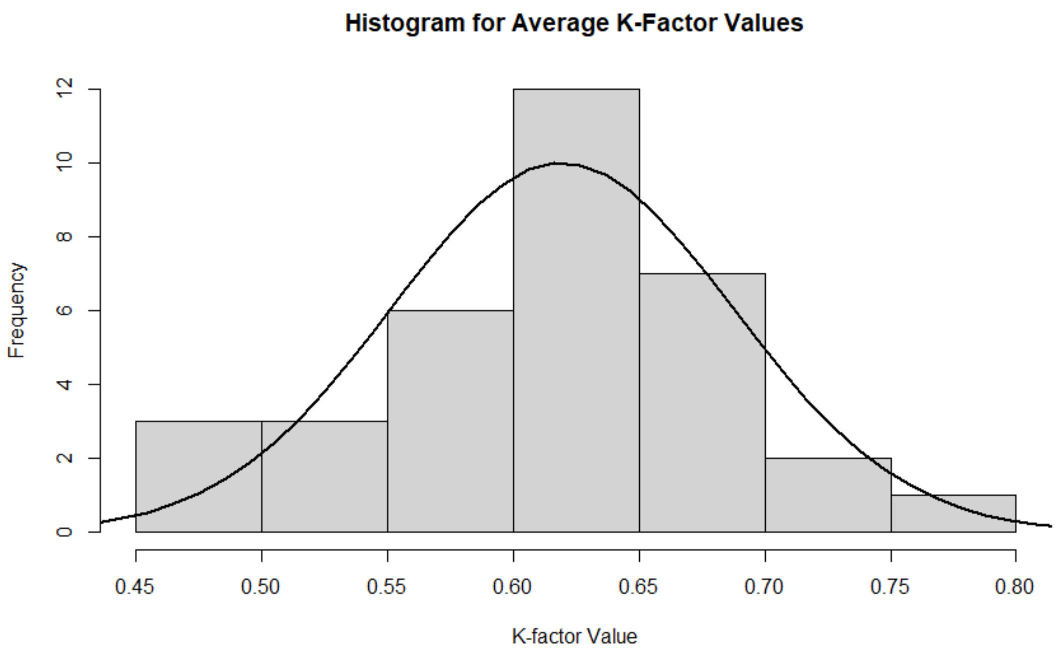


Figure 28: Histogram of average K-factor values for count locations, with normal curve overlay

Variable selection was performed using a mixed stepwise-Akaike's Information Criterion (AIC) minimization approach. Because of the small sample size of just 34 count locations, the goal of this process was to reduce the number of variables to three or fewer per model, or approximately 10 observations per model term.

First, a model was calibrated which included all the available variables described by Table 27. Then, stepwise AIC minimization was performed. Insignificant variables were removed to produce a further reduced model, which was again subjected to AIC minimization. The resulting model was finalized by removing any remaining insignificant variables and testing the assumptions of homoscedasticity and normality of the residuals. A relatively high significance level of 10% was chosen, to reflect the combination of a small sample size and a noisy dataset. Following this process, separate models were developed for K (Table 29) and TWT (Table 30) factors.

Table 29: OLS regression model for direct K-factor estimation

	Estimate	Pr (> |t|)
(Intercept)	0.649	< 2e-16 *
posa	0.061	0.029 *
licenses	-0.030	0.054*
trailhead	-0.068	0.01 *
** Significant at α = 0.1*		
$R^2 = 0.229$	F(3, 30) = 4.27, p=0.013	

Table 30: OLS regression model for direct TWT-factor estimation

	Estimate	Pr (> |t|)
(Intercept)	0.9422	1.37e-12 *
aadt	0.0041	0.031 *
age	-0.0045	0.017 *
poi	0.0948	0.053 *
** Significant at α = 0.1*		
$R^2 = 0.2462$	F(3, 30) = 4.59, p=0.009	

The final model for K-factor values includes the presence of a safer alternative (or an off-road path) within 900 meters, the number of commercial licenses within a 900-meter buffer, and the presence of a trail access point within 900 meters. The proportion of variance explained by these variables is relatively low (R^2=0.2462), but the overall F-statistic significance (p=0.013) shows that we can be confident that the model fits the data better than an intercept only model at a 95% level of confidence.

The final model for TWT-factor values included the annual average motor vehicle traffic, the median age of the census tract containing the count location, and the presence of any places of interest within 900 meters. Again, the model R^2 is relatively low (0.2462), indicating the predictive power of the model is weak, but the overall model is significant at a 95% level of confidence (p=0.009).

5.1.3 Model Evaluation

The models developed in Section 5.1.2 were applied to the 34 Pima County count locations to test the viability of using multiple-linear regression to predict STCL group membership. This process followed an iterative structure, as illustrated by Figure 29.

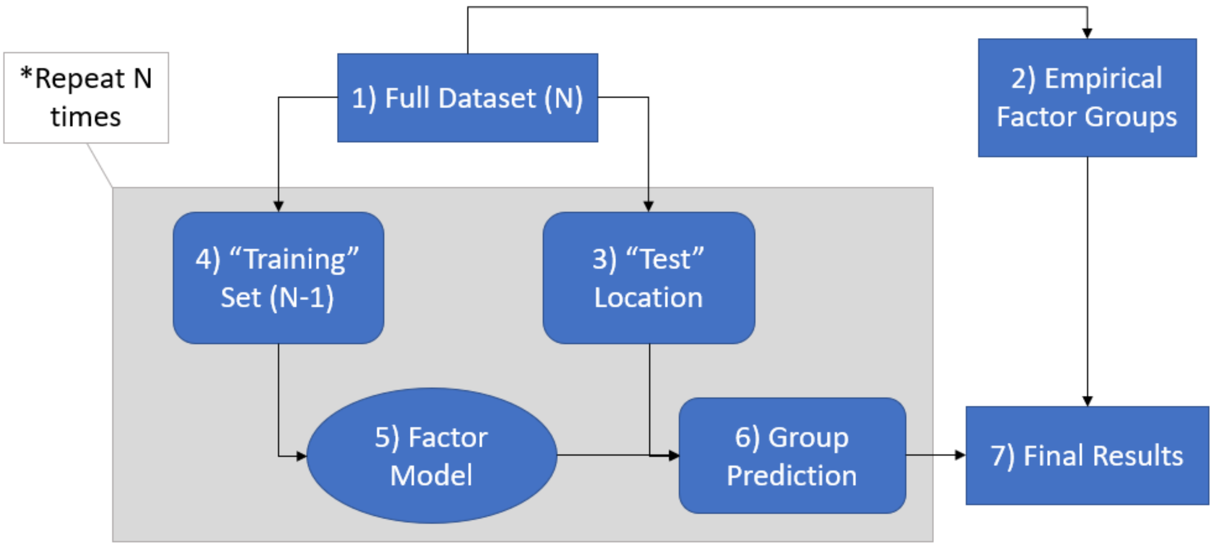


Figure 29: Model evaluation process for STCL group prediction

Starting from the full set of 34 locations (1), empirical clusters were calculated to establish ground "truth" (2). One by one, locations were removed from the dataset to form the "test" set (3), leaving the remaining locations to be the "training" set (4). Model parameters were then calibrated using the training set (5) and used to predict a factor value for the test location. This value was used in turn to predict factor group membership (6). Steps (3) through (6) were repeated once for every location in the sample, meaning that 34 different models were calibrated. Predicted factor groups were compared to the empirical groups established in step (2) to generate final results (7). This analysis was repeated twice, for K-factors and for TWT-factors. For both factor types, two, three, and four-cluster scenarios were all considered.

The predicted factor group for the test location was determined by finding the nearest neighbour to the predicted factor value within the set of cluster centroids. Take for example a two-group k-factor example where $k_1 = 0.34$ and $k_2 = 0.45$; a test location with a predicted k-factor value of 0.41 will be assigned to group k_2. Results of the model evaluation process are quantified in terms of the prediction accuracy (i.e. fraction of the test locations that have been assigned to the correct factor group) and summarized in Table 31.

Table 31: STCL group prediction results using multivariate linear regression

# of Groups	**Group Prediction Accuracy**	
	K-Factor	*TWT-Factor*
2	76%	76%
3	50%	41%
4	41%	35%

As would be expected, overall prediction accuracy was the highest for the 2-group scenario (76% accuracy for both K and TWT factor groups) and declined as the number of groups was increased. Prediction accuracy for both factor types was at or below 50% for the three and four group scenarios. It is possible that a model calibrated from a larger dataset would produce better results, but it's also possible that in reality there are no more than two distinct groups in the Pima County context. As has been noted, one of the shortcomings of the dataset used for this study is the homogeneity of the sample count locations, both in terms of cyclist traffic patterns and in terms of locational characteristics. To accurately model variation in cyclist count patterns, there needs to be a significant amount of non-random variation in the data.

Note that MOY factors were not considered for this analysis. While K and TWT factors are aggregated to a single value for every location, a location's relative MOY traffic patterns are expressed by a vector of month-specific values (in this analysis, each location has three associated values, for January, February, and March). It is not reasonable to calibrate separate regression models for every month of the year; for a full dataset, this would mean calibrating twelve models. Even if individually the models were relatively accurate, each additional model would introduce error into the prediction of MOY factor groups. This method is not likely to produce acceptable results. Furthermore, initial attempts were unable to produce regression models for MOY factor values that were significantly better than intercept-only models. The stepwise regression process described above was not able to produce models for January, February, or March MOY factors that contained any statistically significant independent variables. Clearly, alternative methods will need to be explored for the determination of MOY factor groups for STCLs.

5.1.4 Impact on AADB Estimation

Grouping schema produced by the OLS regression models for K and TWT-factor values were used to generate estimates of AADB using the factor-based estimation method proposed in Chapter 4. These results were compared to grouped and ungrouped results for the proposed method, to assess the error associated with "incorrect" grouping resulting from STCL group prediction. Results for the OLS factor-value prediction model are shown in Table 32, along with results obtained from empirically grouped and un-grouped schemas (which were presented in previous chapters). Note that, because no method for predicting MOY factor groups was developed, it was assumed for the purposes of this evaluation that the "correct" MOY-groups were known. So, the K and TWT-group schema were predicted using the OLS models, but MOY-groups were determined empirically.

Table 32: Summary of AADB estimation results from OLS-predicted factor groups

Method	**Groups**	**MAPE**	**SD of APE**	**Weighted MAPE**	**p-value (T<=t)**
OLS Regression	2	43.6%	35.5%	38.8%	-
Empirical Clustering	2	42.9%	37.6%	38.5%	0.602
Ungrouped	1	44.5%	37.9%	38.8%	0.328

**Significant at a 90% level of confidence*

Using the 2-group K and TWT factor grouping schema predicted by the OLS models described in the previous two sections, AADB estimates were generated with a MAPE of 43.6% across all 34 Pima County locations. Overall estimation results were very similar to results from the 2-group empirical clustering schema, and the ungrouped schema. Two-sample paired T-tests were used to compare MAPE between the "OLS Regression" results to the other sets of estimation results. A significant difference was not found between either of the two pairs. Although it would appear that "correct" empirical clustering resulted in very slightly more accurate estimates than using either the predicted or ungrouped schemas – and that even the "incorrect", predicted grouping schema resulted in more accurate estimates than using just a single factor group – this conclusion is not statistically supported. It would seem as if, in this case, results would be similar no matter how locations were grouped.

5.2 Logit Regression Approach

An alternative to the OLS approach to linking empirical clusters to land use variables described in the previous section is proposed by Medury [13]. Pointing out that "…activity patterns can vary between sites so that there is potential to create more accurate estimates by grouping similar long-term count trends into factor groups," they suggest a probabilistic approach to link empirical clusters of mixed pedestrian-cyclist count locations to land use data collected using the Google Places API. Under this method, the labels for empirical clusters become the categorical response variable in a logistic regression. The resulting model can be used to calculate the probability that a location with a given set of characteristics belongs to group *g*, and the maximum probability among the set of all *g* is used to predict group membership.

This section will outline an attempt to calibrate similar logit models for the Pima County count locations, with separate models developed for each of the K, TWT, and MOY empirical grouping schema. The same set of data was used as was for the OLS method (variables are summarized in Table 27).

5.2.1 Establishing Binary Empirical Factor Groups

We showed in the previous section that prediction accuracy under the OLS method dropped off steeply as the number of clusters increased. Assuming that this would also be the case under a logistic approach, it seems acceptable to limit this evaluation to a binary grouping scenario. As well, increasing the number of groups would decrease the sample size per group, compromising the model quality and reducing the already small number of independent variables which could be included.

K-means clustering was used to establish binary empirical factor groups for K, TWT, and MOY factors. Each of the 34 locations was assigned three group IDs. As will be discussed in more detail, these IDs became the dependent variable in the logistic regression. The group labelled as "0" will be used as the reference group in the regression.

In order to better understand the logit model outputs, the calculated empirical factor groups are summarized in Table 33 below.

Table 33: Summary of binary empirical factor groups for a logistic regression analysis

Factor	**Group 0** ***(reference group)***	**Avg. Factor Value**	**Group 1**	**Avg. Factor Value**
K	Off-peak	0.55	On-peak	0.66
TWT	Weekend	0.79	Weekday	0.97
MOY	Decreasing	Jan – 1.22 Feb – 0.98 Mar – 0.79	Constant	Jan – 0.94 Feb – 0.97 Mar – 1.07

Two distinct groups were identified based on K-factors: an "off-peak" group, the locations in which experience relatively lower relative cyclist volumes during the 8-hour TMC period; and an “on-peak” group, which experience relatively higher 8-hour volumes. For TWT factors, the two empirical factor groups were: a “weekend” group, where average location TWT volume is relatively lower compared to AADB; and a “weekday” group, where TWT volume is relatively higher. And finally, for MOY factors: a “decreasing” group where average cyclist volumes decreased month over month during the January to March study period; and “constant” locations where average monthly volume relative to AADB was consistent.

5.2.2 Model Development

The small sample size for this study of 34 count locations substantially limited certain elements of the development of a logistic regression model. Logistic regression usually requires a large sample size compared to standard OLS. Peduzzi [47] recommends that the sample size for a binomial experiment, N, be equal to ten times the number of covariates divided by the proportion of positive cases. As per these guidelines, N would need to be 10(1/0.65) = 15 to confidently include just one variable in a model for K-factor groups; for two variables, N = 30; for three variables, 46. Long [48] furthermore recommends a minimum N of 100, irregardless of the number of variables. Under this recommendation, we would not have a large enough sample size to conduct a logistic regression study. Medury [13], for reference, used a sample of 153 count locations spread across the southeastern USA. Considering these limitations, models were developed subject to a maximum number of independent variables of two.

The generalized functional form of the logit model is shown in Equations (*34*) and (*35*). These equations are the same for all three of the K, TWT, and MOY models, except that different combinations of linear predictors ($x_j\beta$) will be calibrated for each. Note that the logit model does not directly represent the linear relationship between a set of predictors – in this case the land use variables associated with a given location, such as the number of surrounding commercial buildings – and a probability value. Rather, it models the linear relationship between a set of predictors and the log of the probability that a location belongs to Group 1 over the probability that the location belongs to Group 0. Thus, the calibrated regression models will provide predicted values for the log-odds that a given STCL will belong to Group 1, which can be converted to a probability value, and thus used to predict group membership under some binary classification logic.

$$\theta_j = \frac{e^{x_j\beta}}{1 + e^{x_j\beta}} \tag{34}$$

$$logit[\theta_j] = \ln\left[\frac{\theta_j}{1 - \theta_j}\right] = x_j\beta \tag{35}$$

Where

$\theta_j =$ The probability that location *j* belongs to Group 1

$x_j =$ A vector of covariates associated with location *j*

$\beta =$ A vector of regression coefficients

Model specification followed a very similar process to that described in the previous section for the linear regression models. First, models were calibrated that included the full set of available predictors. These models were, of course, extremely overfit; fitted model values were either 1 or 0, indicating perfect predictive power. Attempts to condense these models using a stepwise AIC minimization approach produced unsatisfactory models as well. So, a reduced set of predictors was manually selected. Variables selected were meant to represent those land use attributes and locational characteristics identified in the literature (and in previous work in this study) as being best correlated with cyclist traffic patterns. These were: business licenses within 900 meters; traffic lanes; the presence of a bike lane; population density (per km^2); the presence of industrial land uses; the motor vehicle traffic weekend-weekday index of the intersection; and the presence of AT trailheads within 900 meters.

The stepwise AIC minimization process was then applied to a model containing this subset of variables to generate final logit models**.** Final models were checked for the statistical significance of their components using Wald tests at a 10% level of significance, with insignificant variables discarded (although some flexibility was applied in this regard). As well, sequential likelihood-ratio tests were used to compare the full models against less complex models in order to simplify them until only two variables remained. Final model specifications are shown in Table 34.

Table 34: Logit regression models for K, TWT, and MOY factor groups

K-factor

Variable	*β*	*SE of β*	*Z-ratio*	*p*	*Odds Ratio*
Intercept	1.406	1.519	0.926	0.355	4.081
lanes	1.370	0.632	2.168	0.030	3.935
perkm	-1.781	0.913	-1.950	0.051	0.168

Model χ^2 = 10.217, $p = 0.006$
Pseudo R^2 = 0.259
n = 34

TWT-factor

Variable	*β*	*SE of β*	*Z-ratio*	*p*	*Odds Ratio*
Intercept	-7.257	2.401	-3.022	0.003	0.001
licenses	2.676	1.220	2.193	0.028	14.529
lanes	1.514	0.582	2.599	0.009	4.544

Model χ^2 = 19.696, $p = 0.000$
Pseudo R^2 = 0.440
n = 34

MOY-factor

Variable	*β*	*SE of β*	*Z-ratio*	*p*	*Odds Ratio*
Intercept	2.260	1.011	2.235	0.025	9.581
licenses	-2.175	1.332	-1.633	0.102	0.114
trailhead	2.661	1.708	1.558	0.119	14.313

Model χ^2 = 7.758, $p = 0.021$
Pseudo R^2 = 0.204
n = 34

Notes: Cox-Snell statistic provided as Pseudo-R^2

Remember that the response variable in a binary logit model is the log-odds of observing a "success". In this case this means observing a location which belongs to the factor group which is coded to be "1". Refer to Table 33 for a summary of empirical factor groups. Interpreting the model outputs can be made somewhat easier by exponentiating coefficients, which are log-odds, to give an odds-ratio. This is the expected multiplicative increase in the odds of a location belonging to group "1" for a one unit increase in the associated variable.

A likelihood ratio test was performed on all three models to evaluate overall model significance; model chi-square (χ^2) values are provided. In each case, the model was shown to be a significant improvement over the “null”, or intercept-only model at the 5% level. Pseudo R-squared statistics were also calculated to give a rough impression of goodness-of-fit (although these values should not be interpreted as direct measures for variance explained, as they would be in OLS regression). These show the model for TWT-groups to be relatively better fit than either of the K or MOY-group models, which is supported by the other evaluation criteria. From this we might conclude that the land-use variables collected for this study are best correlated with temporal patterns in relative weekday-weekend cyclist volumes, out of the three pattern-groups studied.

The model for K-groups indicates that the odds of a location belonging to group *K-1*, the “on-peak” factor group, are about four times higher for every additional traffic lane, all else being equal. Odds of *K-1* membership also decrease by about 83% for every one-unit increase in population density (or every increase in density of 1000 pop/km^2).

TWT group membership seems to be strongly associated with commercial land uses; the odds of a location belonging to group TWT-1, of the “weekday” group, is a full 14 times higher for every additional 100 commercial licenses within 900 meters. As well, TWT-1 membership becomes significantly more likely with every additional traffic lane. In fact, this seems to be the strongest relationship out of any of the variables tested across all three models. Weekend riders may then be trying to avoid wide, high volume roads, as one might expect to be the case.

Finally, the MOY-group model, which may be the most difficult to interpret. It seems as though commercial licenses are associated with lower odds of a location having a “constant” MOY pattern, which is somewhat incongruous. One would expect locations with heavily commercial surroundings to see consistent month-over-month traffic, as commercial uses usually don’t fluctuate seasonally in the way that school or recreational land uses would. Furthermore, the presence of “trailheads” within 900 meters is associated with a large increase in the odds of a location belonging to the “constant” MOY group. Again, this makes little sense, as recreational traffic associated with trailheads would be expected to fluctuate seasonally. It seems likely that, in truth, there are more than two distinct MOY traffic patterns in the Pima County data. The binary approach may not be sufficient, and the missing group(s) may be obfuscating the results. An even more likely limitation is that the sample size (just three months of data from 34 locations) is too small for a robust logistic analysis. Future study could make substantial improvements by addressing this issue.

It should be noted that logistic regression is known to be positively biased in studies with small to moderate sample sizes [49], with odds ratios being systematically overestimated as sample size is decreased. Any conclusions made in this section should be considered carefully with this in mind.

5.2.3 Model Evaluation

The models developed in Section 5.2.2 were applied to the 34 Pima County count locations to test the viability of using logit regression to predict STCL membership. An iterative approach was taken, identical to that presented in Figure 29 for OLS models, but with the addition of a MOY model to the K and TWT models. The empirical factor groups established in Section 5.2.1 were used as ground "truth". One by one, locations were removed from the dataset to form the "test" set, and the remaining locations were used to calibrate three logit models. These models were then used to predict group membership for the test location.

Because binary grouping schema were evaluated, the probability threshold for classifying test locations was set to 0.5. If the expected probability given by the model was less than 0.5, the test location was assigned to Group 0; if greater than 0.5, Group 1. To illustrate the application of the logit models, take *Location 52* as an example. This location belongs to the empirical TWT-factor group "0", or the "weekend" factor group; the number of lanes at this location is 3, and the number of commercial licenses within 900 meters is 42 (see Table 27 for more details). By applying the model coefficients shown in Table 34, an estimate of the log-odds of this location belonging to Group 1 can be calculated as -7.257 + (2.676 x 3) + (1.514 x 0.42) = -1.59. This can be exponentiated to an odds-ratio, 0.20, which can be converted to a probability as 0.20 / (1 + 0.20) = 16.9%. Since the predicted probability is less than the classification threshold of 0.5, *Location 52* is placed into the predicted group "0", which is the same as its "true" empirical group. This process is summarized in Table 35. It should be noted that the model coefficients applied to *Location 52* during the model evaluation were actually very slightly different than the coefficients shown in Table 35, since a model was calibrated specific to each location which excluded it from the calibration dataset.

Table 35: Example of TWT-group Logit Model Application

Location	Empirical TWT-Group	lanes	licenses	Log-odds	Odds	Prob.	Predicted Group
52	0	3	0.42	-1.59	0.20	16.9%	0

Prediction results are shown below in Table 36.

Table 36: STCL group prediction results using multiple logistic regression

Factor	Overall	Group 0	Group 1
K	70.6% (24/34)	81.8% (18/22)	50.0% (6/12)
TWT	79.4% (27/34)	80.0% (16/20)	78.6% (11/14)
MOY	73.5% (25/34)	22.2% (2/7)	92.0% (23/25)

Overall prediction accuracy was relatively high, between 70%-80% for all three models. However, accuracy was noticeably imbalanced between groups. For the K-group model, the smaller Group 1 had a prediction accuracy of only 50%. For MOY-groups the smaller Group 0 had a prediction accuracy of 22.2%. It seems from this that there may be some bias in the models towards the group with the larger sample size. This could be solved both by using a much larger overall sample size, and by equalizing sample sizes between groups in future studies.

5.3 Generalized Manual Approach

Part of the fundamental purpose of this research is to help develop widely applicable methods for local transportation planning agencies to enhance their planning processes. Relatively complex statistical analysis requiring somewhat large and detailed datasets may be less useful in this sense than a more generalized approach which uses local knowledge and more readily available information. Such an approach may in fact be more than sufficient for the purposes outlined here.

The following sections describe the establishment of empirical factor groups for the 34 Pima County locations based on their average K, TWT, and MOY factor values. Then, these groups are evaluated under a quasi-subjective framework using data gathered manually from Google Maps aerial and street view imagery. Ideally, it would be possible to translate these evaluations to a generalized, manual approach to assigning STCLs to an appropriate factor group, for the purpose of estimating AADB using a set of expansion factors calibrated from PCLs which have similar temporal cyclist patterns. However, the attempts to do so below find mixed results, with some factor groups exhibiting more obvious relationships to their physical surroundings than others.

5.3.1 Data Collection

A short set of variables was established, based on previous work and on expectations of which land use variables would be most relevant to temporal patterns, as well as considering the practicality of assessing those variables from Google maps imagery. This evaluation framework is outlined in the subsequent sections. Note that this framework is subjective in nature, predicated on researcher judgement, and could be improved through local knowledge.

5.3.1.1 Commercial / Employment Uses

Intersection surroundings were scanned for major commercial or employment land-uses, such as large shopping plazas, malls, or office complexes. Consideration was given to the expected sphere of influence or different land-uses. The size of commercial generators was assumed to be positively related to their influence on cyclist traffic patterns, with larger commercial or employment centres assumed to have a larger sphere of influence. As a general rule, one major block (or approximately 750 – 900 meters) was considered the limit in terms of direct impacts. For example, the count location indicated by the red chevron in Figure 30 below (Location 36: Camino de la Tierra & Ina Rd) was considered to be in the sphere of influence of the large commercial plaza at the next intersection to the east, especially given that the plaza extends nearly halfway down the block towards it.

Figure 30: Google Maps. (2021) [W Ina Rd & N Camino De la Tierra, Casas Adobes, AZ]

5.3.1.2 Educational / Institutional Uses

Intersection surroundings were also scanned for educational and institutional land uses, such as schools, libraries, and community centres. Note that elementary schools were not considered, both because they are much smaller than junior high or high schools, and so aren't expected to have significant impacts on travel demand; and because most of the count locations are at major intersections, which young children wouldn't be expected to pass through even if they were cycling to school. Also note that healthcare, such as hospitals or medical centres, were considered to be "employment" and not "institutional" land uses, and so were captured under the "Commercial / Institutional" land use category. Similar 'sphere of influence' guidelines were applied as was done for commercial land uses. Larger schools, especially high schools or post-secondary institutions, were assumed to have the largest sphere of influence, up to an approximate maximum of one major block.

5.3.1.3 Recreational Uses and Trail Access

The final land-use category which was scanned for at study locations was recreational generators, such as large parks and conservation areas. Also considered were access points to major trail systems, such as the "Tucson Loop River Path" system. In some cases, consideration was given as to whether the count location was on-route to a major recreational destination. For instance, *Location 97: Colossal Cave Rd & Mary Ann Cleveland Wy* is well into the south-eastern exurbs of the Tucson area; there are no major generators at or within a wide radius of the intersection, but the location is close to the urban boundary and on route to several major recreational areas, such as the "Colossal Cave Mountain Park". The surrounding area was thus classified as 'recreational' in nature, and this effect could indeed be seen in the count data which exhibited a recreational pattern.

5.3.1.4 Relative Geographic Area

Study locations were generally clustered in three semi-distinct areas: Casas Adobes/North Tucson, the Catalina Foothills, and South-East Tucson. Each of these areas had somewhat distinct characteristics and land-use patterns (as per the researcher's judgement) which were thought to possibly have some influence on traffic patterns:

- *North Tucson*: Generally flat, mid-density suburbs mixed with large commercial plazas. Not close to the urban boundary and associated recreational areas but interspersed by an extensive paved multi-use path network following the Rillito River basin.

- *Catalina Foothills*: Most count locations located along a single west-east roadway corridor, Skyline Drive / Sunrise Drive with some intermittent commercial uses surrounded by extremely low density, high income residential land-uses. Characterized by hilly topography, and close to the north-eastern urban boundary and the associated recreational areas.
- *SE Tucson*: Somewhat industrial or heavy-commercial land uses in the areas around and between Tucson International Airport and the Davis-Monthan Air Force Base. Limited recreational land uses.

5.3.1.5 Bicycle Level of Service

A large number of bicycle level of service (BLOS) frameworks have been proposed by various researchers as ways to summarize the level of comfort which cyclists feel on a given road segment or at a given intersection. A very subjective, researcher dependent framework was applied here to score intersections on a scale from A (most comfortable) to F (not comfortable at all). The scale could also be expressed as a cyclist's willingness to choose a route featuring a given road segment in the presence of some other hypothetical option. Ultimately, these measures are a matter of the opinion of the researcher, which introduces some considerable bias.

The major and minor approach at every intersection were assessed separately, taking into consideration factors such as the number of lanes, road speed, road geometry, perceived traffic volume and composition, and the presence of bike lanes. An average BLOS for each count location was assigned as the average BLOS of the major and minor approaches (rounded up). Table 37 shows the subjective BLOS framework applied in this study.

Table 37: Bicycle Level of Service Framework

BLOS	Comfort Level	Examples
A	Completely comfortable, would always choose	Physically separated bicycle infrastructure of some kind is available, for example a protected cycle track Low volume, traffic calmed local road
B	Less comfortable, but would usually choose	Painted bike lanes on a low volume, 2-lane collector
C	Comfortable, but would prefer another option	Painted bike lanes on a medium volume, 2-lane collector Wide shoulder on rural road
D	Uncomfortable, but would choose if necessary	Painted bike lanes on 4-lane, medium volume collector 2-lane street with no bike lane or a narrow shoulder
E	Very uncomfortable, would almost never choose	4-lane, medium volume collector with no bike lane or wide shoulder 4-lane high-volume, high-speed arterial with painted bike lanes
F	Not comfortable at all, would never choose	High volume 6+ lane road, with or without bike lanes or painted shoulder, high % of heavy vehicles, merging lanes and slip-lanes for right turns are present

It is expected that different cyclist “classes” will prefer different types of roadways, contributing to observed differences in traffic patterns. For instance, recreational cyclists might prefer higher BLOS roads while utilitarian cyclists may not have a choice but to choose a given route. One of the difficulties with this sort of analysis is that it is extremely rider dependent. A seasoned cyclist will exhibit much higher levels of comfort than an inexperienced one. Another problem, in the Pima County context, is that essentially all the study roads are in the D-E-F range. There is relatively little variation in BLOS between the 34 study locations, at least under this subjective framework.

5.3.2 Evaluation of K-factor Groups

K-factors represent the ratio of average 8-hour cyclist volume – the hours for which TMCs are normally collected; 7:00-9:00, 11:00-14:00, and 15:00-18:00 – to average daily 24-hour cyclist volume. K-factors for Pima-County locations were calculated just for STC days, Tuesday, Wednesday, and Thursday (TWT), so they do not reflect weekend hour-of-day patterns. In previous research on applying a K-factor to non-motorized traffic, count locations were categorized subjectively using a 95% confidence interval plot of K-factor values [23]. The resulting groups were labelled low, medium, and high, with no attempt to further explain the variation between groups, although it is noted that most of the variability in K-factor values can be attributed to the difference between weekends and weekdays. Miranda-Moreno [11] uses an AM-Midday index (AMI) to help classify PCLs based on their hour-of-day profiles, with high AMIs assumed to represent ‘utilitarian’ travel patterns, and low AMIs ‘recreational’ patterns.

Although the K-factor differs significantly from the AMI in its formulation, it might be thought of similarly here. A high K-factor value would indicate that a large proportion of weekday cyclist volume occurs during the 8-hour ‘peak’ period. Trips made during this period can be assumed to be utilitarian in nature, so that a location with high K-factors could be labelled as “utilitarian”. The inverse would also be true, with low K-factor locations seeing more of their volume outside of workday hours, indicating that that location might be “recreational”.

Locations were empirically clustered using their K-factor values for the purposes of STCL group modelling using OLS and logit regression, as outlined in Sections 5.1 and 5.2. Under a binary grouping approach, two semi-distinct groups were established. The average K-factor across one group was 0.55; this group might be deemed “recreational”. The other group had an average K-factor of 0.66 and might be deemed “utilitarian”. A two-sample T-test comparing these two groups showed the difference between their means to be significant ($T[32]=7.71$, $p=8.54E-09$).

However, although these two factor groups could be identified empirically in the Pima County locations, it is not evident that they have any relationship to the locational characteristics outlined in Section 5.3.1. One would expect that the utilitarian group would be more commonly surrounded by commercial land uses, and the recreational group recreational uses, but this sort of relationship is not apparent. Table 38 below shows the average K-factors across land use categories. Remember that land use categories are non-hierarchical, so a location may have all three of commercial, educational, and recreational land uses present (or it may have none of the three). There is very little difference between the average K-factors across locations where a given land-use is present versus locations where it is absent, indicating that none of the land use classifications examined are correlated with the proportion of 8-hour to 24-hour cyclist volume at a location. Means testing also failed to find significant differences between the mean of K-values across locations within different land use classifications.

Chi-square tests were also used to check for association between empirical groups and different land use variables (Table 39). In no case was a significant association found.

Table 38: Means testing for average K-factor values across land-use variable categories

Land Use Variable	**Category**	**Avg. K-Factor**	**T-test (p)**
Commercial/ Employment	Present	0.62	0.758
	Absent	0.61	
Educational/ Institutional	Present	0.62	0.933
	Absent	0.62	
Recreational	Present	0.60	0.335
	Absent	0.63	
Area	1. SE Tucson	0.64	1v2: 0.885
	2. Catalina Foothills	0.63	1v3: 0.391
	3. North Tucson	0.64	2v3: 0.233
BLOS	A/B/C	0.63	0.403
	D/E/F	0.62	

Table 39: Contingency tables, association testing for land use and empirical grouping categories

Land Use		**Group** *Recreational*	*Utilitarian*	**Chi-square (p)**
Commercial	Present	7	13	0.966
	Absent	5	9	
Institutional	Present	4	6	0.711
	Absent	8	16	
Recreational	Present	4	4	0.320
	Absent	8	18	
Area	SE Tucson	2	3	0.128
	Catalina	1	9	
	N Tucson	9	10	
BLOS	A/B/C	5	8	0.761
	D/E/F	7	14	

None of the composite land-use categories identified in the evaluation framework seem to be uniquely related to one or the other K-factor group. For example, although 13 of 21 (61%) of *utilitarian* locations are within one major block of a commercial land use, 7 of 5 *recreational* locations are also within one major block of a commercial land use. Certain patterns seem to match expectations – for example *utilitarian* locations are more often on low BLOS roadways, which would make sense if utilitarian trips are assumed to be non-discretionary – but these patterns cannot be shown statistically to be correlated with the K-group variable.

This represents a substantial barrier to proposing a generalized approach for matching STCLs to K-factor groups. It is not apparent from these results that, in truth, there is more than one factor group in the Pima County sample. Given the relatively small range of observed K-factor values - most locations see between 30% and 50% of their daily volume during the 8-hour period – it may be acceptable to treat these locations as a single group, although this can be confirmed by measuring MAPE for AADB estimation under this condition.

5.3.3 Evaluation of TWT-factor Groups

Tuesday-Wednesday-Thursday (TWT) factors represent the ratio of average TWT cyclist volume to AADB. In this sense, an average TWT factor value for a location can be thought of in a very similar way to the Weekend-Weekday Index proposed by Miranda-Moreno et al. [11]. An average TWT greater than one indicates that weekday cyclist volumes are higher than weekend, in the same way as does a WWI value less than one. An average TWT less than one indicates a location sees more weekend than weekday cyclists, as does a WWI value greater than one. Miranda-Moreno describes count locations as recreational (those with high relative weekend volumes), utilitarian (low relative weekend volumes), mixed utilitarian, or mixed recreational. Of the three factors tested, TWT exhibited the most evident relationships with its surroundings.

One issue with using TWT factor values to classify count locations is that variability in the average values is suppressed, making it more difficult to identify empirical groups. WWI is calculated directly as the ratio of average weekend to weekday volume. A TWT factor though is calculated as average TWT cyclist volume over the AADB, which in turn includes both weekend and weekday volumes in its calculation. Consider two PCLs as an example, one with very high weekend volume and one with very low weekend volume. The relative difference between the WWI calculated for these two PCLs will be very high compared to the relative difference between their two TWT factor values.

Comparing WWI values makes it very clear that there is a difference between the locations but comparing TWT values less so. In a sample of 37 automated counting stations, Miranda-Moreno [11] observed a minimum WWI of 0.43 and a maximum of 2.26. The average WWI for the utilitarian group was 0.65, and 1.97 for the recreational group. Clearly, these two groups exhibited substantially different traffic patterns, giving little doubt that at least two factor groups exist. In the sample of 34 PCLs being examined here, there is a minimum average TWT value of 0.68 and a maximum of 1.01, a range of just 0.33, making it much more difficult to identify pattern groups.

Nevertheless, for the purposes of this study, locations were divided using the K-means algorithm. The result was two empirical factor groups, a "weekend" location with an average TWT value of 0.79; and a "weekday" group with an average TWT value of 0.97. A two-sample students t-Test on the two groups indicates that we can be highly confident that they represent two different populations with some intrinsic differences ($T[32] = -8.919$, $p < 0.05$). An assessment of the two groups was done using the framework described in Section 5.3.1, which is summarized below:

Weekend

- *Weekend* locations are those with low TWT values (between 0.68 and 0.86 for the Pima County sample), indicating that average weekday volumes are a small percentage of AADB. This factor group is analogous to the "recreational" patterns identified in the literature, which assume that weekend cycling trips are usually recreational in nature.
- The majority of weekend locations are *not* near a large school or institutional land use; of those that are, most are also nearby to major recreational uses boosting weekend traffic.
- In general, in the Tucson context, weekend locations are located close to the urban boundary, where recreational destinations abound. The Catalina Foothills area especially has a lower average TWT factor value than the other identified study areas.
- Weekend cyclist patterns appear to be more prevalent in suburban or exurban areas, often at lower volume, high BLOS intersections. Some consideration should be given to popular cycling routes if that sort of local knowledge is available.

Locations:

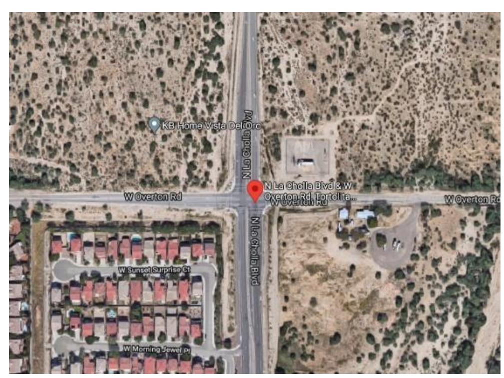
5, 7, 39, 41, 43, 49, 50, 52, 57, 59, 62, 69, 70, 71, 73, 81, 97, 100, 101, 105

Examples:

Location 49: Sabino Canyon Road and Sunrise Drive

Google Maps. (2021) [E Sunrise Dr & N Sabino Canyon Rd, Catalina Foothills, AZ]

Location 105: La Cholla Boulevard and Overton Road

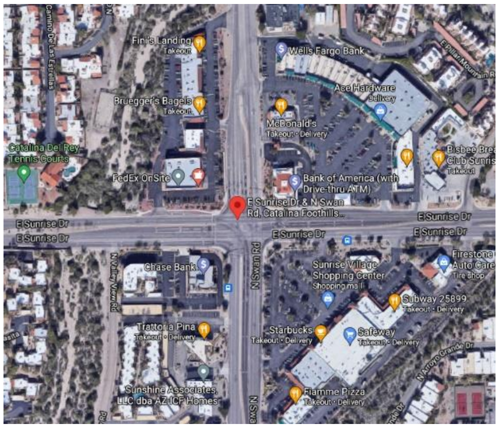


Google Maps. (2021) [N La Cholla Blvd & W Overton Rd, Tortolita, AZ]

	• Low volume, 2-lane roads with painted bike lanes. • Right on the urban boundary, with access to an enormous recreational area off the NE corner. • This location is the archetype for “weekend” count locations.	• Well into the lightly developed northern suburbs. • No major generators of any type, no real reason for utilitarian cyclists to travel through the intersection on an average weekday.
Weekday	• *Weekday* locations are those with high TWT factor values (between 0.89 and 1.01 for the Pima County sample), indicating that weekday cyclist volumes are close to or greater than the AADB. This group is analogous to the “utilitarian” pattern groups identified in the literature, with weekday trips usually assumed to be for a work, school, or retail purpose. • Locations in this group can primarily be identified by the presence of commercial or employment generators; 13 of the 14 PCLs in this group had land use of this type within one major block. • Although some *weekend* locations do have commercial uses nearby, these are also near to major recreational uses, or are further into the suburbs/exurbs where recreational cycling seems to be more prevalent. • Different types of cyclists respond differently to environmental stimuli; recreational trips, being discretionary, are more likely to be put off by weather, for instance [50]. In the same way, we might assume that utilitarian *weekday* cyclists are more willing to use the roads with the lowest BLOS, as they are making non-discretionary trips. This would appear to be supported by the Pima County data; *weekday* patterns seem to be more likely to appear at intersections with a BLOS below D.	*Locations:* 6, 9, 13, 34, 36, 38, 42, 53, 72, 76, 77, 80, 85, 96

Examples:

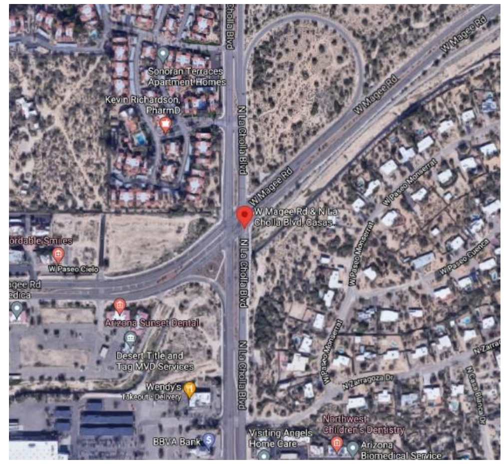
Location 38: Sunrise Drive and Swan Road



Google Maps. (2021) [E Sunrise Dr & N Swan Rd, Catalina Foothills, AZ]

- Low BLOS intersection with large commercial plazas at all corners.
- Area is prone to recreational cycling (Catalina Foothills), but this location isn't close to the urban boundary or any large recreational areas.

Location 53: La Cholla Boulevard and Magee Road



Google Maps. (2021) [N La Cholla Blvd & W Magee Rd, Casas Adobes, AZ]

- Large mall directly to the South-West of intersection.
- BLOS of both approaches is F: high speeds, 6-lanes or more on all approaches, slip lanes and merging lanes on all approaches. There is no apparent reason why a recreational cyclist would ever plan a route through this intersection.

5.3.4 Evaluation of MOY-factor Groups

MOY factors represent the ratio of monthly average daily bicyclists (MADB) to AADB. As to my knowledge, there have been relatively few attempts to characterize month-of-year patterns for the purposes of factor grouping, with most research focusing on hour-of-day and day-of-week profiles [28]. Budowski [31] identifies two traffic pattern groups for PCLs in Winnipeg, Manitoba; a "Winnipeg" group with a higher proportion of July and August volume; and a "Winnipeg Post-Secondary" group, which had relatively higher September and October volumes. Pima County is radically different from Winnipeg in terms of climate, but a context-specific approach considering weather and schools may still be appropriate.

Empirical analysis identified three relatively distinct MOY factor-groups in the 34 Pima County locations, for data from January 1, 2020 to March 15, 2020. The first are those with stable month-over-month volume; the second are those with increasing volumes from January to March; and the third are those with decreasing volumes from January to March. These groups were evaluated using the framework from Section 5.3.1, the results of which are shown below.

Stable

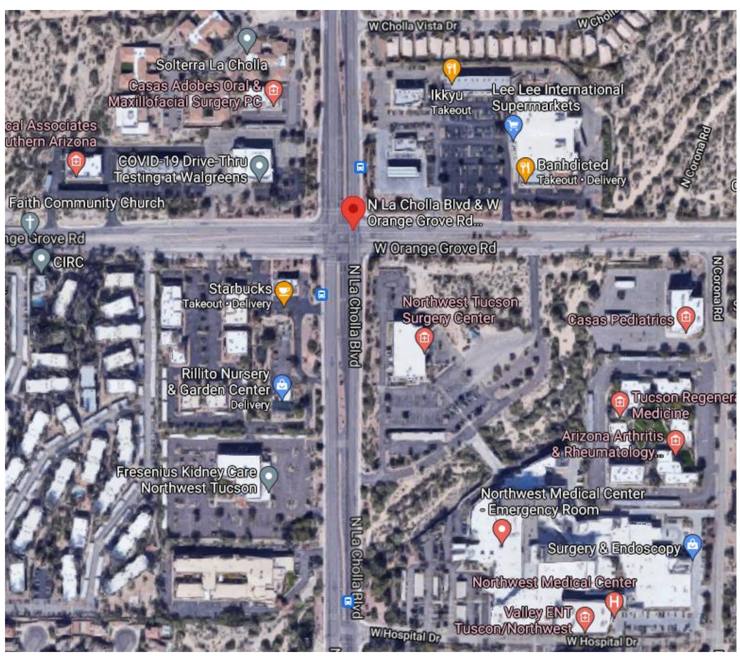
- The *stable* MOY pattern is characterized by stable month-to-month volumes, with MOY factor values for January, February, and March generally being within 0.1 of each other.
- It appears as though the locations with the most stable MOY volumes are those without any major recreational (14 of 19) or educational (also 14 of 19) land uses nearby. This would make some sense as these uses are temporally sensitive; cyclist traffic at an intersection with a nearby school would be expected to vary with the school calendar, and traffic near a major recreational generator with changes in the weather. Traffic related to commercial land uses though would be expected to be the most stable, although the split of *stable* locations between commercial and non-commercial locations is 10 to 9.

Locations:

6, 7, 13, 34, 36, 39, 41, 42, 50, 52, 53, 70, 71, 72, 73, 76, 81, 101, 105

Examples:

Location 13: La Cholla Boulevard and Orange Grove Road



Google Maps. (2021) [N La Cholla Blvd & W Orange Grove Rd, Casas Adobes, AZ]

- Primarily commercial and employment uses surrounding the intersection (see: large commercial plazas and major health-care hub to South-East). These uses are stable trip generators, and would be expected to be less susceptible to MOY variation (somebody who has to bike to work in January probably also has to bike to work in March).

Increasing

- *Increasing* locations show an increasing month-over-month pattern. Only five locations in the Pima County sample exhibited this pattern, with an average January to March change in MOY factor values of 0.34.
- Increasing patterns seem to most often occur nearby to major recreational generators. This would correlate well to climate patterns, with average daily temperatures rising from 11^{o}C in January to 16^{o}C in March [20]. One would assume that this increase in temperature would incentivize increased outdoor activity.

Locations:

5, 38, 49,

59, 97, 100

Examples:

Location 97: Colossal Cave Road and Mary Ann Cleveland Way

Google Maps. (2021) [E Colossal Cave Rd & E Mary Ann Cleveland Way, Rancho Del Lago, AZ]

- Location is far out in the south-east exurbs of the urban region, on route to a major recreational destination (Colossal Caves Mountain Park); recreational activity would be expected to increase as weather warms up, resulting in the increasing MOY traffic pattern.

Decreasing

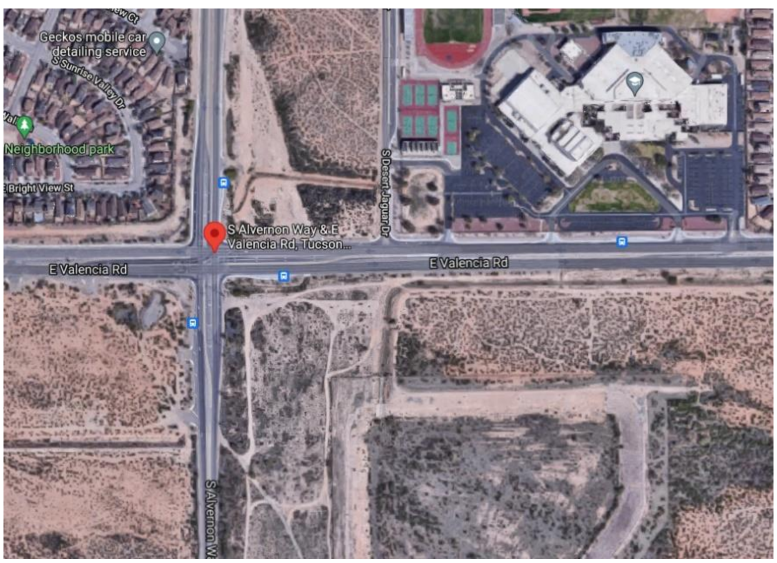
- The *decreasing* MOY pattern is characterized by declining volumes over the January-February-March period. Month-over-month changes in MOY factor values are generally between -0.1 and -0.2, with some extreme declines as high as -0.7. Average January to March change across the group is -0.43.
- Declining patterns appear to be most prevalent at locations that are *not* close to a recreational land use (8 of 9), or that *are* close to a school or institutional land use (6 of 9).
- It is difficult to say for sure that the decreasing pattern is related to the presence of schools, as the factors which would explain it should have been nullified by only using data up to March 15. The Pima County school district March break for 2020 was scheduled for the week of March 15, and the effects of the Corona virus on traffic patterns were not thought to substantially emerge until later towards the end of March.

Locations:

9, 43, 57, 62, 69, 77, 80, 85, 96

Examples:

Location 80: Alvernon Way and Valencia Road



Google Maps. (2021) [E Valencia Rd & S Alvernon Way, Tucson, AZ]

- Enormous high school ~500m to the east of this intersection; hard to say for sure that this is what is responsible for declining traffic pattern, especially since volumes at this location are very low to start with (indicating that not that many students are biking anyways).

One serious issue with the grouping schema identified above is that they pertain to patterns in *relative* MOY-factor values, but not the actual factor values themselves. To illustrate, consider Figure 31 and Figure 32 below. Figure 31 shows the average of MOY factors relative to January 2020 values, across each group. Illustrating the data in this way, the three distinct groups become very apparent. However, examining Figure 32 which shows the average of actual MOY factor values across groups, those patterns are no longer evident. Grouping locations based on their relative patterns does not necessarily minimize variation within groups, as is the goal of factor grouping. However, a more optimal grouping schema does not produce recognizable patterns which can be translated into a generalized approach.

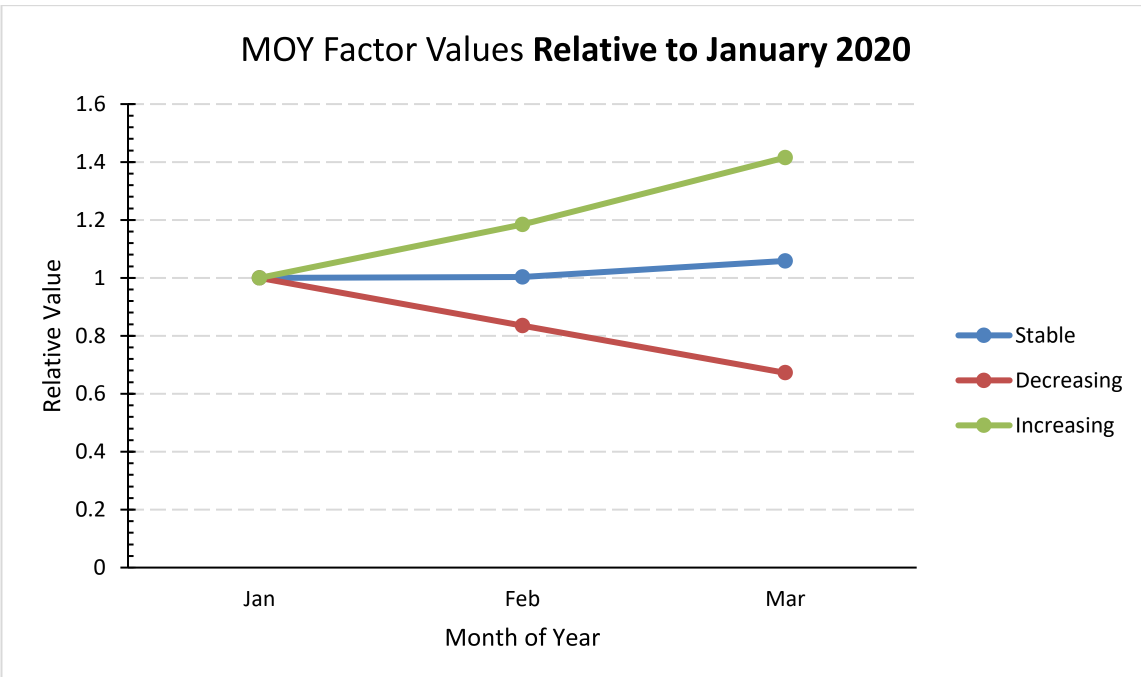


Figure 31: MOY factor values relative to January 2020

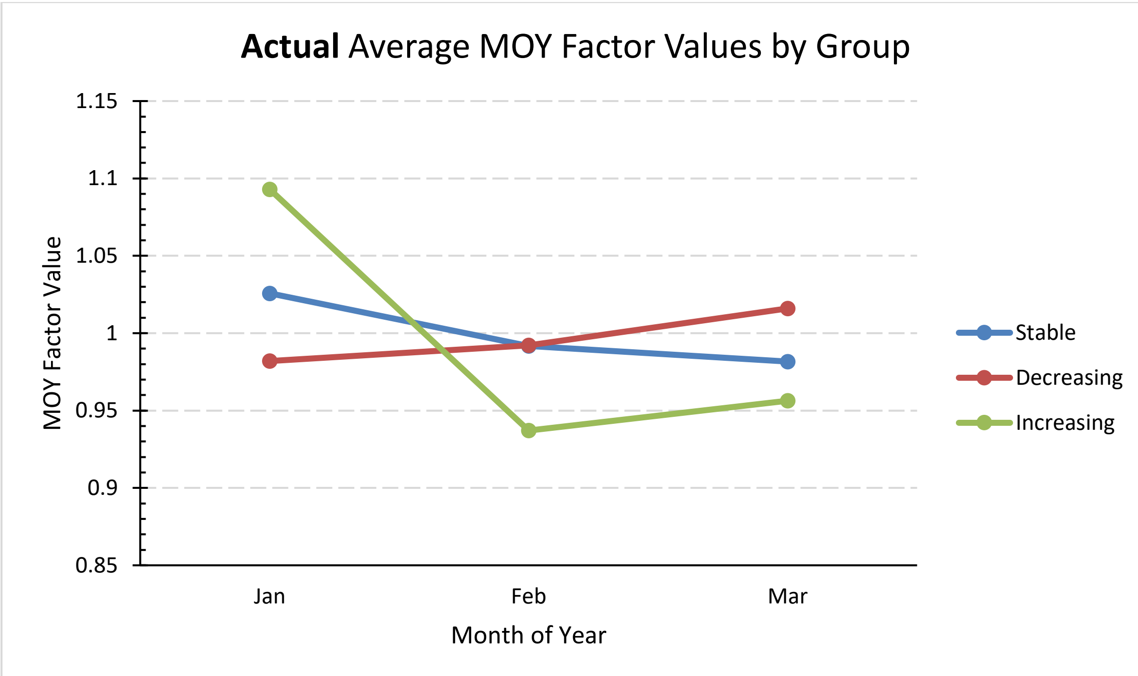


Figure 32: Actual average MOY factor values by group

5.3.5 Conclusions Regarding the "Generalized" Approach

Developing a generalized approach to grouping cyclist count locations according to their shared attributes proved to be somewhat difficult. Although there seem to be some relationships between locational and temporal characteristics of the 34 Pima County locations, there isn't enough evidence to say with certainty that they exist. Furthermore, none of these relationships are universal; we can at best say that some locational characteristic *might* indicate the presence of some temporal traffic pattern.

TWT-factors seem to be the easiest to categorize. There are some relatively clear relationships between land use and weekday-weekend traffic patterns. MOY and K-factor groups are harder to define. It is difficult to identify anything consistently linking the locations in each group. As was mentioned, it may be acceptable to use just a single factor group, especially for K-factors, but further study is needed to determine if this is the case.

Finally, one serious limitation to this methodology seems to be specific to the context in which it was developed. Pima County cyclist traffic, at least according to the sample data, is heavily recreationally biased. Most locations could be classified as fully recreational or mixed recreational-utilitarian under other grouping schema (see: Miranda-Moreno et al [11]). “Utilitarian” count locations are either entirely missing or at least severely underrepresented in this study, making it much more difficult to delineate clear land-use relationships.

Chapter 6
Conclusions and Recommendations

This thesis has examined the problem of estimating annual average daily bicycle (AADB) counts at intersections from short duration counts, specifically 8-hour turning movement counts. The problem can be decomposed into four sub-problems, namely (i) filtering data measurements to identify and exclude erroneous data; (ii) grouping continuous count sites to establish groups that have similar temporal characteristics (factor grouping); (iii) estimating AADB from short-term counts when the associated factor group is known; and finally (iv) determining the appropriate factor group for each short-term count site.

The insights and conclusions found from this research for each of these sub-problems are described in the following sections.

6.1 Filtering Turning Movement Counts

A set of filters for conventional cyclist count data was developed in accordance with the best practice identified in the literature. This included a novel approach to identifying erroneous runs of consecutive identical counts based on the compound probability of observing a given count *N* number of times in a row. However, in trying to apply these filters to a VMU dataset from Milton, Ontario, it was determined that the technological differences between conventional and VMU data collection equipment may make it inappropriate to apply conventional filters to VMU data. Accordingly, Filter 1 (F1) for “null” observations was substantially revised to check for the presence of a valid motor-vehicle observation; and Filters 2 and 5 (F2 and F5) for consecutive zero and 24-hour zero counts were removed outright. The remaining filters were found to be appropriate and carried through. The resulting set of four filters were used to check the quality of the Milton dataset, as well as an additional VMU dataset from Pima County, Arizona. This filtering procedure could be useful in any subsequent study looking to make use of VMU data, as standardized procedures for this data type have yet to be established to our knowledge.

6.2 Baseline Methods for AADB Estimation

A review of literature on AADB estimation from conventional short term cycling counts (STCs) identified three primary baseline factoring methods: a “traditional” day-of-week and month-of-year method (DOW/MOY), a day-of-week-of-month method (DOWOM) and a “disaggregate” day-of-year method (DOY). All three were applied to both the Milton and Pima County VMU datasets separately, using an iterative procedure that removed locations from the datasets one by one to act as short-term count locations (STCLs).

Estimation error on average across all Tuesdays, Wednesdays, and Thursdays in TMC months was extremely high for all three methods, for both Milton and Pima County, and significantly higher than compared to average results from the literature. However, it is expected that this result can be at least partially explained by the extremely low cyclist volumes observed at nearly all of the Milton and Pima County intersections. The lowest AADB at a Milton count location was 10.4 daily cyclists; at a Pima County location, 4.7. For comparison, Nordback [25] classified any count station where AADB was less than 200 as a “low” volume location. This reveals a limitation in using TMC data for AADB estimation, which is that cyclist volumes are typically much lower than on the dedicated infrastructure where “conventional” counts are collected. Relative variation among counts will be higher when counts are lower, contributing to higher estimation error, and furthermore relative error will be higher even where absolute error is small. Future study might try to determine if conventional cyclist counts, which might be higher and more consistent, can be used to calculate factors which can be applied to 8-hour TMCs.

Some baseline factor grouping methods were also trialed. Based on the literature, it was expected that factor grouping would have a positive effect on estimation results. However, for the Milton dataset, factor grouping using temporal indices as suggested by Miranda-Moreno [11] actually increased estimation error. It is possible that the already small number of study locations was responsible for this result. For Pima County, although a slight decrease in MAPE was observed as the number of factor groups increased, it could not be shown that there was any statistical difference between any of the trials. Further study on this subject using expanded datasets may be necessary.

Although Milton AADB estimates for “on-season” summer months were much more accurate than for the “shoulder” months, the lowest MAPE achieved using the DOW/MOY method was 40% for September, which is still prohibitively high. We can see then another significant limitation in the use of TMCs for AADB estimation, at least in a cold-climate North American context. It is commonly recommended in the literature that STCs be collected during the peak on-season months, when cycling activity is at its highest and most stable [25] [31]. In Milton, for instance, we would expect estimation error to be lowest in July and August, when counts are highest and day-to-day variation is at a minimum. However, in practice, TMCs would not be collected during these months. Standard procedures for TMC collection would mean that the STC used for AADB estimation would be just as likely to come from April, May, or November – when error would be very high – as it would be to come from June, September, or August – when error could be expected to be lower. It may be necessary when applying these methods to consider the relationship between the month-of-year that a TMC was collected and the expected AADB estimation accuracy; further study is needed to establish procedures for compensating for this effect.

6.3 Adapting Baseline Methods for Turning Movement Counts

Drawing from the conclusions made in Chapter 3 regarding the baseline estimation methods identified in the literature, several updates were proposed to the "traditional" day-of-week/month-of-year (DOW/MOY) method to address some of the core issues posed by TMCs to factor-based AADB estimation. First, it was proposed that a separate "K-factor" representing the average ratio of daily 8-hour to 24-hour cyclist volume be added. This was done in order to separately represent 8-hour/24-hour variation from day-of-week and month-of-year variation. Statistical analysis showed that 8-hour/24-hour factors are generally consistent across Tuesdays, Wednesdays, and Thursdays and across January-February-March in Pima County, indicating that a single, aggregated K-factor is appropriate (rather than calculating separate day-of-week or month-of-year K-factors).

Second, it was found that day-of-week (DOW) factors do not vary substantially across Tuesdays, Wednesdays, and Thursdays within given months in Pima County. This again indicates that a combined Tuesday-Wednesday-Thursday (TWT) factor is appropriate for expanding TMCs, which are generally only collected on these days. Finally, it was proposed that every permanent count location (PCL) should be assigned to three separate empirical factor groups for the purposes of calculating expansion factors, one for each of the K, TWT, and month-of-year (MOY) factor vectors. This differs somewhat from conventional methods, where a single grouping schema is usually established.

This updated method was tested using the Pima County VMU dataset using the same iterative procedure as was used to test the baseline methods. Comparing methods, the "proposed" method was found to be at least or more accurate than all three of the baseline methods. The proposed method also has the additional advantage of facilitating the linking of empirical and land-use factor groups – which is more difficult under the baseline methods.

Conclusions regarding the best factor group methods should be considered carefully, remembering that the data used for the analysis was very limited (both by the emergence of the Covid-19 pandemic in early 2020, and by overall low cyclist volumes). It would be useful to repeat the statistical analysis used to establish the K and TWT factors using more complete datasets, ideally from several different metropolitan regions.

6.4 Modeling Factor Group Characteristics

A limitation to factor-based AADB estimation methods commonly identified in the is that while it is relatively straightforward to establish empirical factor groups for permanent count locations (PCLs) for which multiple days, weeks, or months of data may be available, it is much more difficult to assign short-term count locations (STCLs) to the correct factor group. This is especially true for TMC locations, for which only 8-hours of count data is available. An attempt was made to address this limitation using three different methods.

A multi-variate linear modelling approach was tested on the Pima County dataset, trialing a large number of land-use variables. Under a two-group approach, the resulting models were able to predict the correct empirical factor group 76% of the time, although accuracy dropped significantly for the three and four-group scenarios. Similarly, a logit regression approach was tested, and the resulting models were ~75% accurate in predicting empirical factor group membership in a two-group scenario. These methods represent reasonable solutions to the problem of assigning STCLs to a factor group and make use of commonly available geospatial data.

As an alternative to statistical approaches, a generalized “manual” approach was proposed. Although it would be very difficult to evaluate the accuracy of this method, it was useful to show practically that temporal cyclist activity patterns can be linked to easily identifiable locational characteristics such as the presence of commercial land uses, or perceived traffic intensity of a given intersection.

Additional research may be needed to establish more transferrable guidelines for STCL factor group modelling. The work done here was highly context specific, and there is no reason to assume that the land use patterns correlated with a given activity pattern in Pima County are the same as in

Milton. It would be useful then to repeat the analysis described here using data from several different metropolitan regions, or even using a combined dataset containing data from different geographic areas.

6.5 Final Recommendations

This thesis outlines in detail a set of procedures by which data which is already being collected regularly in numerous transportation jurisdictions – specifically turning movement counts (TMCs) – could be used to generate estimates of annual average daily bicycle traffic (AADB) at any intersection where an up-to-date TMC is available. This work expands on existing methods which are already being applied in practice across North America [51], showing that there is potential for real-world application.

The main factor limiting the analysis and outcomes of this research is the limited set of data that was available. This was largely due to the significant impacts that the COVID-19 pandemic had (starting in March 2020) on travel patterns, including cycling patterns. As a result, the first key recommendation is to assemble a larger dataset that would include at least 40 sites per geographic/climactic region for which continuous count data are available for at least one year, and to have data from a minimum of three different regions.

Second it is recommended to explore the application of machine learning (ML) techniques for the estimation of AADB from TMCs. The techniques would provide the advantage that they are highly non-linear and do not require prior assumptions about the relationships between explanatory factors and AADB. This may permit the integration of both land use characteristics and data from continuous count sites into a single ML model to estimate AADB.

Third, it has been observed in this work that when the AADB is very low, measures of estimation error may be very high, suggesting that the AADB estimates are unreliable. However, from a practical engineering and planning perspective, it may be sufficient to know that the AADDB is low. Consequently, it is recommended to examine more robust measures of estimation errors.

Finally, it is recommended to examine how AADB estimates from TMCs at intersections can be utilized by engineers and planners (e.g. planning and prioritizing infrastructure and road safety studies) and can be integrated with measurements and AADB estimates for dedicated cycling infrastructure.

Bibliography

[1] Statistics Canada, *Commuters using sustainable transportation in census metropolitan areas,* Ottawa, Ontario, 2017.

[2] R. Beuhler and J. Dill, "Bikeway Networks: A Review of Effects on Cycling," *Transport Reviews,* 2015.

[3] A. C. Nelson and D. Allen, "If You Build Them, Commuters Will Use Them: Association between bicycle facilities and bicycle commuting," *Transportation Research Record,* vol. 1578, pp. 79-83, 1997.

[4] J. Dill and T. Carr, "Bicycle Commuting and Facilities in Major U.S. Cities," *Transportation Research Record No. 03-4134,* pp. 116-123, 2003.

[5] H. J. Mohammed, "Using Turning Movement Counts (TMC) at Intersections," University of Kansas Transportation Center for Road, Street & Bridge Agencies, Lawrence, Kansas, 2019.

[6] R. Schneider, "Comparison of Turning Movement Count Data Collection Methods for a Signal Optimization Study," URS Corporation, Grand Rapids, Michigan, 2011.

[7] OpenStreetMap contributors. (2021) Planet dump. Retrieved from https://planet.openstreetmap.org..

[8] National Academies of Sciences, Engineering, and Medicine, "Guidebook on Pedestrian and Bicycle Volume Data Collection," The National Academies Press, Washington, DC, 2014.

[9] K. Nordback, K. Tufte, N. McNeil, M. Harvey and M. Watkins, "Development of an Online Nonmotorized Traffic Count Archive," National Institute for Transportation and Communities, Portland, OR, 2017.

[10] S. Turner, T. Qu and P. Lasley, "Strategic Plan for Non-Motorized Traffic Monitoring in Colorado," College Station, TX, 2012.

[11] L. F. Miranda-Moreno, T. Nosal, R. J. Schneider and F. Proulx, "Classification of BIcycle Traffic Patterns in Five North American Cities," *Transportation Research Record,* vol. 2339, pp. 68-79, 2013.

[12] N. McNeil, K. Tufte, T. Lee and K. Nordback, "Biking and Walking Quality Counts: Using “BikePed Portal” Counts to Develop Data Quality Checks," Portland State University: Transportation Research and Education Center, Portland, Oregon, 2019.

[13] A. Medury, J. B. Griswold, L. Huang and O. Grembek, "Pedestrian Count Expansion Methods: Bridging the Gap between Land Use Groups and Empirical Clusters," *Transportation Research Record,* vol. Vol. 267, no. 5, pp. 720-730, 2019.

[14] E. Minge, C. Falero, G. Lindsey, M. Petesch and T. Vorvick, "Bicycle and Pedestrian Data Collection Manual," St. Paul, MS, 2017.

[15] A. Bergstrom and R. Magnusson, "Potential of transferring car trips to bicycle during winter," *Transportation Research Part A,* vol. 37, pp. 649-666, 29 January 2003.

[16] J. Mergel, "Case Studeis of Traffic Monitoring Programs in Large Urban Areas," Federal Highway Administration, Office of Highway Information Management, Washington, DC, 1997.

[17] "Enhancing Arizona Department of Transportation Traffic Data Resource," Arizona Department of Transportation, Phoenix, AZ, 2001.

[18] Pima County Transportation, "Roadway and Development Street Standards Manual," Tucson, AZ, 2020.

[19] Environment Canada, *Daily Data Report: OAKVILLE TWN,* Oakville, Ontario: Co-operative Climate Network, 2019.

[20] National Oceanic and Atmospheric Administration, *Daily Weather Summaries,* 2020.

[21] U.S. Department of Transportation, "Traffic Monitoring Guide," U.S. Department of Transportation, Washington, D.C., 2016.

[22] K. Nordback, "Estimating Annual Average Daily Bicyclists and Analyzing Cyclist Safety at Urban Intersections," Denver, 2012.

[23] M. El Esawey and A. Ibrahem Mosa, "Determination and Application of Standard K Factors for Bicycle Traffic," *Transportation Research Record,* vol. 2527, pp. 58-68, 2015.

[24] S. Hankey, G. Lindsey and J. Marshall, "Day-of-Year Scaling Factors and Design Considerations for Nonmotorized Traffic Monitoring Programs," *Transportation Research Record,* vol. 2468, pp. 64-73, 2014.

[25] K. Nordback, W. E. Marchall, B. N. Janson and E. Stolz, "Estimating Annual Average Daily Bicyclists: Error and Accuracy," *Transportation Research Record,* vol. 2339, pp. 90-97, 2013.

[26] T. Nosal, L. F. Miranda-Moreno and Z. Krstulic, "Incorporating weather: a comparative analysis of Average Annual Daily Bicyclist estimation methods," *Transportation Research Record: Joural of the Transportation Research Board,* vol. 2468, pp. 100-110, 2014.

[27] D. Beitel and L. F. Miranda-Moreno, "Methods for Improving and Automating the Estimation of Average Annual Daily Bicyclists," in *TRB 95th Annual Meeting Compendium of Papers*, Washington, D.C., 2016.

[28] 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, no. 10, pp. 295-310, 16 May 2019.

[29] M. Figliozzi, P. Johnson, C. Monsere and K. Nordback, "Methodology to Characterize Ideal Short-Term Counting Conditions and Improve AADT Estimation Accuracy Using a Regression-Based Correcting Function," *Journal of Transportation Engineering,* vol. 140, no. 5, 7 March 2014.

[30] M. El Esawey, "Toward a Better Estimation of Annual Average Daily Bicycle Traffic: Comparison of Methods for Calculating Daily Adjustment Factors," *Transportation Research Record,* vol. 2593, pp. 28-36, 2016.

[31] A. R. Budowski, "Developing Expansion Factors to Estimate Cyclist Seasonal Average Daily Traffic in Winnipeg, MB," Winnipeg, MB, 2015.

[32] J. B. Griswold, A. Medury, R. J. Schneider and O. Grembek, "Comparison of Pedestrian Count Expansion Methods: Land Use Groups versus Empirical Clusters," *Transportation Research Record,* August 2018.

[33] K. Nordback, W. Marshall and B. Janson, "Development of estimation methodology for bicycle and pedestrian volumes based on existing counts," Colorado Department of Transportation, Denver, 2013.

[34] D. Johnstone, K. Nordback and S. Kothuri, "Annual Average Nonmotorized Traffic Estimates from Manual Counts: Quantifying Error," *Transportation Research Record,* vol. 2672, no. 43, pp. 134-144, 2018.

[35] K. Nordback, S. Kothuri, G. Gibson, W. Marshall and N. Ferenchak, "Improving Bicycle Crash Prediction for Urban Road Segments," National Institute for Transportation and Communities (NITC), Portland, OR, 2018.

[36] G. Lindsey, M. Petesch and S. Hankey, "The Minnesota Bicycle and Pedestrian Counting Initiative: Implementation Study," Hubert H. Humphrey School of Public Affairs, Minneapolis, 2015.

[37] K. Nordback, W. E. Marchall, B. N. Janson and E. Stolz, "Estimating Annual Average Daily Bicyclists: Error and Accuracy," *Transportation Research Record,* vol. 2339, pp. 90-97, 2013.

[38] S. Olejnik and J. Algina, "Generalized Eta and Omega Squared Statistics: Measures of effect size for some common research designs," *Psychological Methods,* vol. 8(4), pp. 434-477, 2003.

[39] M. El Esawey, "Estimation of Annual Average Daily Bicycle Traffic with Adjustment Factors," *Transportation Research Record,* vol. 2443, pp. 106-114, 2014.

[40] J. B. Griswold, A. Medury and R. J. Scheider, "Pilot Models for Estimating Bicycle Intersection Volumes," *Transportation Research Record,* vol. 2247, pp. 1-7, 2011.

[41] S. Hankey, G. Lindsey, X. Wang, J. Borah, K. Hoff, B. Utecht and Z. Xu, "Estimating use of non-motorized infrastructure: Models of bicycle and pedestrian traffic in Minneapolis, MN," *Landscape and Urban Planning,* vol. 107, pp. 307-316, 2012.

[42] J. Strauss, L. F. Miranda-Moreno and P. Morency, "Cyclist activity and injiry risk analysis at signalized intersections: A Bayesian modelling approach," *Accident Analysis and Prevention,* vol. 59, pp. 9-17, 2013.

[43] D. J. Fagnant and K. Kockelman, "A direct-demand model for bicycle counts: the impacts of level of service and other factors," *Environment and Planning B: Planning and Design,* vol. 93, pp. 93-107, 2016.

[44] M. Tabeshian and L. Kattan, "Modeling Nonmotorized Travel Demand at Intersections in Calgary, Canada," *Transportation Research Record: Journal of the Transportation Research Board,* vol. 2430, pp. 38-46, 2014.

[45] L. F. Miranda-Moreno and T. Nosal, "Weather of Not to Cycle: Temporal Trends and Impact of Weather on Cycling in an Urban Environment," *Transportation Research Record,* vol. 2247, pp. 42-52, 2011.

[46] K. K. Mukoko and S. S. Pulugurtha, "Examining the influence of network, land use, and demographic characteristics to estimate the number of bicycle-vehicle crashes on urban roads," *IATSS Reaserch,* vol. 44, pp. 8-16, 2020.

[47] P. Peduzzi, J. Concato, E. Kemper, T. Holford and A. Feinstein, "A simulation study of the number of events per variable in logistic regression analysis," *Journal of Clinical Epidemiology,* vol. Vol. 49, no. 12, pp. 1373-1379, 1996.

[48] J. S. Long, Regression Models for Categorical and Limited Dependent Variables, vol. 7, Indiana University: SAGE Publications Inc., 1997.

[49] S. Nemes, J. M. Jonasson, A. Genell and G. Steineck, "Bias in odds ratios by logistic regression modelling and sample size," *BMC Medical Research Methodology,* vol. 9 (56), 2009.

[50] T. Thomas, C. Jaarsma and S. Tutert, "Temporal variations of bicycle demand in the Netherlands: The influence of weather on cycling," in *Transportation Research Board Annual Meeting*, Washington, D.C., 2009.

[51] K. Nordback and M. Sellinger, "Methods for Estimating Bicycling and Walking in Washington State," Washington State Department of Transportation, Olympia, Washington, 2013.

[52] Statistics Canada, *Commuters using sustainable transportation in census metropolitan areas,* Ottawa, Ontario: 2016 Census of Population, 2017.

[53] Institute of Transportation Engineers (ITE) Pedestrian and Bicycle Council, "Count Adjustment Factors," March 2009. [Online]. Available: http://bikepeddocumentation.org/.

[54] M. El Esawey, A. Ibrahem Mosa and K. Nasr, "Estimation of daily bicycle traffic volumes using sparse data," *Computers, Environment and Urban Systems,* vol. 54, pp. 195-203, 2015.

[55] J. Dowds and J. Sullivan, "Applying a Vehicle-Miles of Travel Calculation Methodology to a County-Wide Calculation of Bicycle and Pedestrian Miles of Travel," in *TRB Annual Meeting*, 2012.

[56] J. Cohen, *Statistical power analysis for the behavioural science,* 2nd ed ed., Hillsdale: Lawrence Erlbaum Associates, 1988, pp. 284-288.

[57] M. Haynes and S. Andrzejewski, *GIS Based Bicycle and Pedestrian Demand Forecasting Techniques,* 2010.

[58] R. Cervero, O. L. Sarmiento, E. Jacoby and L. F. Gomez, "Influences of Built Environments on Walking and Cycling: Lessons from Bogota," *International Journal of Sustainable Transportation,* vol. 3, pp. 203-226, 2009.

[59] R. Sherry, D. E. Sidelinger, S. Saitowitz, D. Browner, S. Vance and L. McDermid, "Designing and Implementing a Regional Active Transportation Monitoring Program Through a County-MPO-University Collaboration," *American Journal of Health Promotion,* vol. Vol. 28, no. No. 3, pp. 104-111, January/February 2014.

[60] K. Nordback, S. Kothuri, P. Theodore, P. Mcleod, E. Rose and H. Twaddell, "Exploring Pedestrian Counting Procedures," U.S. Department of Transportation, Washington, DC, 2016.