2018-05-23

Estimating the Impact of Weather Variables on Cycling Infrastructure Usage

Shahriari, Siroos

UNIVERSITY OF CALGARY

Estimating the Impact of Weather Variables on Cycling Infrastructure Usage

by

Siroos Shahriari

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN CIVIL ENGINEERING

CALGARY, ALBERTA

May, 2018

Abstract

Cycling has gained increased attention among both researchers and public city planners in the last two decades as a sustainable means of transportation. Compared to users of other modes of transportation, cyclists are more exposed to weather conditions. As a result, cycling infrastructure usage changes throughout the year because of highly variable weather conditions. The objective of this study is to evaluate the impact of weather variables on cycling infrastructure usage. The study develops a model that estimates changes in cycling infrastructure usage for different months of the year using weather variables namely, temperature, precipitation, snow on the ground, wind and sunlight. Cycling infrastructure usage data from the city of Calgary was modeled using generalized estimating equations (GEE) to suite unbalanced and correlated data. The model that was developed can help municipalities for cycling network planning such as decision making on investments, maintenance and modifications to the cycling infrastructure.

Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Farnaz Sadeghpour for her excellent guidance and support at each step of my research. This work would have not been possible without her guidance.

I would also like to thank the members of my examination committee, Dr. John Douglas Hunt, Dr. Alexandre de Barros, and Dr. Nigel Waters for their time and useful comments on my work.

I would also like to thank my friends who have been my family here, for all the great moments that we shared and all the great memories that I will never forget.

Last but not the least, thanks to my parents for their continuous support throughout my life.

Dedication

To my family and Sahba, for their constant and unconditional support.

Table of Contents

List of Tables

List of Figures and Illustrations

CHAPTER 1: INTRODUCTION

1.1 BACKGROUND

Cycling has gained increased attention among both researchers and public planners around the world in the last few decades as a sustainable means of transportation. The three main pillars of sustainability are environment, social, and economic (Bell and Morse, 2012). Increasing the share of cycling usage can help cities to move toward a more sustainable transportation system due to cycling environmental, social and economic benefits.

Cycling is an environmentally friendly mode of transportation as it produces no emissions, and therefore, increasing the share of cycling will help with current pressing matters such as global warming. The Intergovernmental Panel on Climate Change (IPCC) has announced that using a combination of walking, cycling, and public transit for transportation can reduce greenhouse gas emissions from vehicles by 25% at a cost of only US $33 million per tonne of CO_2. In contrast, reductions caused by the use of high-efficiency vehicles would cost approximately US $110 million per tonne of CO_2 (Publique Toronto, Santé, 2012).

In addition to environmental benefits, cycling is a cost-efficient mode of transportation. Maintenance fees are lower compared to the fees for vehicles and using bicycle for transportation does not consume fossil fuel. In addition to cycling's low usage cost, investment in cycling infrastructure has economic benefits. Research has shown that investing in cycling infrastructure would provide more jobs per dollar compared to investments in vehicle-only infrastructure. Approximately 11.5 local jobs are created for each $1 million spent on cycling infrastructure, compared to investments on vehicle-only infrastructure which creates 9.6 local jobs per $1 million (Garrett-Peltier, 2011).

An increase in cycling usage will lead to social benefits for users as it increases social interaction and livelihood on the streets (Publique Toronto, Santé, 2012). Cycling provides greater mobility for people who do not own cars. Increasing cycling usage in a neighbourhood will lead to safer roads because of a reduction in vehicular traffic. As a result of lower vehicular traffic, the neighbourhood becomes less dangerous and safer for children and elderly people (Jacobsen 2013). In addition, cycling has several health benefits for users (Deenihan and Caulfield, 2014; Cavil and Davis, 2007). Despite cycling benefits, share of cycling trips around the world is still low compared to other modes of transportation. Table 1.1 shows the share of cycling in all transportation trips according to Pucher and Buehler (2008) for different countries around the world. As shown in the table, the highest cycling mode share is attributed to European countries, such as the Netherlands and Denmark, while North American countries, such as Canada, have a low share of cyclists. As a result, due to cycling's above-mentioned benefits, countries with a low share of cycling are trying to increase the share of cycling as a mode of transportation (Pucher et al., 2010).

Table 1.1 Bicycle share of trips around the world (Pucher and Buehler 2008)

Country	Cycling mode share (%)
Netherlands	27
Denmark	18
Finland	11
Sweden	10
Germany	10
Belgium	8
Switzerland	6
Austria	5
Norway	4
France	3
Italy	3
Ireland	2
Canada	2

1.2 PROBLEM STATEMENT

In the last few decades, transportation planning efforts in North America have tried to increase the share of cycling as an active mode of transportation. Research has shown that a more expanded network of cycling infrastructure with better cycling facilities (e.g. parking facilities) attracts more people to cycle (Dill and Carr, 2003; Titze et al., 2008). Hence, to encourage more people to cycle, municipalities have improved cycling infrastructure network conditions (i.e. providing more facilities for existing cycling infrastructure) and expanded the infrastructure network (i.e. adding new infrastructure sections to the existing network).

The available transportation budget is limited within municipalities. To optimize the use of the budget, typically cycling's total kilometers traveled is used when making decisions on allocating budget monies for cycling infrastructure. The total kilometers traveled for each infrastructure section (a part of infrastructure separated from other parts with intersections) is computed by multiplying the infrastructure section annual usage with length of the infrastructure section. The infrastructure annual usage is used to identify where to spend money to improve or expand the infrastructure network more efficiently. While using infrastructure usage collected for an entire year is more reliable for decision making, currently short count usage (partial data) is used. Short counts are collected for a limited period of the time (i.e., one month). The annual usage is calculated by scaling up short count usage with usage estimate factors. Usage estimate factors are calculated based on previous usage data and represent the ratio of usage of each month to the annual usage. The reason for using short count data and estimate factors is that collecting infrastructure usage for an entire network throughout a year is not economically feasible. Data collection requires either employees assigned to each infrastructure section for an entire year to count cyclists or numerous bicycle counters installed across the network. The cost of a counter varies depending on their functionality. The cost of a counter starts from $8,000 CAD for counters installed on cycling pathways counting only cyclists to $30,000 CAD for counters installed on full streets counting cyclists, pedestrians and cars.

While scaling of short count usage is currently used for decision making, it has its limitations for decision making as it does not consider the usage variation throughout a year. Usage variation throughout the year is needed in cities around the world. For example, in cities with cold climates such as most Canadian cities there is always a debate on whether to keep or remove cycling infrastructure during the cold months of the year. Because it is assumed that cycling infrastructure will remain unused or will have a low usage during cold months.

As literature showed and it will be tested in chapter 3, the cause of usage variation of cycling infrastructure is due to the effect of weather conditions. Therefore, a more reliable and accurate method is required to estimate the cycling infrastructure usage and the variation of usage throughout the year using weather variables.

1.3 RESEARCH OBJECTIVE

The objective of this study is to develop a methodology for estimating impact of weather on cycling infrastructure usage and estimate the usage with short count data (monthly basis). To achieve this objective, two sub-objectives are identified as follows:

1. Identify a methodology to evaluate the magnitude of impact of weather variables on cycling infrastructure usage for each month of the year.

2. Develop an estimation model to estimate cycling infrastructure usage throughout the year based on weather variables using short count.

1.4 RESEARCH METHODOLOGY

To identify the methodology for evaluating the impact of weather variables on cycling infrastructure usage (sub-objective 1), data from the city of Calgary was used. Impact of five weather variables namely temperature, precipitation, snow on the ground, wind speed and sunlight hours was assessed on cycling infrastructure usage. Data was collected with 26 counters mainly located in downtown Calgary. Data included around 15000 points where each point represents the usage of one infrastructure section for one day. The impact of above-mentioned weather variables was assessed separately on weekday and weekend usage. Due to data nature (correlated over time with unbalanced categories) Generalized estimating equations (GEE) was chosen to create a statistical model to evaluate the impact of weather variables on cycling infrastructure usage.

Weather impact was computed separately for weekday and weekend usage. The proposed estimation model was developed (sub-objective 2) with defining criteria for selecting between weekday and weekend weather coefficients to estimate the usage for specific infrastructure section. Weather coefficients were selected for each infrastructure using a selection criterion that was calculated for infrastructure sections based on their type of users.

The model was developed with data from the core area of the city of Calgary. To validate the developed model, a group of infrastructure sections locating across the city (core and outside core area) was used. The estimation model was validated for both core and outside core area of the city to find if the model is capable of estimating the usage for any location in the city and determine that it is not limited to the core area.

1.5 THESIS ORGANIZATION

This thesis consists of 5 chapters. Chapter 2 presents an overview of previous studies related to variables that can impact cycling and cycling infrastructure usage trends. The first part of the chapter discusses studies that investigate the impact of variables on cycling with focus on weather variables. The second part of the chapter focuses on studies that investigate cycling infrastructure usage trends and studies that develop models to estimate usage.

Chapter 3 initially elaborates on the first sub-objective of this study and identifies the weather variables that affect cycling infrastructure usage throughout a year. Then, the magnitude of the impact of those weather variables on infrastructure usage is evaluated and discussed. Then, two estimation models are developed based on infrastructure weekend and weekday usages.

Chapter 4 presents applications of the developed models and the process for selecting the best model to estimate usage. Then, the chapter validates the accuracy of proposed models.

Chapter 5 summarizes the results of the study with contributions and discussion on this work. The chapter also presents the limitations of the study and recommendations for future work.

CHAPTER 2: LITERATURE REVIEW

This chapter presents a review of past studies that have been conducted on cycling as one of the active modes of transportation. The first part of this chapter goes through studies on the variables that impact cycling and how they impact cycling; particular emphasis is given to the affect of weather variables. The reason for this emphasis is due to the affect of weather variables on cycling infrastructure usage throughout the year which will be explained at the end of the chapter. The second part of this chapter reviews studies that investigate cycling infrastructure usage trends and prediction models, which have been developed to predict cycling infrastructure usage.

Past studies that have been conducted on cycling can be categorized into three categorise. The first category includes studies that focus on crash risk, accidents and fatalities related to this mode of transportation (Maring and Schagen, 1990, Stutts and Hunter, 1999, Hall and Kaltenecker, 1999). The second category includes studies that focus on variables that impact cycling while the last category contains studies that develop estimation models to estimate the usage of this mode of transportation. In this chapter the focus is on the second and third categories referred to as variables that impact cycling and usage estimation models. Figure 2.1 shows the map of literature conducted on cycling.

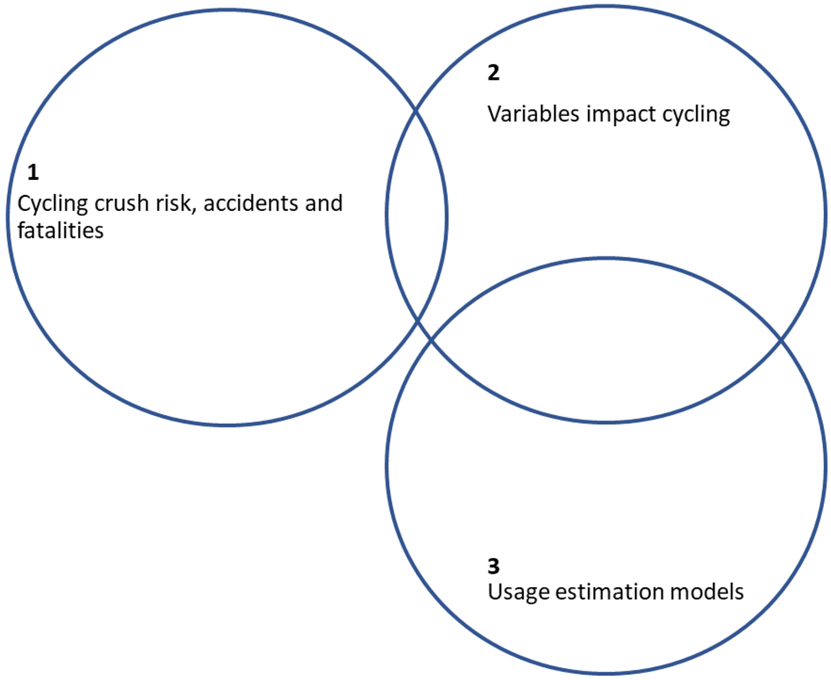


Figure 2.1 Map of literature conducted on Cycling

2.1 VARIABLES IMPACTING CYCLING

Studies have shown that several variables affect cycling. In this chapter, these variables are categorized into five groups: cyclist characteristics, attitude toward cycling, cycling infrastructure, built environment and network connectivity and weather conditions. Table 2.1 provides a summary of the studies that examine the effect of these variables on cycling. Each group of variables and the associated literature will be discussed in detail.

Table 2.1 An overview of the literature on the variables that affect cycling frequency

Article	Cyclist characteristics	Attitude toward cycling	Cycling infrastructure	Built environment and network connectivity	Weather condition
Zhao et al. 2018					✓
Meng et al. 2016					✓
Amiri and Sadeghpour 2015	✓	✓			✓
Spencer et al. 2013			✓		✓
Amiri and Sadeghpour 2014	✓	✓	✓		✓
Wegman et al. 2012	✓				
Buehler 2012			✓		✓
Flynn et al. 2012	✓				✓
Caulfield et al. 2012	✓	✓		✓	
Li et al. 2012			✓	✓	
Saneinejad et al. 2012					✓
Larsen and El-Geneidy 2011	✓		✓		
Twaddle et al. 2011	✓	✓	✓	✓	
Heinen et al. 2011		✓			
Daley and Rissel 2011		✓			
Heinen et al. 2010	✓	✓	✓	✓	✓
Parkin and Meyers 2010			✓		
Vandenbulcke et al. 2009	✓				
Dill 2009	✓	✓	✓		
Titze et al. 2008	✓	✓	✓	✓	
Garrard et al. 2008	✓				
Sisson and Tudor-Locke 2008	✓				
Parkin et al. 2008	✓		✓		✓
Walker 2007	✓				
Hunt and Abraham 2007	✓		✓	✓	
Gatersleben and Appleton 2007	✓	✓			
Brandenburg et al. 2007					✓
Plaut 2005	✓				
Krizek and Roland 2005			✓		
Brandenburg et al. 2004					✓
Dickinson et al. 2003	✓		✓		
Bergstrom and Magnusson 2003	✓	✓	✓		✓
Nankervis 1999					✓
Osberg et al. 1998					✓
Moritz 1997	✓		✓		
Rodgers 1997	✓				
Rodgers 1995	✓				✓

2.1.1 Cyclist Characteristics

Cyclist characteristics includes variables that describe or relate to cyclists' demographics such as gender, age, owning a bicycle, education, annual income and variables such as travel distance and trip duration. The following sections present some of the important cycling characteristics that impact cycling.

Gender

Majority of studies investigated the impact of gender on cycling found that the percentage of male cyclists was usually higher than female cyclists (Garrard et al., 2008; Dickinson et al., 2003; Gatersleban and Appleton, 2007; Titzea et al., 2008). Studies found men were more likely to be regular cyclists than women. Also, cyclists had different preferences based on their gender. Researchers in Melbourne conducted a census of cyclists that showed about 80% of cyclists were men. One of the main reasons that may explain the difference between the number of male and female cyclists was safety concerns about the roads and facilities; this concern was higher among female cyclists. In addition, the study found that females had a lower mean travel distance than males. Females also showed higher preferences for bicycle pathways compared to other bicycle facilities such as on-road lanes or streets without bicycle facilities (Garrard et al., 2008). Another study from Calgary, Canada that surveyed downtown commuter cyclists corroborated the findings of Garrard et al. (2008) and showed that more than 75% of the cyclists who commuted to downtown were male. The study also showed that the main barrier that prevented women from cycling was road safety, while men had concerns about destination facilities such as parking, showers, and changing rooms (Twaddle et al., 2010).

A case study from the United Kingdom (the UK) conducted on commuters of three companies in Hertfordshire showed that women had shorter travel distances (the mean travel distance to work was 8.3 miles) compared to men (12.5 miles). While women lived closer to their destination and a larger portion of them believed that they lived near enough to cycle, they were significantly less likely to cycle compared to men. The women surveyed cited personal security during the journey, ease of use, and dropping off and picking up children as the main reasons for using a car (Dickinson et al., 2003). Another study from the UK surveyed people from the University of Surrey and categorized respondents into 5 groups. The respondents from the first group had never used a bicycle (never-cyclists) and respondents in last group used a bicycle regularly (regular cyclists). Sixty-four percent of respondents from the never-cyclist group were women, while only 24% of the respondents from the last group were women. The result indicated that females were more likely to be never-cyclists, while men were more likely to be regular cyclists (Gatersleban and Appleton, 2007). This finding agreed with the results from Dickinson et al. (2003). In contrast, some case studies did not find a significant difference between the number of male and female cyclists. For example, a study from Graz, Austria surveyed adult participants and found no difference in the number of male and female cyclists (Titzea et al., 2008).

Age

Studies investigate impact of age on cycling showed that elderly people tended to cycle less compared to younger age groups (Vandenbulcke et al., 2009; Moritz, 1997). A case study in Belgium indicated that as the percentage of people less than 25 years of age increased in the urban population, bicycle usage increased. Also, the study found that the distance traveled decreased as the age of a cyclist increased for those over 50 years of age (Vandenbulcke et al., 2009). In the same line, according to Moritz (1997), who surveyed respondents from the United States and Canada, the majority of the surveyed cyclists were between 26 and 45 years old.

Researchers from the United States surveyed general cyclists in the United States, and the results showed about 90% of the surveyed cyclists were younger than 45 years old. The study also showed that people who were more than 45 tended to cycle less compared to other age groups. Moreover, cyclists who were older than 45 had a lower mean annual riding time compared to other age groups (Rodgers, 1995).

Some studies investigated the effect of age on accidents and safety. Researchers from the United States used data from a national survey of cyclists, and they showed that the risk of being in an accident had a negative correlation with the cyclists' age. Individuals who were younger than 24 years old faced a greater risk of being in an accident compared to other age groups. The risk of having an accident decreased as cyclists' age increased (Rodgers, 1997). The higher risk for younger cyclists could be associated with lack of experience. In contrast, another study that used bicycle fatality data from Europe mentioned that as the age of cyclists increased the number of fatalities increased (Wegman et al., 2012). These studies showed that although younger generations faced a higher risk of accident, older generations had a higher fatality rate.

Bicycle ownership

Studies found that owning a bicycle can positively impact cycling. A survey-based study from the UK showed that 100% of regular cyclists own a bicycle, while only 43% of people who never cycle own a bicycle (Gatersleban and Appleton, 2007). Similarly, another study showed that as the rate of bicycle ownership increased, the cycling rate increased (Buehler, 2012). In contrast, as car ownership increased, the cycling rate decreased (Buehler, 2012; Plaut, 2005). Consequently, owning a bicycle eases access to a bicycle for people, which leads to an increase in cycling.

Travel distance and duration

Studies that investigate the effect of travel distance and duration show that travel distance and duration have an inverse relationship to cycling frequency. As travel distance increases cycling frequency decreases (Vandenbulcke et al., 2009; Larsen and El-Geneidy, 2011; Hunt and Abraham 2007). In addition, studies showed that cyclists preferred shorter trips compared to longer trips. According to Dill (2009) the most important factor for cyclists in choosing a route was to minimize their travel distance.

Using commuting data from the United States, another study mentioned that commuter cyclists had a lower mean travel distance to work compared to car users. Also, they had a higher mean travel distance to work compared to people who walk to work. Among the respondents, most people who lived close to their workplace (a mean distance of 0.2 miles) usually walked to their workplace. Respondents with a mean distance of 2.54 miles from their workplace used a bicycle to commute, and car commuters had a mean distance of 13.5 miles (Plaut, 2005). Similarly, another study that used an online transportation survey of students from Arizona State University who lived off-campus showed that most of the respondents who cycled regularly lived within 1 mile of their destination (the university campus). This distance was higher for those respondents who commuted by car. The respondents who cycled lived a mean distance of 0.4 miles from their destination, while this number for motorists was 2 miles (Sisson and Tudor-Locke, 2008).

In Sweden, a survey showed that for travel distances of less than 3 km, more than 60% of these trips were made by bicycle from April to October and more than 40% were made by bicycle from November to March (Bergstrom and Magnusson, 2003). Similarly, Vandenbulckea (2009) conducted a case study in Belgium and found that a travel distance of about 2 kilometers had the highest number of commuter cyclists. Lower travel distances encouraged people to walk to their destinations, and for higher travel distances, individuals were inclined to use cars and public transportation.

Another study conducted a survey in Dublin and asked cyclists about their preferences in terms of trip duration. Cyclists indicated that they preferred shorter trips compared to longer trips (Caulfielda et al., 2012). Moreover Garrard et al. (2008) mentioned that most cyclists, especially women cyclists, were inclined to cycle for destinations with shorter travel times.

2.1.2 Attitudes toward Cycling

Attitude toward cycling played an important role in an individuals' decision to cycle. Studies showed that people with a positive attitude about cycling tended to cycle more (Gatersleben and Appleton, 2007; Heinen et al., 2011; Bergstrom and Magnusson, 2003; and Daley and Rissel, 2011). Regular cyclists believed cycling was an efficient, cheap, and environmentally friendly mode of transportation. In contrast, one of the main reasons that prevented non-cyclists, especially women, from cycling was an individual's concern about the safety aspects of this type of transportation according to Daley and Rissel (2011).

Another study found that non-cyclists believed that cycling was not a comfortable type of transportation, while cyclists believed that cycling was both mentally and physically relaxing and a cheap type of transportation (Heinen et al., 2011). According to Gatersleben and Appleton (2007), 100% of regular cyclists liked cycling and believed that cycling was good for one's health and the environment. However, among the people who did not cycle, only 34% liked cycling. According to Bergstrom and Magnusson (2003), travel time was mentioned by people who never used a bicycle as the most important factor for decision making. Further, the most important factor for winter cyclists was exercise.

Moreover, Heinen et al. (2011) found that the habit of cycling positively impacted the likelihood of using a bicycle to commute. In addition, they showed that the subjective norm (the social expectation to accept and follow a behavior) positively affected the likelihood of cycling for those who cycled for distances less than 5km. Similarly, research from Austria mentioned that social support had a positive impact on cycling rate. Their model showed that significant social support could lead to about a 60% increase in cycling frequency (Titzea et al., 2008).

2.1.3 Cycling Infrastructure

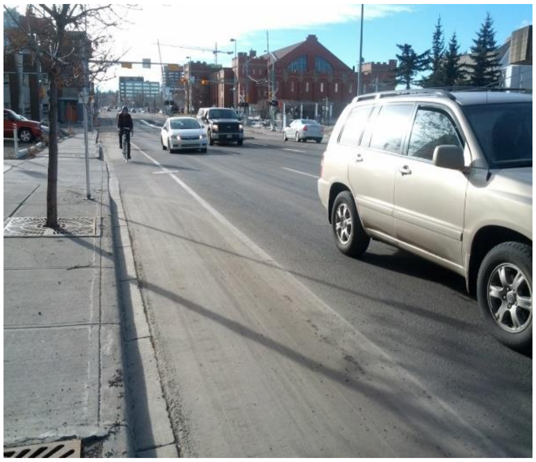
Studies found that infrastructure quality (i.e. type of cycling facility, road surface condition, geometric design of the infrastructure) was one of the principal factors affecting a cyclist's decision to choose a particular route.

Infrastructure type

Cycling infrastructure can be classified into four groups based on their cycling facilities. Figure 2.2 shows pictures of the four above mentioned infrastructure types.

1. Infrastructure without cycling facility: cyclists use the road with other users without any special facility.
2. Infrastructure with on-road bike lane: lane allocated to cyclists, separated from street with paint lane.
3. Infrastructure with physically separated bike lane: lane allocated to cyclists, separated from street with physical barrier.
4. Off-road pathways: A pathway dedicated only to cyclists.

a) Without cycling facility



b) On-road bike lane

c) Physically separated bike lane

d) Off-road pathways

Figure 2.2 Cycling infrastructure types

Studies found that cyclists preferred cycling pathways followed by on-street lanes (Dill, 2009; Caulfield et al., 2012; Twaddle et al., 2010). A case study from the Oregon metropolitan area in Portland showed that although roads with cycling facilities account for only 8% of the total road network, they account for 49% of total cycling usage (Dill, 2009). Another study discussed the effect of infrastructure type on usage. Cyclists preferred off-road paths compared to other types of facilities followed by green lanes and on-road lanes. Streets without any facilities for cycling were the least favored route for cyclists (Caulfield et al., 2012). Similarly, Bernhoft and Carstensen

(2008) conducted a survey in Denmark that showed that cyclists mentioned the presence of pathways for cycling as the most important factor for their comfort. Moreover, Twaddle et al. (2010) showed that cyclists mentioned that providing more off-street paths and cycling lanes were the two most effective factors for bicycle network improvement. According to Li et al. (2012), using data on cyclists from China, off-road paths were more comfortable for cyclists when cycling traffic was light. However, on-street cycle lanes were more comfortable for cyclists in heavy traffic.

Road surface conditions

Some studies discussed the effect of infrastructure condition on cyclists' choice. Slippery road surfaces and roads that had not been cleared of snow were the two most important road conditions that affected cyclists' mode choice (Bergstrom and Magnusson., 2003). In the same line, Amiri and Sadeghpour (2014) conducted a survey in Calgary in March 2012 with 103 respondents. Respondents indicated that their first priority for cycling facility improvement was snow and gravel removal.

Infrastructure width and slope

Li et al. (2012) showed that infrastructure slope and width of cycling infrastructure were factors that impacted cyclists' comfort. While Li et al. (2012) mentioned that cycling lane width had a positive correlation with cyclists' comfort and wider lanes provided more space for cyclists, one study showed that insufficient cycle lane width could have a reverse effect on crossing distance between vehicles and cyclists. In other words, in streets with insufficient cycle lane width, vehicles passed cyclists with a lower distance separation compared to streets without facilities. This reverse effect could result in a higher risk of accident for cyclists (Parkin and Meyers, 2010).

2.1.4 Built Environment and Network Connectivity

Research on the effect of built environment and network connectivity on cycling showed that cyclists preferred connected networks with a low number of intersections. Besides, cyclists preferred recreational neighborhoods compared to residential neighborhoods (Titze et al., 2008; Caulfield et al., 2012; Li et al., 2012).

Network connectivity

Network connectivity relates to how easily users can transport between different sections of the network. A well-connected network has links between each pair of sections while a low connected network does not. In the well-connected network, users have direct access to other parts of the network through the links. Studies indicated that cyclists preferred connected networks of cycling infrastructure, and network connectivity was positively associated with cycling frequency (Osama et al., 2017; Titze et al., 2008). A study from Graz, Austria found that highly connected networks had nearly a double odd ratio (100% increase in usage) compared to low connected networks in term of cyclists' usage (Titze et al., 2008).

Number of intersections

Cyclists greatly preferred roads with fewer intersections according to Caulfield et al. (2012). Also, destination facilities, such as showers, lockers, and bicycle parking, were mentioned as the most important improvements for improving destinations and encouraging cycling as a mode of transportation. (Twaddle et al., 2010). Also, according to Li et al. (2012) the presence of bus stops was negatively associated with cyclists' comfort for cyclists who used cycling pathways and positively associated with cyclists' comfort for cyclists who used on-street lanes. Consequently, bus stops provided more space for cyclists who cycle in on-road lanes.

Neighborhood

Li et al. (2012) showed that residential and commercial neighborhoods had a negative impact on cyclists' comfort for cyclists who used off-street pathways. However, for cyclists who cycled in on-road lanes, the type of neighborhood did not have a significant impact on their comfort. Moreover, according to Osama et al. (2017), infrastructure located in recreational neighborhoods attracted more cyclists compared to infrastructure located in residential neighborhoods.

2.1.5 Weather Conditions

A considerable amount of literature has been focused on the impact of weather conditions on cycling. These studies indicated that weather variables significantly impacted cycling frequency. Harsh weather conditions, such as days with low temperatures and high precipitation, reduced the number of cyclists (Bergström and Magnusson, 2003; Buehler, 2012; Flynn et al., 2012; Saneinejad et al., 2012). Table 2.2 shows the literature on impact of weather variables on cycling and the variables examined in each study.

Table 2.2 Literature of impact of weather variables on cycling

Article	Temperature	Precipitation	Snow on ground	Wind speed	Sunlight hours
Zhao et al. 2018	✓	✓		✓	
Meng et al. 2016	✓	✓			
Amiri and Sadeghpour 2015	✓		✓		
Amiri and Sadeghpour 2014	✓		✓		
Spencer et al. 2013		✓	✓	✓	✓
Buehler 2012	✓				
Saneinejad et al. 2012	✓	✓		✓	
Flynn et al. 2012	✓	✓			
Parkin et al. 2008	✓	✓			
Brandenburg et al. 2007	✓				
Brandenburg et al. 2004	✓	✓			
Bergstrom and Magnusson 2003	✓	✓	✓		
Nankervis 1999		✓			
Osberg et al. 1998					✓
Rodgers 1995					✓

Temperature

Temperature was positively correlated with cycling frequency according to relevant studies. Studies conducted in cities with cold climate showed that days with higher temperatures caused more people to cycle. A Swedish study that surveyed employees of four companies in two Swedish cities showed a difference in mode choice between seasons. People cycled more during summer compared to winter. Instead of cycling, the majority of people used their cars in winter, which indicated that the main alternative mode of transportation in cold days was a car. Also, distance became a more important factor when choosing the mode of transportation during winter. For long distances, the number of cyclists remarkably dropped as distance increased in winter. The study showed that people almost never took bicycle trips of more than 10 km in winter (Bergström and Magnusson, 2003). Similarly, Buehler (2012) showed that the likelihood of commuting to work by bicycle in summer was 73% more than the likelihood of using a bicycle to commute in winter.

Another study from Vermont showed that bicycle usage had a positive correlation with temperature. People tended to cycle more on warmer days. The model presented in the study showed that an increase in one degree of Fahrenheit resulted in a 3% increase in the likelihood of cycling (Flynn et al., 2012). Another study using UK census data, showed that higher temperature encourages more people to cycle to work (Parkin et al., 2008). Similarly, another study from Toronto showed that cold temperatures negatively affected people's tendencies to cycle. The study also showed younger people were more affected by cold temperatures than older age groups. The researchers found that using a bicycle was only sensitive to temperatures below 15° C. They also investigated the impact of weather conditions on both male and female cyclists. The results showed that the impact of temperature on the two genders was not the same; women were more affected by temperature compared to men (Saneinejad et al., 2012). However, Zhao et al. (2018) investigating the impact of weather on the usage of two cycling infrastructure sections (i.e. cycling pathway, physically separated bike lane) located in Seattle showed that temperature is positively correlated with usage and the usage is sensitive to temperature below 20° C. Brandenburg et al. (2007) investigated the impact of weather conditions on commuter and recreational cyclists. They found, using data from a suburban recreation area in Vienna, that recreational cyclists were more affected by weather conditions compared to commuter cyclists. Consequently, it is possible that recreational cyclists checked the weather conditions beforehand, and as a result, they avoided cycling in harsh weather.

Sunlight hours

Studies showed that individuals tended to cycle during daylight hours rather than in the dark (Osberg et al., 1998; Rodgers, 1995). A study that observed cyclists in Paris and Boston showed that the majority of cyclists, cycle during daylight hours. Sixty-five and 76% of cyclists were observed during daylight hours in Paris and Boston, respectively (Osberg et al., 1998). Another study using a bicycle exposure survey showed that 54% of cyclists avoided cycling in the dark. Also, the risk of death for cyclists who cycled in the dark was about 3.8 times higher than for cyclists who cycled during daylight hours (Rodgers, 1995). Respondents in another study mentioned that during some parts of the year, they could not cycle safely due to insufficient daylight hours. Also, some respondents indicated that the effect of weather variables, such as rain, could compound the negative effects of cycling in the dark, which further prevented them from cycling (Spencer et al., 2013).

Wind

Another weather variable discussed in the literature is wind. Wind seems to have a relatively minor negative impact on a cyclist's decision to cycle. Researchers believe that the main reason for the negative impact is that wind makes the temperature feel colder. Further, a study from Toronto found that wind speed negatively affects cyclists. The magnitude of wind was twice that for cyclists compared to pedestrians (Saneinejad et al., 2012). However, another study surveyed 24 adult bicycle commuters, and the results showed that wind could also have a positive impact on cycling when it was at a cyclist's back (Spencer et al., 2013).

Precipitation

According to relevant studies, precipitation is one of the preventing factors for cyclists; the number of cyclists drops on days with precipitation. A case study in Vermont found that people cycle less during cold weather and on rainy days. Among the weather factors, rain was found to have the greatest negative effect on cycling. Days without rain had an odds ratio equal to 1.91 compared to rainy days. Also, each inch of snow resulted in a 10% decrease in the likelihood of bicycle commuting (Flynn et al., 2012). Another study from Melbourne showed that rain was the most important factor as it caused more than half of the people surveyed (50.7%) to choose not to ride a bicycle on rainy days; the second most important factor was cold weather, as it caused about one-fifth of the people surveyed (20.9%) to choose not to ride a bicycle (Nankervis, 1999). In the same line Zhao et al. (2018) investigating impact of weather on the usage of two cycling infrastructure sections (i.e. cycling pathway, physically separated bike lane) showed that as amount of precipitation increases the usage of infrastructure decreases. Moreover, people who never cycled mentioned precipitation as the third most important factor for their mode choice (Bergstrom and Magnusson, 2003). Another study using observations in two recreation areas in Austria showed that percentage share of bicyclists drops from 60% (days without precipitation) to 40% in days with precipitation (Brandenburg et al. 2004). Similarly, according to Saneinejad et al. (2012), precipitation in the form of rain and showers negatively impacted cyclists; showers showed a larger impact on cyclists followed by rain. In the same line Meng et al. (2016) using a survey conducted on 553 cyclists in Singapore showed that rain makes cyclists feel that the weather is not good for cycling. Their results showed almost 3 out of 10 cyclists check the weather before their trip and 2 out of 3 would change their plan if it is going to rain. However, rain was not always considered a negative factor for cyclists. Some respondents described rain as refreshing and helpful (Spencer et al., 2013).

Snow on the ground

Snow on the ground was also considered to be one of the factors that prevented cyclists from cycling. Amiri and Sadeghpour (2015) conducted a survey with cyclists during winter showed that the most important preventing factor for winter cyclists was ice or snow on the ground. According Bergstrom and Magnusson (2003), road surfaces that were not cleared of snow was the most important road condition for mode choice according to the opinion of the surveyed cyclists.

Also, Spencer et al. (2013) showed that using salt to melt ice on cycling infrastructures had a negative impact on cyclists. Commuters described salt on roads as hazardous for cycling. In the same line, Amiri and Sadeghpour (2014) conducted a survey in Calgary in March 2012 with 103 respondents. Respondents indicated that their first priority for cycling facility improvement was snow and gravel removal. The details of the variables that affect cycling, including their groupings, is shown in Table 2.3.

Table 2.3 An overview of the literature on factors affecting cycling frequency

Characteristics	Attitude toward Cycling	Infrastructure	Weather Conditions	Built Environment
Age	Belief in benefits	Infrastructure type	Temperature	Neighborhood
Gender	*Environmental benefit*	*Off-road pathways*	Sunlight hours	*Residential*
Education	*Fitness and exercise*	*On-road Sep. physically*	Wind	*Commercial*
Annual income	*Enjoyment*	*On-road Sep. by lane*	Precipitation	*Industrial*
Body mass index	Being outside	Without cycling facilities	*Rain*	Parks
Own a car	Flexibility	Cycling road width	*Hail*	Network connectivity
Own a bicycle	Cost saving	Road conditions	*Snow*	Number of intersections
Having children	Habit	Surface snow clearance	Snow on ground	Bus stop, parking
Trip distance	Subjective norm	Slippery surface		Destination facilities
Trip duration	Perceived behavior	Occurrence of grit/debris		*Bike parking*
Trip purpose	Culture	Surface cracks		*Showers*
	Safety concerns	Road Slope		*Lockers*
				Free car parking

2.2 CYCLING INFRASTRUCTURE USAGE

Cycling usage trends shows the usage increase and decrease depending on the time of year. In recent years, studies investigated cycling infrastructure usage trends and developed estimation models to estimate infrastructure usage. Investigating cycling infrastructure usage trends helps us to better understand usage variation and the possible causes of the variation. A better understanding of bicycle usage trends enables us to develop more accurate estimation models.

2.2.1 Usage Trends Over Time

Some recent studies investigated cycling infrastructure usage trends; one study in particular used data from Minneapolis and found that traffic volumes for cyclists and pedestrians (non-motorized usage) varied in different locations (Lindsey et al., 2013). Also, the usage varied depending on the month and day of the week. However, usage trends were consistent for different locations with different magnitudes. Also, the researchers found that cycling trends had a higher variation within a year than pedestrian trends. As it is assumed that the cause of this variation is effect of weather conditions, this finding agreed with the results of the study conducted by Saneinejad et al. (2012) that found weather variables such as wind affected cyclists more than pedestrians. In addition, Lindsey et al. (2013) found that cycling usage increased more on the weekend compared to weekdays than did pedestrian usage.

Miranda-Moreno et al. (2013) studied cycling patterns in five North-American cities, Montreal, Ottawa, Portland, San Francisco, and Vancouver, and classified the existing infrastructures based on the type of user. The study showed that bicycle patterns could be categorized as commuter, mixed commuter, recreational, and mixed recreational. The commuter pattern usually indicated morning and afternoon peak usage similar to motorized usage patterns, while recreational patterns indicated only one peak usage at midday both on weekends and weekdays. They found that within their classification of cycling infrastructure, hourly and daily usage were fairly consistent. However, they indicated that monthly usage appeared to be impacted by weather conditions as monthly usage in cities with cold climates decreased more compared to usage in warmer cities. Also, they mentioned that the impact of weather on recreational and commuter cyclists was not the same; infrastructure usage by recreational cyclists decreased more compared to the usage of infrastructure by commuter cyclists.

2.2.2 Usage Estimation Models

In recent years, there has been an increasing amount of literature on estimating cycling infrastructure usage. To estimate cycling infrastructure usage, Osama et al. (2017) developed a model using network indicators, land use variables, and road facilities. They used data from 134 traffic analysis zones in the city of Vancouver. Their estimation model showed a good fit with R-Squared equal to 0.62. Among the variables that affected infrastructure usage, their analysis showed that recreational neighborhoods attracted more cyclists compared to residential neighborhoods; this finding agreed with the findings of Caulfield et al. (2012). Also, the results from the Osama et al. study showed that with an increase in length, coverage, and continuity of bike network, cycling infrastructure usage increased in agreement with Titze et al. (2008). Osama et al. (2017) model showed that slope had a negative impact on cyclists. This finding was consistent with the findings of Li et al. (2012). Further, their model showed that cycling pathways encouraged more people to cycle.

Other studies estimated cycling infrastructure usage by developing models using partial data (Nordback et al., 2015; El-Esawey, 2014; El-Esawey et al., 2013; El-Esawey et al., 2015; El-Esawey and Mosa, 2017). These studies estimated bicycle traffic by creating daily, monthly, and seasonal factors. Bicycle volume was estimated by multiplying the known short-term count by those factors. To estimate cycling infrastructure usage, El-Esawey et al. (2013) developed daily factors, which could be used to estimate monthly usage when daily usage was available. They developed factors for each day of the week, each month, and two weather conditions, which resulted in 168 factors. They used cycling data from 74 road sections in Vancouver in 2010 and 2011. Since data for the study were collected in Vancouver, which had a moderate temperature during year, they only considered the impact of precipitation. They categorized days into wet and dry days; wet days had more than 5 mm of precipitation, while dry days had less than 5 mm. El-Esawey et al. (2013) focused on estimating monthly usage using daily factors. Their model estimated monthly cycling infrastructure usage with an average error of 18%. Also, they found that developing different factors for different types of infrastructure did not improve the accuracy of estimation.

In another study, El-Esawey (2014) estimated the annual cycling infrastructure usage using daily, monthly, and seasonal factors. The study used daily bicycle volume data for a full year in Vancouver. Similar to his previous work, El-Esawey defined two weather conditions, wet and dry. His analysis showed that using monthly factors provided better results to estimate annual usage compared to using seasonal factors. Estimating annual usage with monthly factors had an average error of 11%, while the error increased to 17% when using seasonal factors. Moreover, their analysis showed that 15% of the annual usage estimation error was due to using daily factors, while 11% of the error was due to using monthly factors. The overall error for using both daily and monthly factors was about 23%. In addition, the study showed that for estimating annual usage of a specific year, using factors that were calculated from other years led to larger errors. Hence, to estimate cycling infrastructure usage for a year, researchers should use that specific year's factors.

El-Esawey et al. (2015) provided a framework to estimate annual cycling infrastructure usage using bicycle volume data. They defined two weather scenarios and described factors for each weather scenario. Wet weather had more than 5 mm and dry weather had less than 5 mm of precipitation. They validated their factors for estimating daily usage using two consecutive hours between 6 am and 9 pm. Their error analysis showed that they had about 40% error for estimating daily usage using data from a two-hour window.

In another study, El-Esawey and Mosa (2017) developed two factors to estimate cycling infrastructure usage. The first factor (Kp/d) was developed using data of peak hour usage and total daily usage. The second factor (Kp/AADB) was developed using data of peak hour usage and annual average usage. These two factors were developed to estimate daily and annual usage, respectively, using peak hour usage. Different Kp/d factors were developed for different links, for weekdays and weekends, and for each year. Different Kp/AADB factors were developed for weekdays and weekends, for each month of the year, and for each year. El-Esawey and Mosa (2017) validation showed that the average error when using Kp/d to estimate daily usage was 16.6%. The estimated daily usage could be used with daily and monthly factors to estimate annual usage. Using Kp/AADB to estimate annual cycling infrastructure usage with peak hour usage led to an error of about 28.3%.

Nordback et al. (2013) developed daily and monthly factors to estimate annual cycling infrastructure usage. They used cycling data from Boulder, Colorado. Data were collected from 26 stations from 1999 to 2012. The analysis of the data showed that the estimation error is lowest when the annual estimation was based on one or more weeks of data: their average model estimation error was less than 30% when they used one to three weeks of data, and when four weeks of data was used to estimate annual usage, the lowest average estimation error (15%) was obtained. Another study developed day-of-the-year factors, which resulted in 365 factors for each day of the year. The study showed that using day-of-the-year factors to estimate annual usage provided greater accuracy than using factors for days of the week or for months of the year (Hankey et al., 2014). Table 2.4 provides an overview of usage estimation models and their average estimation errors.

Table 2.4 Overview of the usage estimation models and their estimation errors

	Usage Estimation		Error (%)
	Observed usage	Estimated usage	
El-Esawey et al. 2013	Daily	Monthly	18
El-Esawey 2014	Daily	Annual	15
	Monthly		11
	Seasonal		17
Hankey et al. 2014	Daily	Annual	20
El-Esawey et al. 2015	Two-hour	Daily	40
Nordback et al. 2015	Hourly (1h, 2h, 3h)	Annual	54,46,40 respectively
	Daily		38
	Weekly (1week, 2week)		22,19 respectively
	Monthly		15
El-Esawey and Mosa 2017	Peak hour	Annual	28.3
	Daily		16.6

2.3 SUMMARY

Studies have shown that several variables affect cycling. In this chapter, these variables are categorized into five groups: cyclist characteristics, attitude toward cycling, cycling infrastructure, built environment and network connectivity and weather conditions.

Research on infrastructure usage trends showed that cycling infrastructure usage during a year was affected by weather conditions (Miranda-Moreno et al., 2013). In other words, studies showed that usage variation was related to weather variation. The trend was similar for different locations with different magnitudes of usage. While the usage variation is due to impact of weather variables, recent studies estimated annual cycling infrastructure usage including the effect of only one weather variable (precipitation). Those studies estimated cycling infrastructure usage using short count duration factors. The developed factors could only estimate annual usage and were not able to estimate usage variation during a year. Also, when estimating the usage for years other than the year factors were developed on, their estimation accuracy drops significantly since the variation in weather conditions is different in each year. Hence using short count factors is limited for only one year.

Consequently, in this thesis a model is developed to estimate annual cycling infrastructure usage and usage variation over a year. The estimation model estimates cycling infrastructure usage while considering the effect of five weather variables. The proposed model estimates annual usage and the variation in usage during a year and can be used for any year with known weather data.

CHAPTER 3: THE ESTIMATION MODEL

The objective of this study is to presents a generic framework for developing a model capable of estimating cycling infrastructure usage variation using short count data. To develop the model actual data was required. Hence, cycling infrastructure usage data from the city of Calgary was used. The applicability of the model to other locations will be discussed in chapter 5.

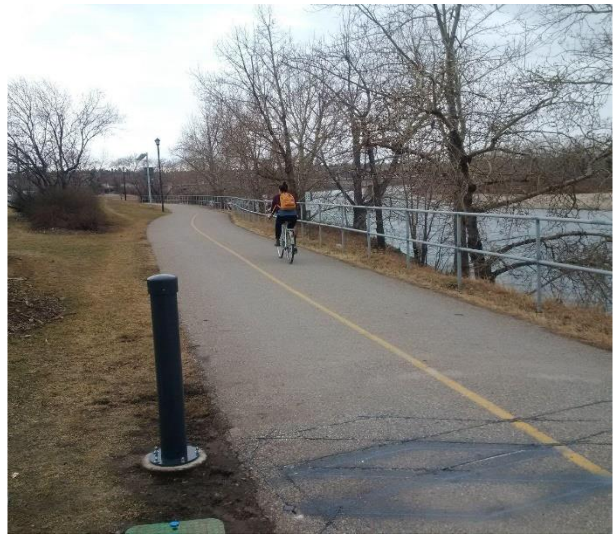
First, potential variables that impact usage over a year were identified to develop the model. Previous studies suggested that weather variables had a significant impact on cycling infrastructure usage variation. Investigating the association between two weather variables namely temperature and precipitation with infrastructure usage, confirmed the speculation about the effect of weather conditions on infrastructure usage. Hence, an estimation model was developed based on weather variables that impact cycling. During the development of the model, the magnitude of the impact of those variables on cycling infrastructure usage was determined, and infrastructure usage was estimated based on those variables.

3.1 DATA SOURCES

Data used for this study was provided from three sources.

1: The Transportation Department at the City of Calgary counts daily infrastructure usage throughout the year. They provided daily bicycle counts obtained from several counters at different locations throughout Calgary. This source of data includes 15750 data points; each data point represents specific infrastructure section usage for a specific day of the year. Each infrastructure section is a part of the infrastructure with different characteristics from other parts of infrastructure which is separated from other sections by intersections. The counters contain a diamond loop installed under the road surface and a pole placed beside the road. Counters count each bicycle by analysing the electromagnetic signature of bicycle wheels. These counters are 97% accurate for detecting bicycles. Figure 3.1 shows some of the counters.

a) Lindsay Park counter



b) Bow River Pathway counter

c) Nose Creek at Bow River counter

d) Parkdale counter

Figure 3.1 Photos of automated bicycle counters

Bicycle counters are installed in different locations across the city of Calgary. The location of counters can be seen in Figure 3.2 As shown in the figure, the majority of counters are installed in the downtown area. A few counters are installed outside the downtown area mostly in recreational areas and parks. Table 3.1 shows the assigned number for each infrastructure section.

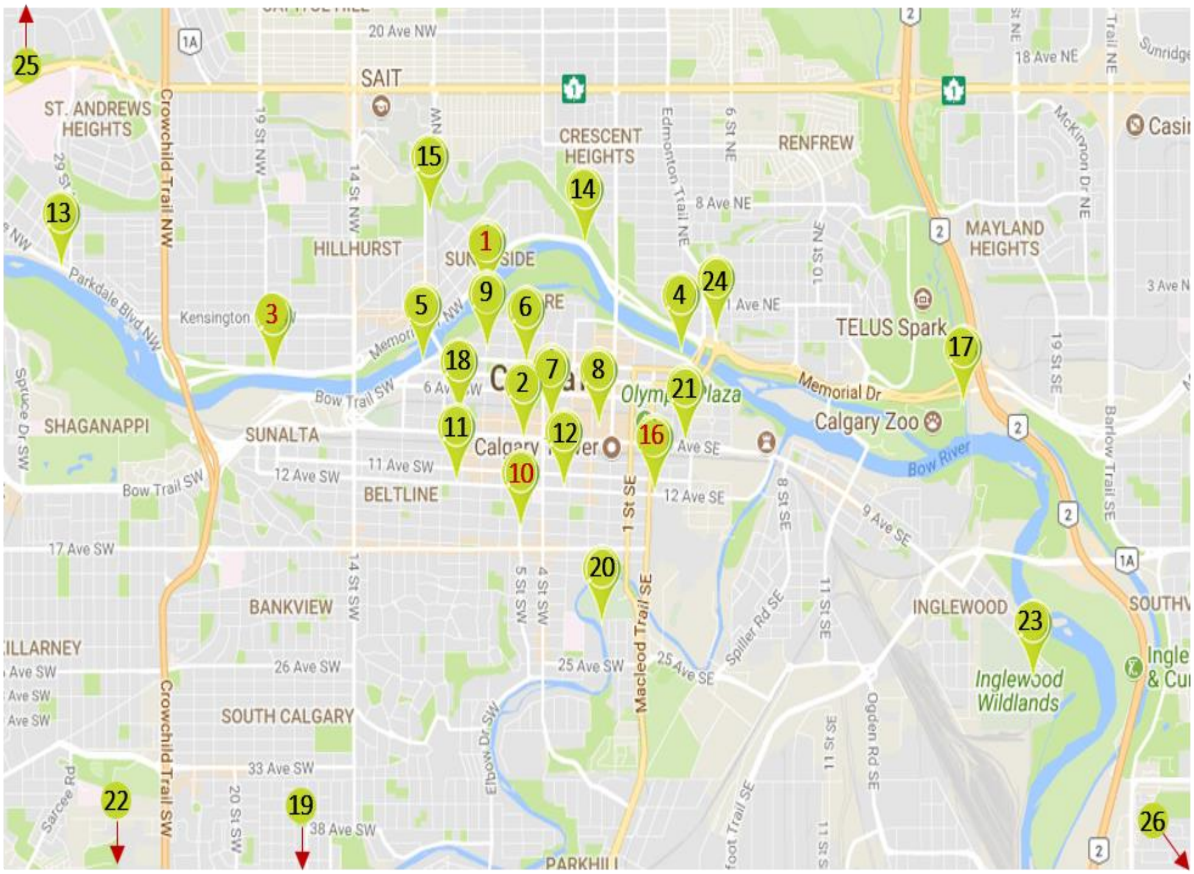


Figure 3.2 Locations of automated bicycle counters throughout the City of Calgary

Table 3.1 Infrastructure sections assigned number and location

Infrastructure Section	Infrastructure assigned number
Peace Bridge	1
5 St at 10 Ave SW	2
Memorial Drive and 19 St NW	3
River Walk	4
Bow River Pathway	5
5 St and 5 Ave SW	6
8 Ave and 3 St	7
Stephen Avenue Bicycle Counts	8
7 St and 3 Ave SW	9
5 St and 15 Ave SW	10
12 Ave and 8 St SW Solar Display	11
12 Ave and 2 St SW	12
Parkdale	13
Memorial Drive at Prince's Island Park	14
10 St and 5 Ave NW	15
12 Ave and 3 St SE	16
Nose Creek at Bow River	17
8 Ave and 8 St	18
South Glenmore	19
Lindsay Park	20
9 Ave and 4 St SE	21
North Glenmore	22
Inglewood Bird Sanctuary	23
Edmonton Trail	24
Nose hill Park	25
Wetland Trail	26

2: The Canadian Government provides historical weather data on its website. The data includes mean temperature, precipitation, accumulated snow on the ground, and wind speed, all of which are measured daily throughout the year at the Calgary International Airport station.

3: The public website (http://www.sunrise-and-sunset.com) provides data for the number of sunlight hours and times of the sunrise and sunset for every day of the year in Calgary.

3.2 ANALYSIS OF CYCLING USAGE VARIATION

The first step was to find the potential causes of cycling infrastructure usage variation. Therefore, cycling infrastructures usage trends were investigated. It was noted that usage trends for different infrastructure sections during a year had several similarities regardless of the difference in characteristics of infrastructure sections. Next, infrastructure usage trends were compared with usage trends of different weather variables. It was noted that cycling infrastructure usage was associated with weather variables.

3.1.2 Hourly Cycling Infrastructure Usage Trends

Hourly usage trends for cycling infrastructure sections were developed using counter data for 2016. For each section of infrastructure, the usage in 2016 was calculated for each hour on both weekdays and weekends. Then, the ratio of the usage for each hour to the total usage of infrastructure sections in 2016 was calculated. Figure 3.3 and Figure 3.4 show cycling infrastructure hourly usage trends for different sections of infrastructure during weekdays and weekends. The vertical axis represents the ratio of usage for each hour compared to the total usage, and the horizontal axis represents the hours of a day. As shown in Figure 3.3, the weekday usage pattern includes morning and afternoon peak periods that are similar to motorized peak periods. This pattern is approximately the same for all of the infrastructure sections. The morning peak roughly starts at 6:30 a.m. and lasts until 9:30 a.m., and the afternoon peak roughly starts at 3:30 p.m. and lasts until 6:30 p.m. The lowest usage of infrastructure sections is from 12:00 a.m. to 5:00 a.m.

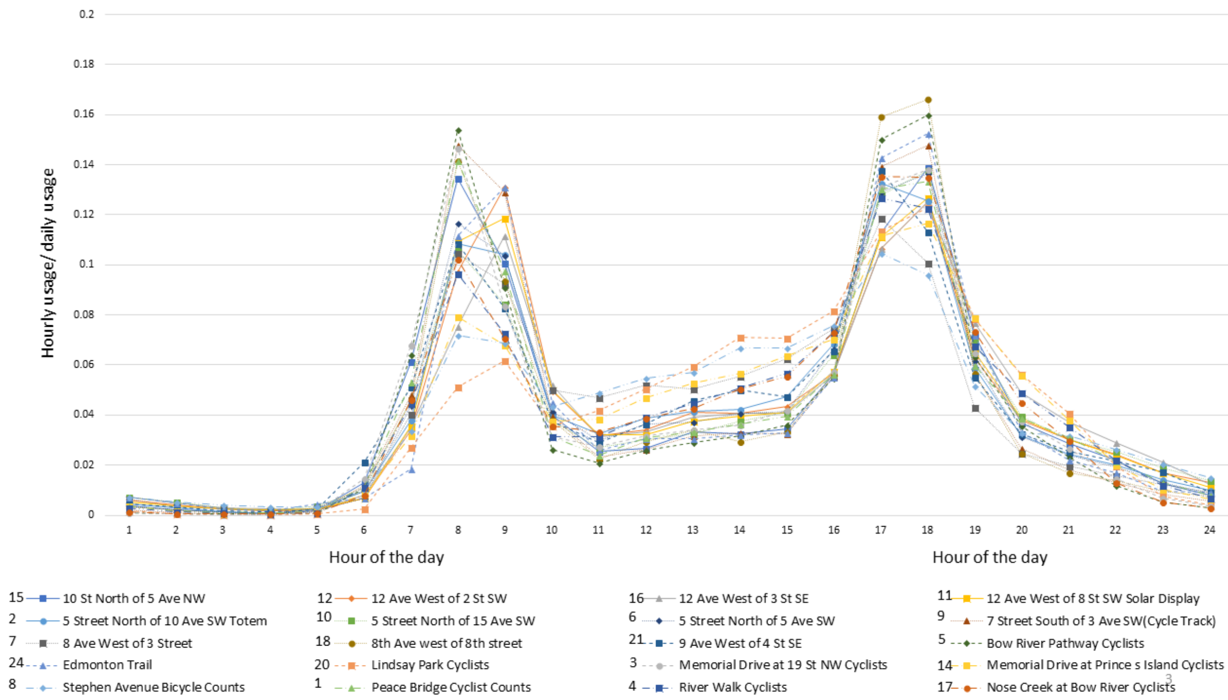


Figure 3.3 Infrastructure hourly usage trends during weekdays

Figure 3.4 shows the usage pattern during weekends. The weekend pattern includes only one peak at midday, which roughly starts at 1:30 p.m. and lasts until 5:30 p.m. The usage patterns for different sections of infrastructure are similar. However, the peak period at midday for some of the infrastructure sections are higher compared to others. These highly-used infrastructure sections are located in parks and recreational places and are mainly used for recreational purposes. The lowest usage of infrastructure sections is from 4:00 a.m. until 6:00 a.m.

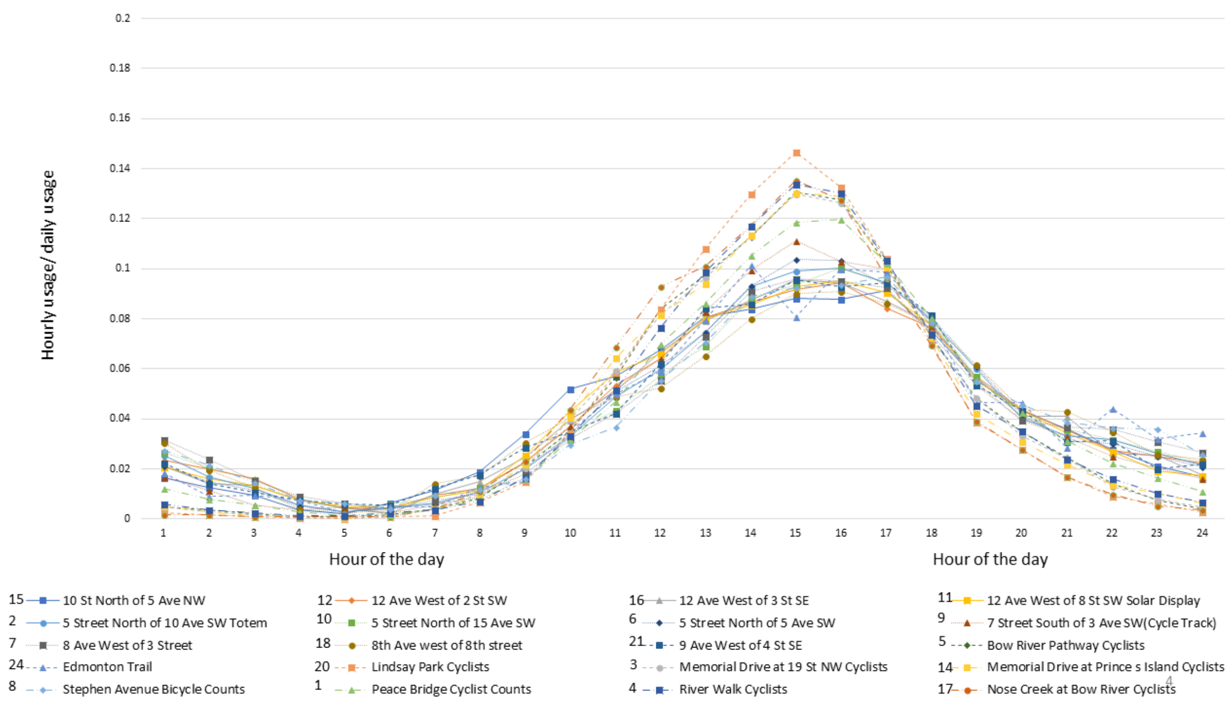
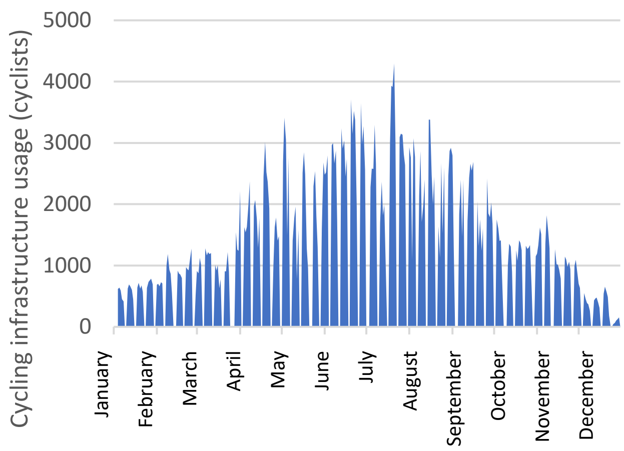


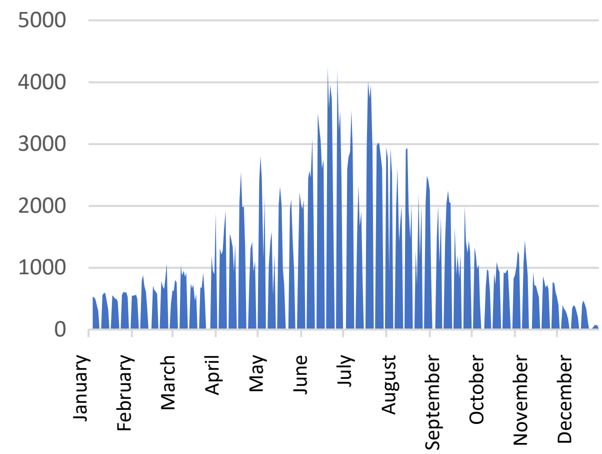
Figure 3.4 Infrastructure hourly usage trends during weekends

3.1.3 Daily Cycling Infrastructure Usage Trends

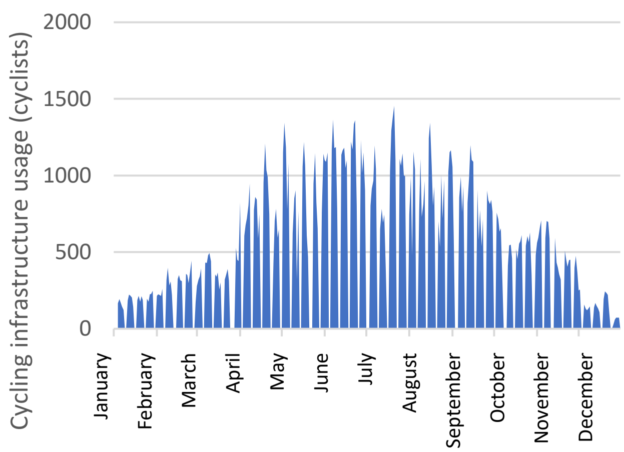
Daily infrastructure usage trends were developed from infrastructure usage data from 2016. Since weekend and weekday usages were different in values, they were plotted separately. Figure 3.5 shows the weekday usage in 2016 for four selected sections of infrastructure. Figure 3.6 shows the weekend usage in 2016 for the four selected sections of infrastructure.



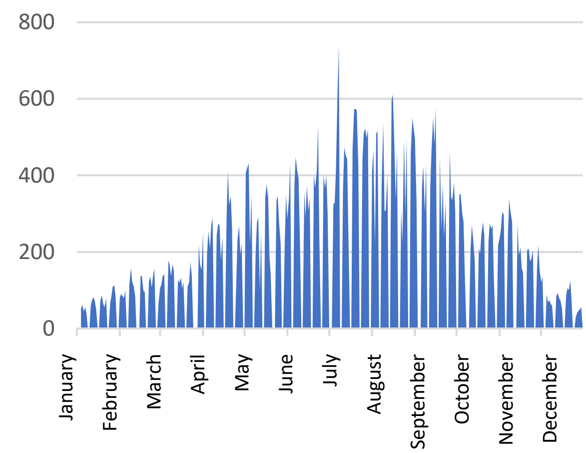
a) Peace Bridge (Counter 1)



b) Memorial Drive and 19 St NW (Counter 3)

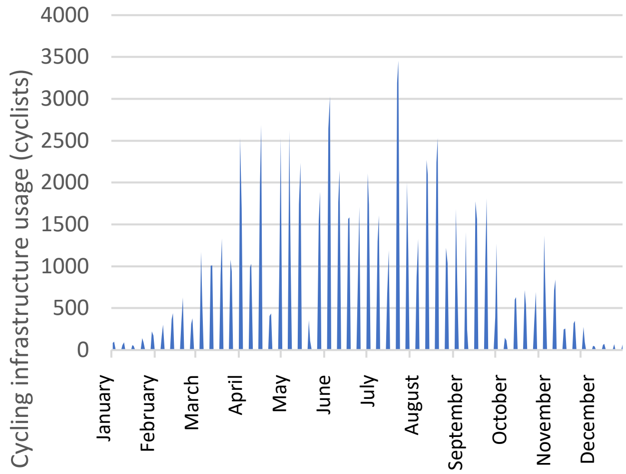


c) 5 St and 15 Ave SW (Counter 10)

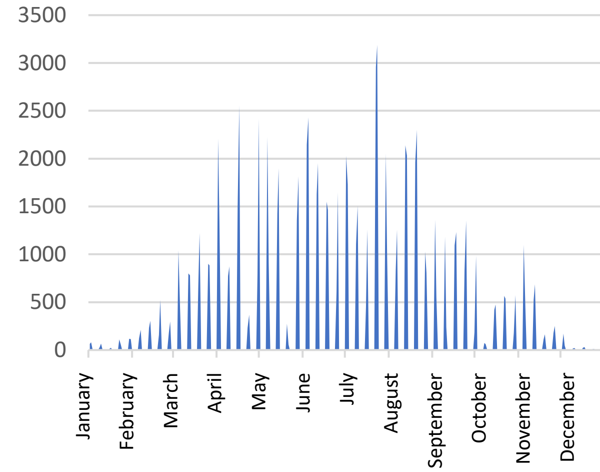


d) 12 Ave and 3 St SE (Counter 16)

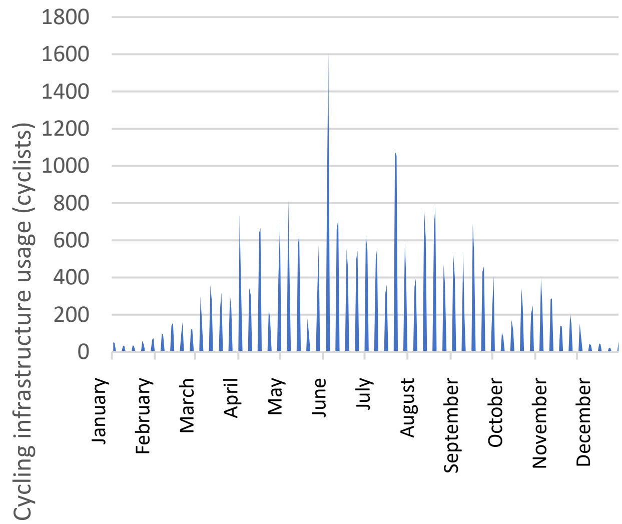
Figure 3.5 Daily infrastructure usage trends during weekdays over a year



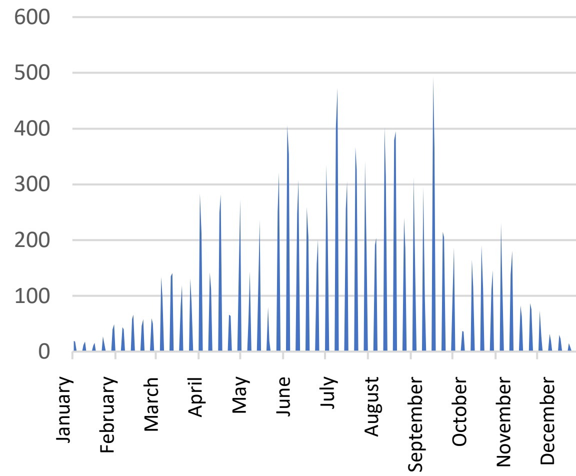
a) Peace Bridge (Counter 1)



b) Memorial Drive and 19 St NW (Counter 3)



c) 5 St and 15 Ave SW (Counter 10)



d) 12 Ave and 3 St SE (Counter 16)

Figure 3.6 Daily infrastructure usage trends during weekends over a year

Table 3.2 Infrastructure characteristics

Infrastructure sections	Type of infrastructure	Location	Infrastructure neighborhood
Peace Bridge	Off-road pathways	Downtown	Park
Memorial Drive and 19 St	Off-road pathways	Outside downtown	Residential
5 St at 15 Ave SW	On-road Sep. physically	Downtown	Commercial & residential
12 Ave and 3 St SE	On-road Sep. physically	Downtown	Commercial & residential

Table 3.2 shows the characteristics of the selected infrastructure sections. The selected infrastructure sections are shown in Figure 3.2 with different colors (1, 3, 10, 16). Since we wanted to cover a variety of infrastructure sections with different characteristics four infrastructure sections were selected with different locations and different characteristics, such as infrastructure type and neighborhood. Selected infrastructure sections showed similar trends in their usage over a year. These infrastructure sections show several peaks and drops in their usage at the same times of year. Figure 3.7 shows the usage trends during weekdays for the selected infrastructure sections. For example, the graph shows that there is a drop in usage in late October for all infrastructure sections. Also, there is a similar drop in mid-December in usage for all infrastructure sections. Figure 3.8 shows the weekend usage trends for the selected infrastructure sections. Again, similar behaviors can be seen in all infrastructure sections. For example, usage drops in October and rises in early November.

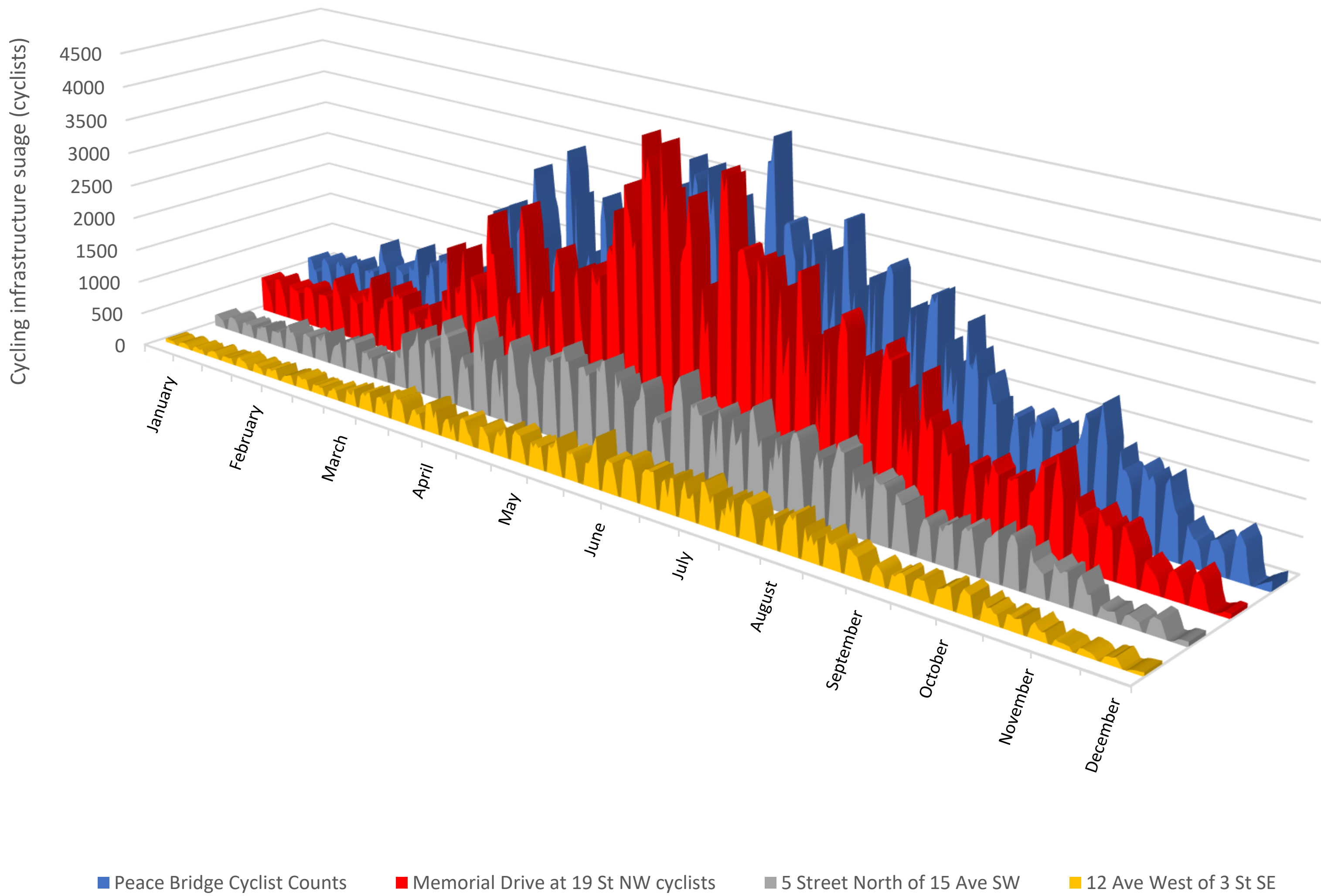


Figure 3.7 Weekday usage trends 2016

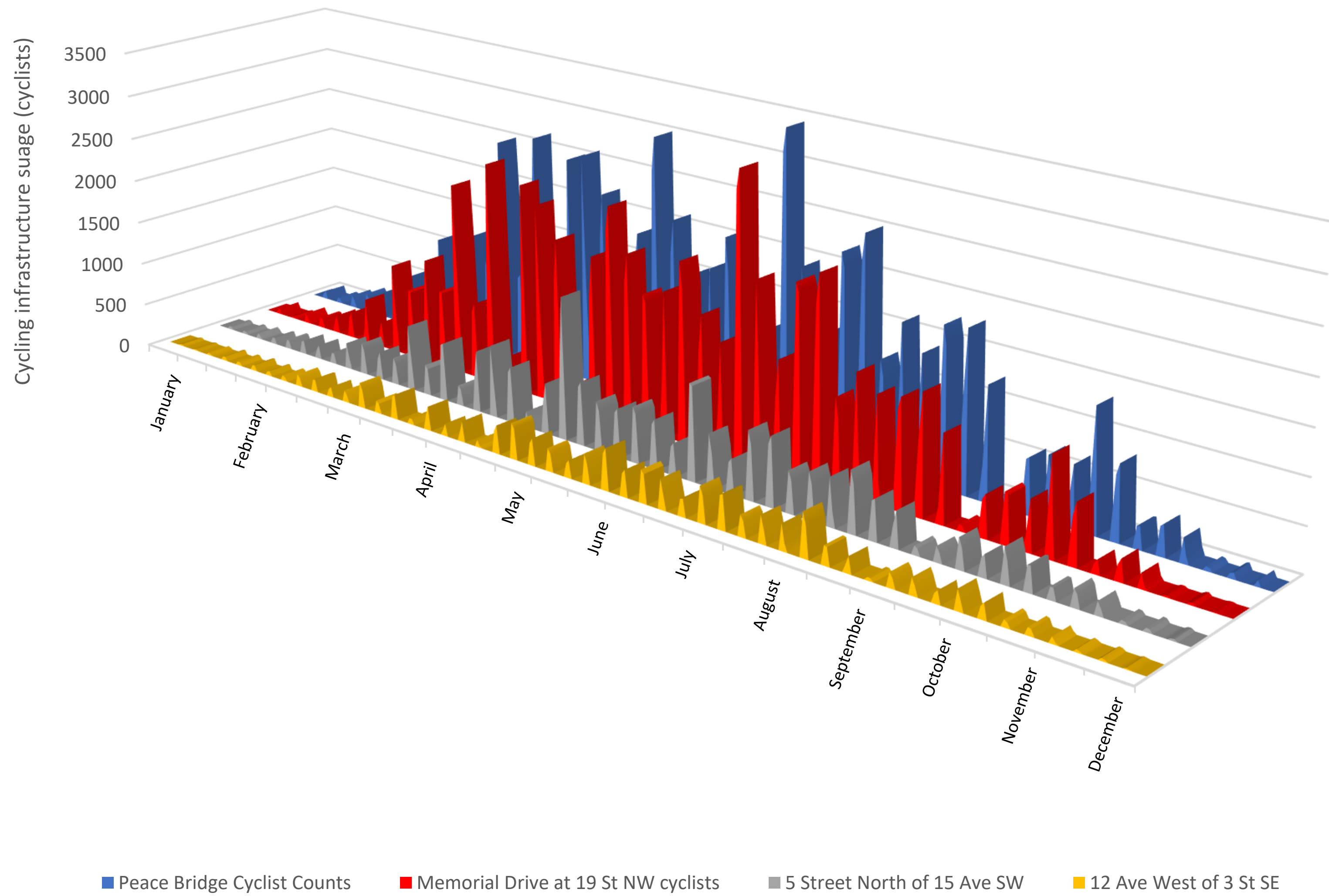


Figure 3.8 Weekend usage trends in 2016

3.1.4 Infrastructure Usage and Weather Conditions

One of the hypotheses of the causes of cycling infrastructure usage variation is the effect of weather conditions on cyclists. To discover more about the effect of weather on cyclists, we compared cycling infrastructure usage trends with the trends of the weather variables. Looking at usage trends of one randomly selected infrastructure section (12 Ave at 3 St SE) on weekdays and two weather variable trends, we noted that cycling infrastructure usage is affected by weather conditions. Figure 3.9 shows the selected infrastructure sections usage trend and temperature trend in 2016. As indicated in the figure with red boxes, a drop in temperature causes a drop in usage.

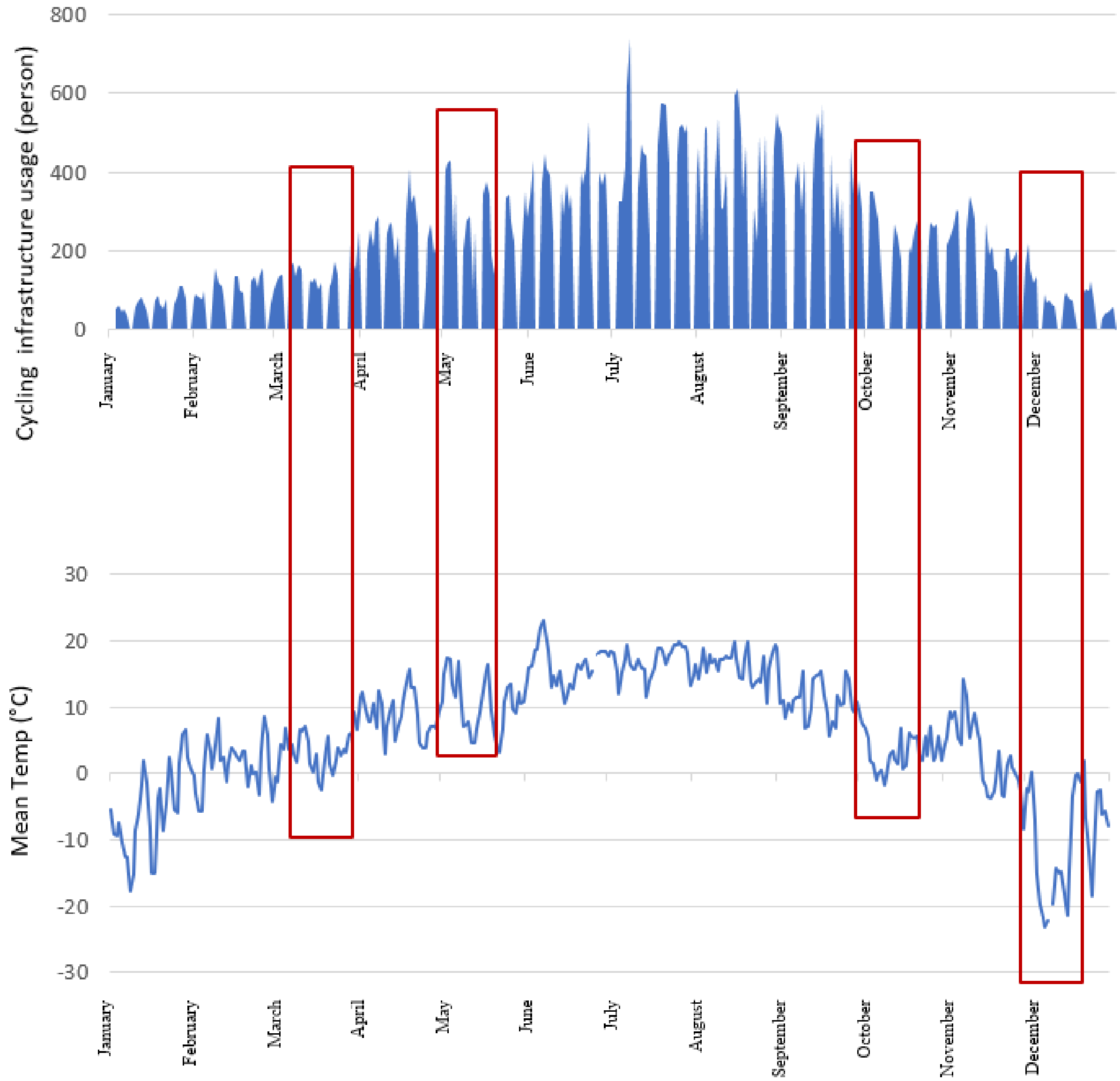


Figure 3.9 Cycling infrastructure usage and temperature

Figure 3.10 shows the selected infrastructure sections usage and precipitation. As shown in the figure, infrastructure usage drops on days with high precipitation.

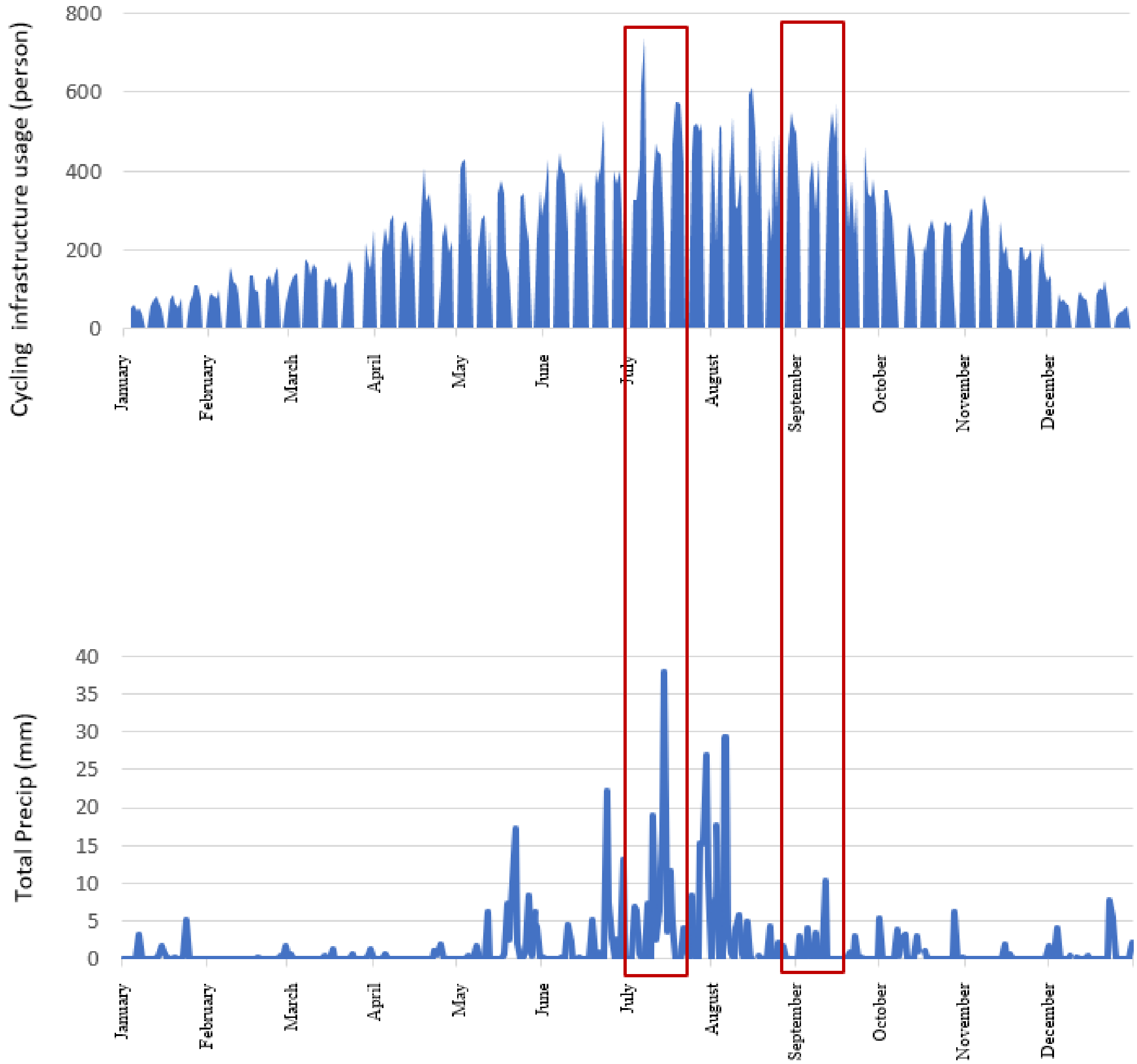


Figure 3.10 Cycling infrastructure usage and precipitation

3.1.5 Summary

Analysis demonstrated that infrastructure sections in different locations across the city with different characteristics, display similar trends in usage during a year. Also, comparing the usage trends with the weather variable trends, we noted that usage was correlated with the weather variables. Since other factors that affect infrastructure usage, such as built environment and infrastructure characteristics, are consistent for specific infrastructure sections during a year, the model was developed based on weather variables to estimate infrastructure usage variation over a year.

3.2 ESTIMATION APPROACH

In this part of the study, we evaluated the impact of weather variables on cycling infrastructure usage and estimated cycling infrastructure usage variation based on weather variables. To develop and test the model, we separate the data into two sections. The majority of the collected data (about 90% of data) was used to develop the model, and a small portion (about 10% of data) was used as test data to validate the model and to determine the model accuracy for estimating cycling infrastructure usage. Figure 3.11 shows the data preparation procedure.

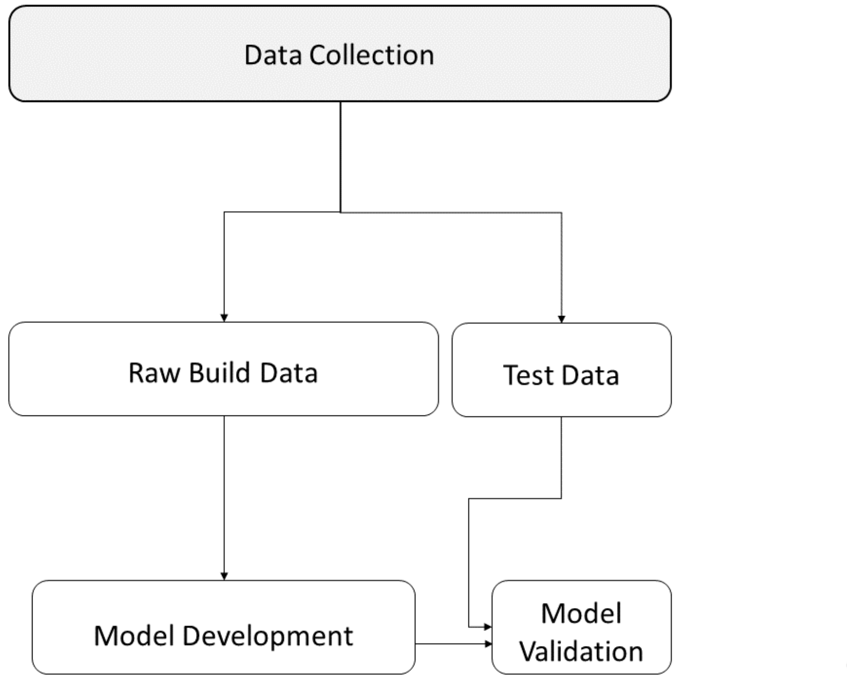


Figure 3.11 Data preparation flowchart

This study estimates the usage of cycling infrastructure compared to a reference month with known usage (the month with observation). The estimation is conducted by evaluating the change in usage as a result of a change in weather variables. The usage change contains five components. Each component represents the impact of each weather variable on usage. Total change in usage is computed by adding up the changes in usage resulting from the effect of each weather variable on usage. The equation can be written as follows where ΔF represents the usage change.

$$\Delta F= \beta_1 (\Delta T) + \beta_2 (\Delta P) + \beta_3 (\Delta W) + \beta_4 (\Delta S) + \beta_5 (\Delta H) + E \qquad (3.1)$$

Where β_1 to β_5 represent the coefficients of weather variables, T represents mean temperature, P represents precipitation, W represents wind speed, S represents snow on the ground, H represents sunlight hours and E represents estimation error. A Generalized Linear Model (GLM) is a statistical approach that is used to estimate the β's. This study used data collected with counters installed on cycling infrastructure sections. Data were collected on a daily basis (discrete data). A common GLM approach used for random discrete data is the Poisson regression model. In the Poisson regression model, the calculated mean and variance are equal. However, usually in real-world data, the mean and the variance are not equal (Cox, D 1983). Data with different values of mean and variance are referred to as overdispersed or underdispersed data. The Poisson model is not suitable for overdispersed or underdispersed data as the coefficients are underestimated and biased. To encounter the difference between mean and variance of data in the analysis, Negative Binomial (NB) regression models have been used. Negative binomial regression is a popular widely used method for discrete data with overdispersion or underdispersion (Poch and Mannering 1996). One limitation of NB models is the assumption of data independence. However, when data is collected from the same units at successive points over time, repeated observations are correlated. As a result, the NB model is not suitable for this study since cycling infrastructure usage was measured at successive times throughout the year. To encounter dependency of data, Generalized Estimating Equations (GEE) was first introduced by Liang and Zeger (1986). GEE is capable of estimating the coefficients with unbalanced and correlated data and have been used in transportation and traffic research with correlated data (Ma et al. 2010; Wang and Abdel. 2008; Lord and Persuad. 2000; Hutchings et al. 2003; Donnell et al. 2013). Generalized Estimating Equations (GEE) was chosen over other GLM analysis methods to analyze the data. Figure 3.12 shows the data measurement diagram.

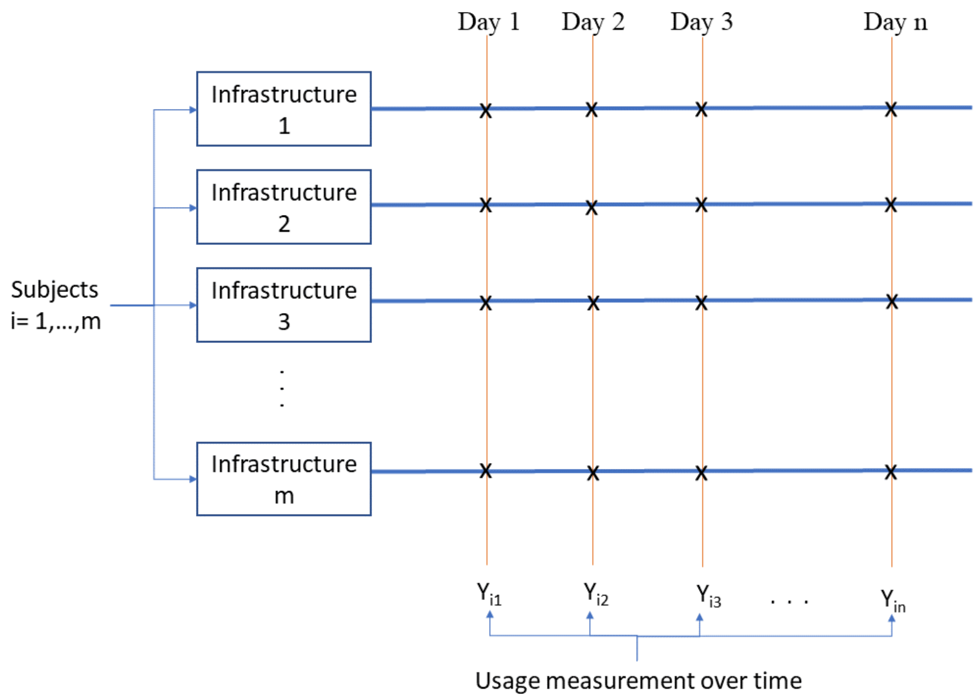


Figure 3.12 Data measurement diagram

In GEE, quasi-likelihood estimates of β's are computed from the maximization of the normality-based log likelihood without assuming that the response is normally distributed. In this study, the parameters β are coefficients for the weather variables, which are estimated by solving the following equation according to Liang and Zeger (1986):

$$\sum_{i=1}^{m} \left(\frac{\partial \mu_i}{\partial \beta}\right) V_i^{-1} (Y_i - \mu_i) = 0 \tag{3.2}$$

where m is the number of infrastructure sections, Y_i represent the vector of observations for each infrastructure section throughout the time, μ_i called the mean vector of Y_i which is the estimated value of Y_i as a function of input variables (observations) and β parameters, and V is the covariance matrix of observations, which takes into the account how changes in one variable are associated with the changes in other variables. The covariance matrix includes three elements:

$$V_i^{-1} = \phi\, A_i^{1/2}\, R_i\, A_i^{1/2} \tag{3.3}$$

where ϕ is an overdispersion parameter (scale parameter) which takes into account the difference in variability of dataset and model estimation; $A_i^{1/2}$ is an $n \times n$ diagonal matrix with square root of variances of observation where n represents the number of observations; and R_i is an $n \times n$ correlation matrix. The correlation between observations is carried out by using a correlation matrix in GEE. Different correlation matrix structures exist. Different correlation matrix structures are shown in Figure 3.13 where t represents the day with observation (e.g. t_1 represents day 1). Each correlation matrix element Rij represents the correlation between the i^{th} and j^{th} observation.

1. An independent correlation matrix indicates that observations do not have any correlation with each other.
2. An exchangeable correlation matrix indicates that all measurements are equally correlated.
3. An auto-regressive correlation matrix indicates that correlation depends on time or distance of measurements between two measurements.
4. An unstructured correlation matrix indicates that there is no specific assumption about the correlations.

$$\begin{array}{cc} & \begin{array}{ccccc} t_1 & t_2 & t_3 & & t_n \end{array} \\ \begin{array}{c} t_1 \\ t_2 \\ t_3 \\ \\ t_n \end{array} & \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \end{array}$$

Independent

$$\begin{array}{cc} & \begin{array}{ccccc} t_1 & t_2 & t_3 & & t_n \end{array} \\ \begin{array}{c} t_1 \\ t_2 \\ t_3 \\ \\ t_n \end{array} & \begin{bmatrix} 1 & \alpha & \alpha & \cdots & \alpha \\ \alpha & 1 & \alpha & \cdots & \alpha \\ \alpha & \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \vdots & & \vdots \\ \alpha & \alpha & \alpha & \cdots & 1 \end{bmatrix} \end{array}$$

Exchangeable

$$\begin{array}{cc} & \begin{array}{ccccc} t_1 & t_2 & t_3 & & t_n \end{array} \\ \begin{array}{c} t_1 \\ t_2 \\ t_3 \\ \\ t_n \end{array} & \begin{bmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^n \\ \alpha & 1 & \alpha & \cdots & \alpha^2 \\ \alpha^2 & \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \vdots & & \vdots \\ \alpha^n & \alpha^2 & \alpha & \cdots & 1 \end{bmatrix} \end{array}$$

Auto regressive

$$\begin{array}{cc} & \begin{array}{ccccc} t_1 & t_2 & t_3 & & t_n \end{array} \\ \begin{array}{c} t_1 \\ t_2 \\ t_3 \\ \\ t_n \end{array} & \begin{bmatrix} 1 & \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_1 & 1 & \alpha_m & \cdots & \alpha_t \\ \alpha_2 & \alpha_m & 1 & \cdots & \alpha_v \\ \vdots & \vdots & \vdots & & \vdots \\ \alpha_n & \alpha_t & \alpha_v & \cdots & 1 \end{bmatrix} \end{array}$$

Unstructured

Figure 3.13 Four common correlation structures used in GEE

The GEE model is first fitted by computing initial estimates of β, for example, using a naive linear regression. After computing initial β estimates, GEE estimated the dispersion parameter from residuals and computed the correlation matrix based on residuals. Then, the working covariance matrix was computed and β was estimated again. These steps are continued until convergence of the β's is obtained. The flowchart of the GEE analysis is shown in Figure 3.14.

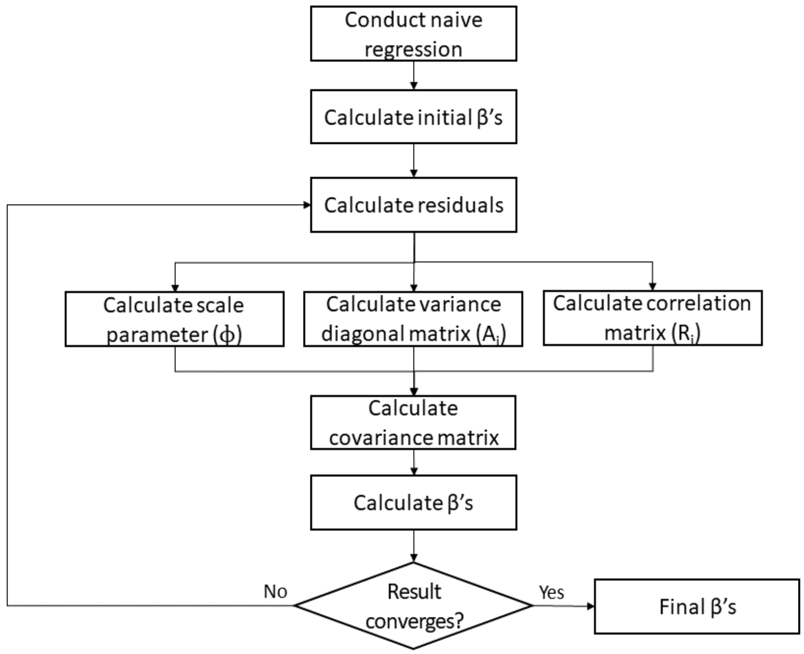
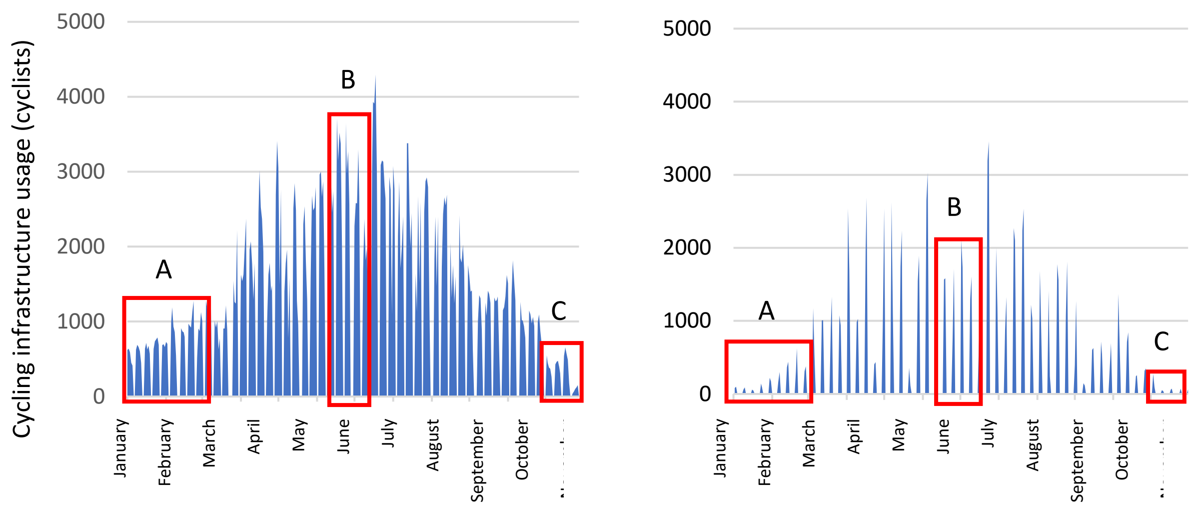


Figure 3.14 GEE flowchart

3.3 MODEL DEVELOPMENT

A closer look at the daily usage revealed that there is a difference in usage during weekdays and weekends. Figure 3.15 shows daily usage of one cycling infrastructure section during weekdays and weekends. For example, as it can be seen in Figure 3.15, during three different periods of the year (early in the year, middle of the year and end of the year) weekend and weekday usages are different in values (A, B and C). Due to differences in usage on weekdays and weekends, two separate models were developed. The first model was developed based on weekday usage, and the second model was developed based on weekend usage.



Weekday usage Weekend usage

Figure 3.15 Difference in infrastructure usage trends during weekdays and weekends

Since the model estimates usage based only on the weather variable, other factors that affect cycling usage were removed to increase model accuracy. For example, national holidays were excluded from the database. To analyze data, the change in usage for each day of the week in each month was compared to the same day of the week in September to eliminate the impact of the day of the week. Also, since cycling infrastructure usages collected in this study had different standard deviations for different infrastructure sections, a logarithmic data transformation was applied. Consequently, the logarithmic value of cycling infrastructure usage was used to develop the model.

Effect of weather variables on people's decision to cycle is not constant over a year. The effect of preventing factors, such as precipitation, are assumed to be greater in cold months due to the compounding effect of rain and cold temperatures. The variation in the magnitude of the effect of the variables makes it necessary to have separate models for different time periods in a year. Miranda-Moreno et al. (2013) found that although hourly and weekly usage patterns were consistent in different cities with different characteristics, monthly usage patterns were different due to the impact of weather variables. Hence, monthly periods were selected for developing sub-models. Therefore, 11 sub-models were developed. Each sub-model evaluates the effect of weather conditions on changes in cycling frequency for each month of the year compared to September.

Correlation structure

Cycling infrastructure usage was measured at repeated times. Since the independent correlation structure is not a reasonable choice for the correlation matrix of data with repeated measurement, it was excluded from the choices. Exchangeable structure indicates that all the observations are equally correlated regardless of how distant in time the observations are made. This assumption is not suitable for this study as it is believed that observations that are closer to each other are more correlated. Auto-regressive correlation is good choice of correlation structure when the time intervals between observations is constant (time interval between the observations this study is 1 day). This structure indicates that the correlation between observations has an exponential decay over time which means the observations next to each other are more correlated than observations far from each other (waters 2017). Unstructured correlation is also a suitable correlation structure for this study with assigning different correlations between each pair of observations. However, it adds several parameters to the model and makes the model more complex with higher number of parameters. Among the correlation structures, auto-regressive and unstructured seems to fit the data best due to data nature. The three correlation structures were tested on model. The most fitted correlation structure was selected for the model based on the Quasi-likelihood under Independence Criterion (QIC). The most well-know criterion that is used for estimating statistical models is Akaike Information Criterion (AIC). AIC is a likelihood-based estimator and it is not suitable for no likelihood-based approaches such as GEE. The QIC criterion was first introduced by Pan (2001), by replacing the likelihood in the AIC with quasi-likelihood. The QIC can be expressed as:

$$QIC = -2\,\mathrm{Q}\,(\mu, \mathrm{I}) + 2\,\mathrm{trace}\,(\Omega_I\, \mathrm{V}_R) \tag{3.4}$$

where Q (μ, I) represent quasi-likelihood computed using independent structure, trace refers to the sum of diagonal elements of the matrix, Ω_I represent the inverse of the variance matrix fitting an independence model, V_R represents a modified sandwich estimate of variance from the model. The model with the smallest value of QIC is preferred. QIC value for all correlation structures was calculated. Within correlation structures, auto-regressive correlation was selected as it showed the best goodness of fit with lowest QIC value. The reason that unstructured correlation showed a higher QIC values is because it adds several parameters to the model and increase the model complexity. Table 3.3 shows the QIC values of three correlation structures on sub-models of weekday and weekend.

Table 3.3 QIC values of GEE correlation structures

	Weekday usage			Weekend usage		
Month	Auto-regressive	unstructured	exchangeable	Auto-regressive	unstructured	exchangeable
January	115.34	125.58	117.92	52.55	54.35	52.82
February	106.10	88.76	98.53	64.26	56.94	64.51
March	91.72	159.38	94.67	41.76	31.39	40.60
April	90.54	1247.51	92.54	31.41	70.66	31.53
May	78.12	203.03	82.10	38.52	44.22	37.18
June	57.36	179.39	72.51	21.82	68.71	24.36
July	89.94	84.11	99.08	21.68	26.84	21.54
August	98.61	1050.30	104.33	33.23	54.43	32.26
September	-	-	-	-	-	-
October	68.60	114.04	64.48	26.94	77.38	26.31
November	84.06	141.72	79.32	19.49	25.95	19.50
December	69.56	54.11	73.54	34.77	45.58	36.85
Average	35.13	50.58	35.22	86.35	313.44	89.00

First Model

The first model was developed using weekday usage and weather data. Table 3.4 provides the results obtained from the analysis of weekday usage using the 11 sub-models. The coefficients represent the impact of each weather variable on the log value of infrastructure usage. Some of the coefficients are not significant nor rational. For example, snow on the ground in winter can not positively impact cycling infrastructure usage. Therefore, those coefficients are removed from the model.

Table 3.4 First model coefficients

Month	Temp	Total Precipitation	Snow on Ground	Wind Speed	Hours of Sunlight
January	0.021	-0.019		-0.003	0.097
February	0.02	-0.029		-0.005	
March	0.017	-0.032	-0.008	-0.006	
April	0.025	-0.017		-0.009	
May	0.02	-0.033		-0.006	
June	0.018	-0.011		-0.005	
July	0.027	-0.009		-0.006	
August	0.022	-0.013		-0.007	
September	-	-	-	-	-
October	0.015	-0.026	-0.052	-0.003[1]	
November	0.014	-0.017	-0.021	-0.006	
December	0.017	-0.018	-0.035	-0.004	

1.significant at p-value <0.05. All the other presented coefficients are significant at p-value <0.01

Second Model

The second model is similar to the first model, but instead of weekday usage, it was developed based on weekend usage. Table 3.5 provides the results obtained from an analysis on weekend usage for the 11 sub-models. The second model has fewer significant coefficients compared to the first model because it has a smaller sample size.

Table 3.5 Second model coefficients

Month	Temp	Total Precipitation	Snow on Ground	Wind Speed	Hours of Sunlight
January	0.034	-0.023		-0.006	
February	0.032	-0.014		-0.015	
March	0.032	-0.046	-0.016		
April	0.04	-0.015		-0.009	
May	0.03	-0.023		-0.012	
June	0.023	-0.033			
July	0.027	-0.028			
August	0.028	-0.016		-0.010	
September	-	-	-	-	-
October	0.025[1]	-0.015	-0.092		
November	0.039	-0.044			
December	0.023	-0.023	-0.017[1]	-0.012	

1.significant at p-value <0.05. All the other presented coefficients are significant at p-value <0.01

Estimation Results

Table 3.6 Shows the p-value and 95% confidence interval obtained from Wald Chi-Square test for GEE scale parameter and variables with significant impact on weekday usage.

Table 3.6 GEE Estimation results on weekday usage

Month	Variable	Estimate (β)	p-Value	95% Confidence Interval	
				Lower	Upper
January	Temperature	.021	.000	.017	.025
	Precipitation	-.019	.001	-.029	-.008
	Wind speed	-.003	.000	-.004	-.002
	sunlight hour	.097	.000	.052	.142
	scale	.106			
February	Intercept	.169	.017	.030	.307
	Temperature	.020	.000	.016	.023
	Precipitation	-.029	.000	-.038	-.019
	Wind speed	-.005	.000	-.007	-.003
	scale	.098		.016	.023
March	Intercept	.193	.000	.147	.238
	Temperature	.017	.000	.012	.022
	Snow	-.008	.005	-.014	-.003
	Precipitation	-.032	.000	-.049	-.015
	Wind speed	-.006	.000	-.008	-.005
	scale	.076			
April	Intercept	-.083	.018	-.152	-.014
	Temperature	.025	.000	.022	.028
	Precipitation	-.017	.000	-.021	-.013
	Wind speed	-.009	.000	-.013	-.006
	scale	.081			
May	Intercept	-.188	.000	-.291	-.085
	Temperature	.020	.000	.016	.024
	Precipitation	-.033	.000	-.041	-.025
	Wind speed	-.006	.000	-.008	-.004
	scale	.071			

Table 3.7 GEE Estimation results on weekday usage (continued)

Month	Variable	Estimate (β)	p-Value	95% Confidence Interval	
				Lower	Upper
	Intercept	-.165	.007	-.286	-.045
June	Temperature	.018	.007	.015	.021
	Precipitation	-.011	.000	-.014	-.008
	Wind speed	-.005	.000	-.007	-.003
	scale	.036			
	Temperature	.027	.000	.022	.031
	Precipitation	-.009	.000	-.010	-.008
July	Wind speed	-.006	.000	-.007	-.004
	scale	.057			
	Temperature	.022	.000	.017	.027
	Precipitation	-.013	.000	-.016	-.010
August	Wind speed	-.007	.000	-.009	-.006
	scale	.065			
	Temperature	.015	.000	.010	.020
	Snow	-.052	.000	-.070	-.035
October	Precipitation	-.026	.000	-.032	-.020
	Wind speed	-.003	.023	-.005	.000
	scale	.055			
	Temperature	.014	.000	.012	.017
	Snow	-.021	.000	-.032	-.010
November	Precipitation	-.017	.000	-.023	-.011
	Wind speed	-.006	.000	-.009	-.004
	scale	.014			
	Intercept	.757	.000	.570	.945
	Temperature	.017	.000	.014	.020
	Snow	-.035	.000	-.042	-.028
December	Precipitation	-.018	.000	-.027[1]	-.010
	Wind speed	-.004	.000	-.005	-.002
	scale	.049			

Table 3.8 Shows the p-value and 95% confidence interval obtained from Wald Chi-Square test for GEE scale parameter and variables with significant impact on weekend usage.

Table 3.8 GEE Estimation results on weekend usage

Month	Variable	Estimate (β)	p-Value	95% Confidence Interval	
				Lower	Upper
January	Intercept	.548	.000	.368	.728
	Temperature	.034	.000	.028	.040
	Precipitation	-.023	.000	-.027	-.019
	Wind speed	-.006	.001	-.010	-.003
	scale	.106			
February	Intercept	.379	.000	.256	.503
	Temperature	.032	.000	.026	.039
	Precipitation	-.014	.006	-.023	-.004
	Wind speed	-.015	.000	-.021	-.009
	scale	.144		.026	.039
March	Temperature	.032	.000	.027	.038
	Snow	-.016	.001	-.026	-.006
	Precipitation	-.046	.000	-.054	-.039
	sunlight hour	.080	.000	.038	.121
	scale	.032			
April	Intercept	-.157	.000	-.231	-.083
	Temperature	.040	.000	.036	.044
	Precipitation	-.015	.000	-.020	-.010
	Wind speed	-.009	.000	-.012	-.006
	scale	.057			
May	Intercept	-.197	.001	-.314	-.079
	Temperature	.030	.000	.026	.035
	Precipitation	-.023	.000	-.031	-.014
	Wind speed	-.012	.000	-.017	-.007
	scale	.066			

Table 3.9 GEE Estimation results on weekend usage (continued)

Month	Variable	Estimate (β)	p-Value	95% Confidence Interval	
				Lower	Upper
	Intercept	-.198	.000	-.302	-.094
June	Temperature	.023	.000	.020	.027
	Precipitation	-.033	.000	-.039	-.027
	scale	.032			
	Intercept	.486	.000	.274	.698
	Temperature	.027	.000	.022	.032
July	Precipitation	-.028	.000	-.031	-.026
	scale	.033			
	Temperature	.028	.000	.023	.033
August	Precipitation	-.016	.000	-.018	-.014
	Wind speed	-.010	.000	-.015	-.005
	scale	.039			
	Intercept	-.973	.032	-1.86	-.136
	Temperature	.025	.000	.017	.034
	Snow	-.092	.000	-.127	-.058
October	Precipitation	-.015	.000	-.020	-.010
	scale	.096			
	Intercept	1.123	.000	.737	1.509
	Temperature	.039	.000	.032	.047
November	Precipitation	-.044	.000	-.054	-.034
	scale	.037			
	Intercept	1.528	.000	.840	2.216
	Temperature	.023	.000	.018	.028
	Snow	-.017	.043	-.034	-.001
December	Precipitation	-.023	.000	-.028	-.018
	Wind speed	-.012	.000	-.015	-.009
	scale	.099			

Analysis menu

The analysis was conducted using Microsoft Excel and SPSS software. First raw data was entered into excel, the logarithmic value of usage was calculated and usage was clustered based on the month of the year in Excel. Figure 3.16 shows weekdays usage data in excel. The first column indicates the infrastructure section assigned number (pat A). Part B in the figure shows the logarithmic value of usage for different months of the year. Part C shows three out of five weather variables (i.e. sunlight hours, temperature, precipitation). Each row in excel includes the usage in one day of the week for all months with weather variables assigned to that specific day in different months.

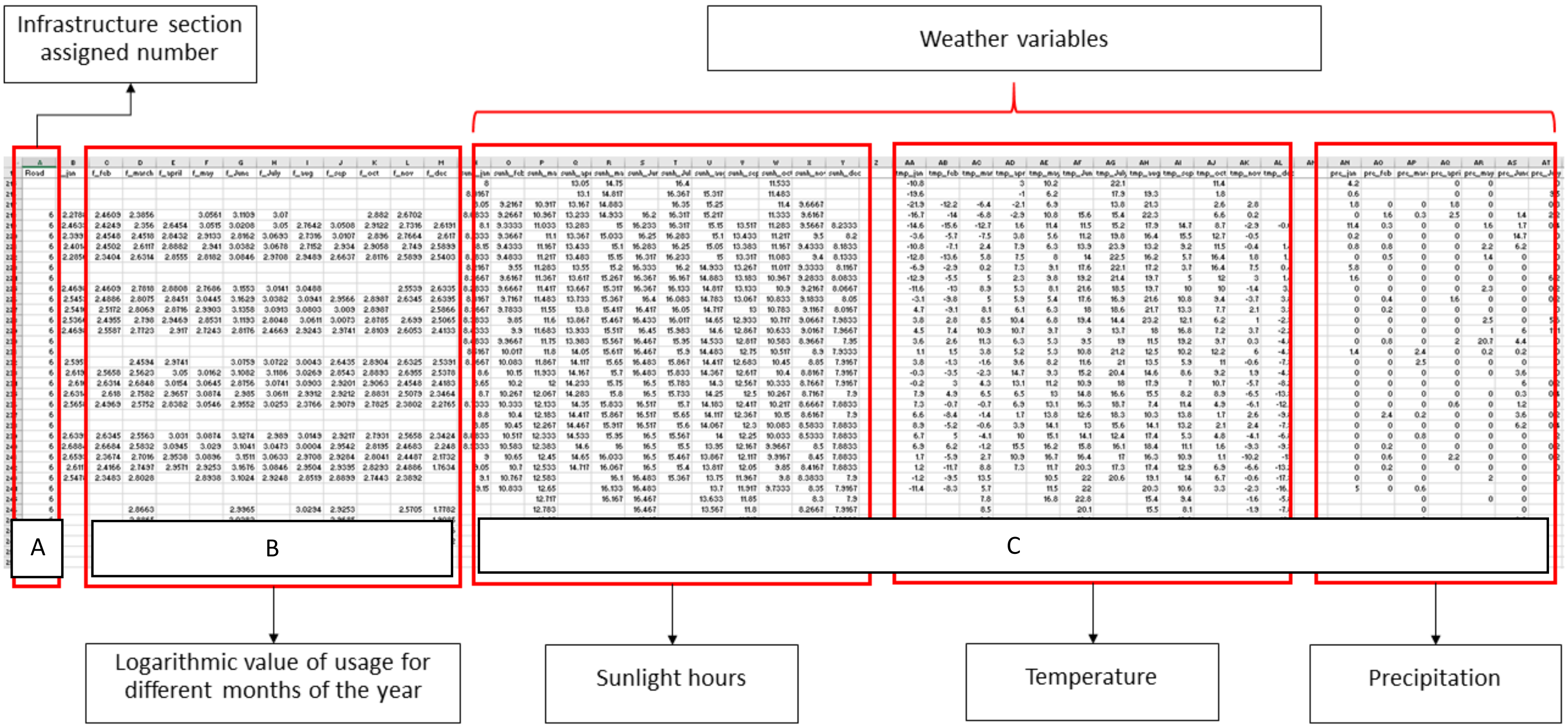
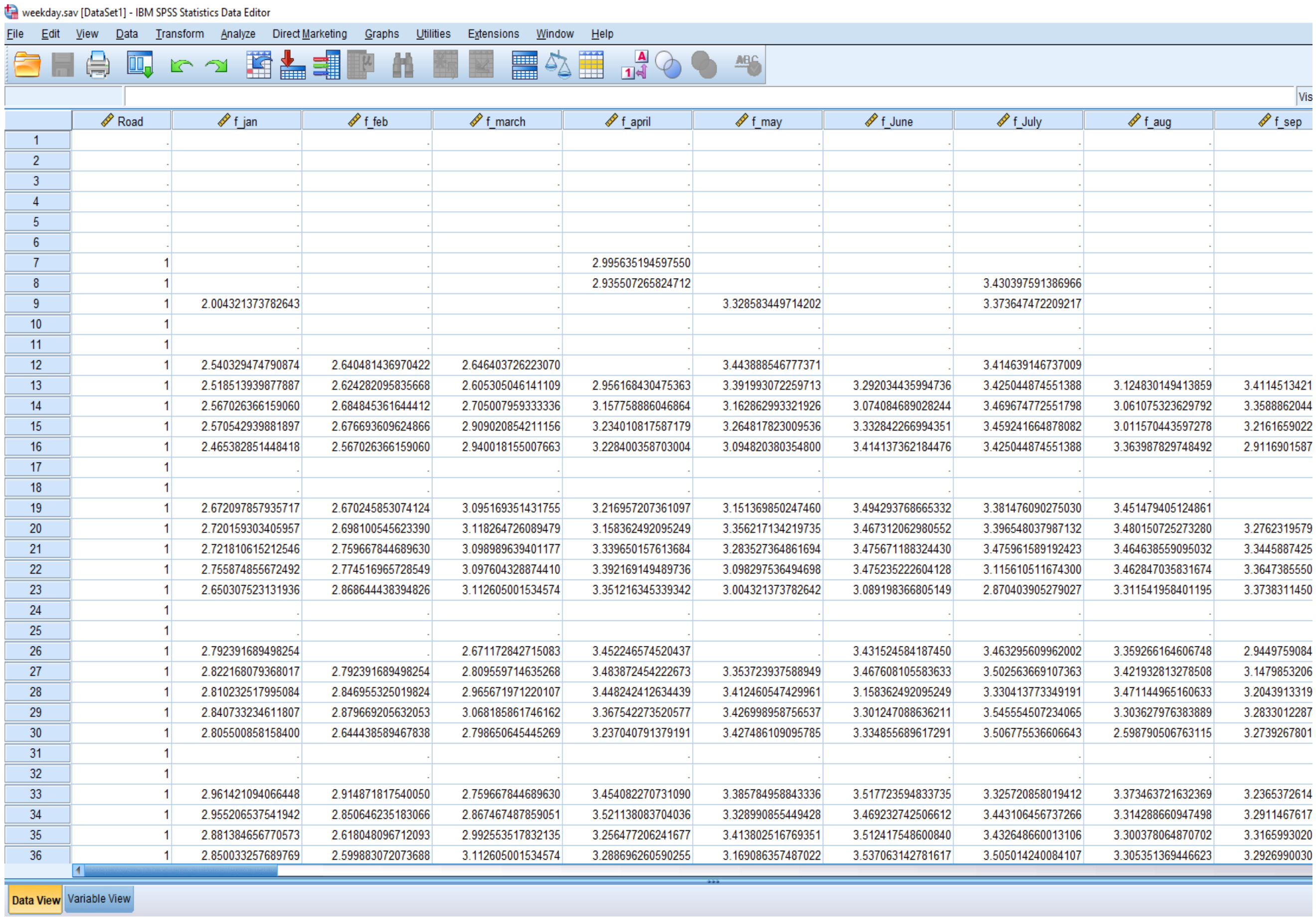


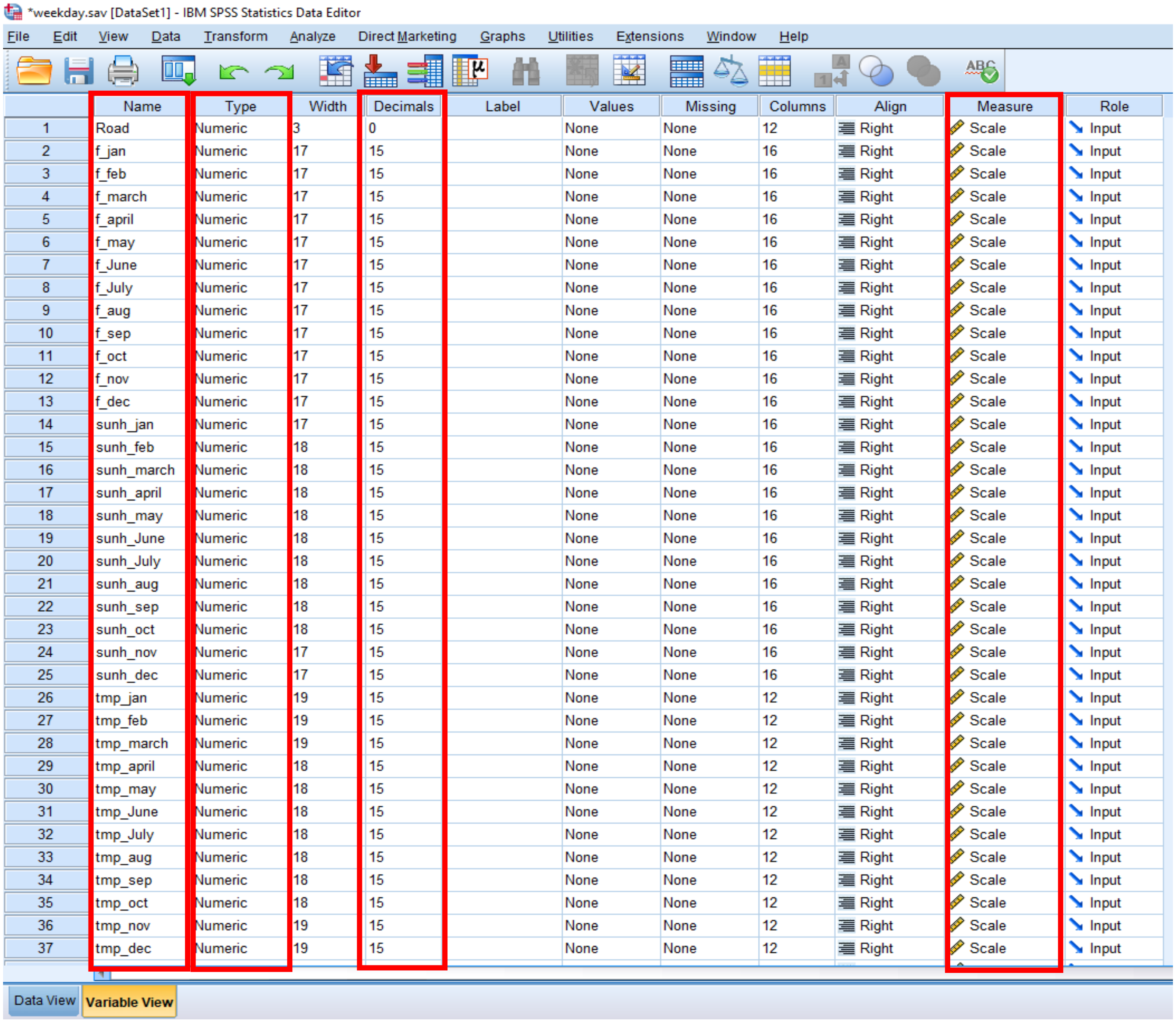
Figure 3.16 Snapshot of data in excel

For the next step data was converted in SPSS file. Figure 3.17 shows weekdays data converted from excel to SPSS file. Figure 3.18 shows the characteristics of variables in SPSS. The characteristics of cycling usage and two weather variables (i.e. temperature and sunlight hours) for different months of the year can be seen in the figure. The first column in the figure shows the variables name and the second column shows variables type. Different variable types such as numeric, restricted numeric (intergern), string and date are available for each variable. This study used numeric data. The fourth column shows variables decimals. Except for the first variable (Road) which is an integer variable that shows road assigned number, all variables are shown with 15 decimals. Variables values can be seen in Figure 3.17. The 10^{th} column in Figure 3.18 shows the type of measurement of variables. Variables measurement can be nominal, ordinal or scale. Nominal measurement refers to categorical variables which it is not possible to rank the categories (e.g. gender). Ordinal measurement refers to categorized variables that categories could be ranked. However, the intervals between categories are not defined (e.g. Likert scale). Scale measurement has the same interval between all the measurements (e.g. temperature). In this study, all the variables have scale measurement.



	Road	f_jan	f_feb	f_march	f_april	f_may	f_June	f_July	f_aug	f_sep
1	.	.	.	.	.	.	.	.	.	
2	.	.	.	.	.	.	.	.	.	
3	.	.	.	.	.	.	.	.	.	
4	.	.	.	.	.	.	.	.	.	
5	.	.	.	.	.	.	.	.	.	
6	.	.	.	.	.	.	.	.	.	
7	1	.	.	.	2.995635194597550	.	.	.	.	
8	1	.	.	.	2.935507265824712	.	.	3.430397591386966	.	
9	1	2.004321373782643	.	.	.	3.328583449714202	.	3.373647472209217	.	
10	1	.	.	.	.	.	.	.	.	
11	1	.	.	.	.	.	.	.	.	
12	1	2.540329474790874	2.640481436970422	2.646403726223070	.	3.443888546777371	.	3.414639146737009	.	
13	1	2.518513939877887	2.624282095835668	2.605305046141109	2.956168430475363	3.391993072259713	3.292034435994736	3.425044874551388	3.124830149413859	3.4114513421
14	1	2.567026366159060	2.684845361644412	2.705007959333336	3.157758886046864	3.162862993321926	3.074084689028244	3.469674772551798	3.061075323629792	3.3588862044
15	1	2.570542939881897	2.676693609624866	2.909020854211156	3.234010817587179	3.264817823009536	3.332842266994351	3.459241664878082	3.011570443597278	3.2161659022
16	1	2.465382851448418	2.567026366159060	2.940018155007663	3.228400358703004	3.094820380354800	3.414137362184476	3.425044874551388	3.363987829748492	2.9116901587
17	1	.	.	.	.	.	.	.	.	
18	1	.	.	.	.	.	.	.	.	
19	1	2.672097857935717	2.670245853074124	3.095169351431755	3.216957207361097	3.151369850247460	3.494293768665332	3.381476090275030	3.451479405124861	
20	1	2.720159303405957	2.698100545623390	3.118264726089479	3.158362492095249	3.356217134219735	3.467312062980552	3.396548037987132	3.480150725273280	3.2762319579
21	1	2.721810615212546	2.759667844689630	3.098989639401177	3.339650157613684	3.283527364861694	3.475671188324430	3.475961589192423	3.464638559095032	3.3445887425
22	1	2.755874855672492	2.774516965728549	3.097604328874410	3.392169149489736	3.098297536494698	3.475235222604128	3.115610511674300	3.462847035831674	3.3647385550
23	1	2.650307523131936	2.868644438394826	3.112605001534574	3.351216345339342	3.004321373782642	3.089198366805149	2.870403905279027	3.311541958401195	3.3738311450
24	1	.	.	.	.	.	.	.	.	
25	1	.	.	.	.	.	.	.	.	
26	1	2.792391689498254	.	2.671172842715083	3.452246574520437	.	3.431524584187450	3.463295609962002	3.359266164606748	2.9449759084
27	1	2.822168079368017	2.792391689498254	2.809559714635268	3.483872454222673	3.353723937588949	3.467608105583633	3.502563669107363	3.421932813278508	3.1479853206
28	1	2.810232517995084	2.846955325019824	2.965671971220107	3.448242412634439	3.412460547429961	3.158362492095249	3.330413773349191	3.471144965160633	3.2043913319
29	1	2.840733234611807	2.879669205632053	3.068185861746162	3.367542273520577	3.426998958756537	3.301247088636211	3.545554507234065	3.303627976383889	3.2833012287
30	1	2.805500858158400	2.644438589467838	2.798650645445269	3.237040791379191	3.427486109095785	3.334855689617291	3.506775536606643	2.598790506763115	3.2739267801
31	1	.	.	.	.	.	.	.	.	
32	1	.	.	.	.	.	.	.	.	
33	1	2.961421094066448	2.914871817540050	2.759667844689630	3.454082270731090	3.385784958843336	3.517723594833735	3.325720858019412	3.373463721632369	3.2365372614
34	1	2.955206537541942	2.850646235183066	2.867467487859051	3.521138083704036	3.328990855449428	3.469232742506612	3.443106456737266	3.314288660947498	3.2911467617
35	1	2.881384656770573	2.618048096712093	2.992553517832135	3.256477206241677	3.413802516769351	3.512417548600840	3.432648660013106	3.300378064870702	3.3165993020
36	1	2.850033257689769	2.599883072073688	3.112605001534574	3.288696260590255	3.169086357487022	3.537063142781617	3.505014240084107	3.305351369446623	3.2926990030

Figure 3.17 Snapshot of data view in SPSS



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Road	Numeric	3	0		None	None	12	Right	Scale	Input
2	f_jan	Numeric	17	15		None	None	16	Right	Scale	Input
3	f_feb	Numeric	17	15		None	None	16	Right	Scale	Input
4	f_march	Numeric	17	15		None	None	16	Right	Scale	Input
5	f_april	Numeric	17	15		None	None	16	Right	Scale	Input
6	f_may	Numeric	17	15		None	None	16	Right	Scale	Input
7	f_June	Numeric	17	15		None	None	16	Right	Scale	Input
8	f_July	Numeric	17	15		None	None	16	Right	Scale	Input
9	f_aug	Numeric	17	15		None	None	16	Right	Scale	Input
10	f_sep	Numeric	17	15		None	None	16	Right	Scale	Input
11	f_oct	Numeric	17	15		None	None	16	Right	Scale	Input
12	f_nov	Numeric	17	15		None	None	16	Right	Scale	Input
13	f_dec	Numeric	17	15		None	None	16	Right	Scale	Input
14	sunh_jan	Numeric	17	15		None	None	16	Right	Scale	Input
15	sunh_feb	Numeric	18	15		None	None	16	Right	Scale	Input
16	sunh_march	Numeric	18	15		None	None	16	Right	Scale	Input
17	sunh_april	Numeric	18	15		None	None	16	Right	Scale	Input
18	sunh_may	Numeric	18	15		None	None	16	Right	Scale	Input
19	sunh_June	Numeric	18	15		None	None	16	Right	Scale	Input
20	sunh_July	Numeric	18	15		None	None	16	Right	Scale	Input
21	sunh_aug	Numeric	18	15		None	None	16	Right	Scale	Input
22	sunh_sep	Numeric	18	15		None	None	16	Right	Scale	Input
23	sunh_oct	Numeric	18	15		None	None	16	Right	Scale	Input
24	sunh_nov	Numeric	17	15		None	None	16	Right	Scale	Input
25	sunh_dec	Numeric	17	15		None	None	16	Right	Scale	Input
26	tmp_jan	Numeric	19	15		None	None	12	Right	Scale	Input
27	tmp_feb	Numeric	19	15		None	None	12	Right	Scale	Input
28	tmp_march	Numeric	19	15		None	None	12	Right	Scale	Input
29	tmp_april	Numeric	18	15		None	None	12	Right	Scale	Input
30	tmp_may	Numeric	18	15		None	None	12	Right	Scale	Input
31	tmp_June	Numeric	18	15		None	None	12	Right	Scale	Input
32	tmp_July	Numeric	18	15		None	None	12	Right	Scale	Input
33	tmp_aug	Numeric	18	15		None	None	12	Right	Scale	Input
34	tmp_sep	Numeric	18	15		None	None	12	Right	Scale	Input
35	tmp_oct	Numeric	18	15		None	None	12	Right	Scale	Input
36	tmp_nov	Numeric	19	15		None	None	12	Right	Scale	Input
37	tmp_dec	Numeric	19	15		None	None	12	Right	Scale	Input

Figure 3.18 Snapshot of Variables view in SPSS

The next step was conducting GEE analysis. Figure 3.19 shows the code used in SPSS for conducting GEE analysis. First, as shown in box 1, the difference between the usage of each month was computed compared to the reference month (September). Second, the difference between the five weather variables for each month was computed compared to the reference month (box 2). Then, GEE analysis was conducted based on the usage change and weather variables change (box 3).

1

```
get file='C:\Users\siroo\Desktop\article\week 44\without 3.10.12.log.sa
compute f_avg=mean(f_sep,f_sep).

compute f_avg_jan=(f_avg-f_jan).
compute f_avg_feb=(f_avg-f_feb).
compute f_avg_march=(f_avg-f_march).
compute f_avg_april=(f_avg-f_april).
compute f_avg_may=(f_avg-f_may).
compute f_avg_june=(f_avg-f_june).
compute f_avg_july=(f_avg-f_july).
compute f_avg_aug=(f_avg-f_aug).
compute f_avg_sep=(f_avg-f_sep).
compute f_avg_oct=(f_avg-f_oct).
compute f_avg_nov=(f_avg-f_nov).
compute f_avg_dec=(f_avg-f_dec).
********************************


*************************************************
compute tmp_avg=mean(tmp_sep,tmp_sep).

compute tmp_avg_jan=(tmp_avg-tmp_jan).
compute tmp_avg_feb=(tmp_avg-tmp_feb).
compute tmp_avg_march=(tmp_avg-tmp_march).
compute tmp_avg_april=(tmp_avg-tmp_april).
compute tmp_avg_may=(tmp_avg-tmp_may).
compute tmp_avg_june=(tmp_avg-tmp_june).
compute tmp_avg_july=(tmp_avg-tmp_july).
compute tmp_avg_aug=(tmp_avg-tmp_aug).
compute tmp_avg_sep=(tmp_avg-tmp_sep).
compute tmp_avg_oct=(tmp_avg-tmp_oct).
compute tmp_avg_nov=(tmp_avg-tmp_nov).
compute tmp_avg_dec=(tmp_avg-tmp_dec).
*************************************************
compute s_avg=mean(s_sep,s_sep).

compute s_avg_jan=(s_avg-s_jan).
compute s_avg_feb=(s_avg-s_feb).
compute s_avg_march=(s_avg-s_march).
compute s_avg_april=(s_avg-s_april).
compute s_avg_may=(s_avg-s_may).
compute s_avg_june=(s_avg-s_june).
compute s_avg_july=(s_avg-s_july).
compute s_avg_aug=(s_avg-s_aug).
compute s_avg_sep=(s_avg-s_sep).
compute s_avg_oct=(s_avg-s_oct).
compute s_avg_nov=(s_avg-s_nov).
compute s_avg_dec=(s_avg-s_dec).
```

2

```
*************************************************
compute pre_avg=mean(pre_sep,pre_sep).

compute pre_avg_jan=(pre_avg-pre_jan).
compute pre_avg_feb=(pre_avg-pre_feb).
compute pre_avg_march=(pre_avg-pre_march).
compute pre_avg_april=(pre_avg-pre_april).
compute pre_avg_may=(pre_avg-pre_may).
compute pre_avg_june=(pre_avg-pre_june).
compute pre_avg_july=(pre_avg-pre_july).
compute pre_avg_aug=(pre_avg-pre_aug).
compute pre_avg_sep=(pre_avg-pre_sep).
compute pre_avg_oct=(pre_avg-pre_oct).
compute pre_avg_nov=(pre_avg-pre_nov).
compute pre_avg_dec=(pre_avg-pre_dec).
*************************************************
compute w_avg=mean(w_sep,w_sep).

compute w_avg_jan=(w_avg-w_jan) .
compute w_avg_feb=(w_avg-w_feb) .
compute w_avg_march=(w_avg-w_march) .
compute w_avg_april=(w_avg-w_april) .
compute w_avg_may=(w_avg-w_may) .
compute w_avg_june=(w_avg-w_june) .
compute w_avg_july=(w_avg-w_july) .
compute w_avg_aug=(w_avg-w_aug) .
compute w_avg_sep=(w_avg-w_sep) .
compute w_avg_oct=(w_avg-w_oct) .
compute w_avg_nov=(w_avg-w_nov) .
compute w_avg_dec=(w_avg-w_dec) .
*************************************************
compute sunh_avg=mean(sunh_sep,sunh_sep).

compute sunh_avg_jan=(sunh_avg-sunh_jan) .
compute sunh_avg_feb=(sunh_avg-sunh_feb) .
compute sunh_avg_march=(sunh_avg-sunh_march) .
compute sunh_avg_april=(sunh_avg-sunh_april) .
compute sunh_avg_may=(sunh_avg-sunh_may) .
compute sunh_avg_june=(sunh_avg-sunh_june) .
compute sunh_avg_july=(sunh_avg-sunh_july) .
compute sunh_avg_aug=(sunh_avg-sunh_aug) .
compute sunh_avg_sep=(sunh_avg-sunh_sep) .
compute sunh_avg_oct=(sunh_avg-sunh_oct) .
compute sunh_avg_nov=(sunh_avg-sunh_nov) .
compute sunh_avg_dec=(sunh_avg-sunh_dec) .
descriptives variables=all.
```

3

```
genlin f_avg_april  with    tmp_avg_april s_avg_april   pre_avg_april  w_avg_april sunh_avg_april
                            /model    tmp_avg_april s_avg_april   pre_avg_april  w_avg_april sunh_avg_april
  distribution=normal link=identity
 /repeated  SUBJECT=road corrtype=ar(1).
********************************
genlin f_avg_may  with    tmp_avg_may s_avg_may   pre_avg_may  w_avg_may sunh_avg_may
                            /model    tmp_avg_may s_avg_may   pre_avg_may  w_avg_may sunh_avg_may
  distribution=normal link=identity
 /repeated  SUBJECT=road corrtype=ar(1).
genlin f_avg_june  with    tmp_avg_june s_avg_june   pre_avg_june  w_avg_june sunh_avg_june
                            /model    tmp_avg_june s_avg_june   pre_avg_june  w_avg_june sunh_avg_june
  distribution=normal link=identity
 /repeated  SUBJECT=road corrtype=ar(1).
********************************
genlin f_avg_july  with    tmp_avg_july s_avg_july   pre_avg_july  w_avg_july sunh_avg_july
                            /model    tmp_avg_july s_avg_july   pre_avg_july  w_avg_july sunh_avg_july
  distribution=normal link=identity
 /repeated  SUBJECT=road corrtype=ar(1).
********************************
genlin f_avg_aug  with    tmp_avg_aug s_avg_aug   pre_avg_aug  w_avg_aug sunh_avg_aug
                            /model    tmp_avg_aug s_avg_aug   pre_avg_aug  w_avg_aug sunh_avg_aug
  distribution=normal link=identity
 /repeated  SUBJECT=road corrtype=ar(1).
********************************


genlin f_avg_sep  with    tmp_avg_sep s_avg_sep   pre_avg_sep  w_avg_sep sunh_avg_sep
                            /model    tmp_avg_sep s_avg_sep   pre_avg_sep  w_avg_sep sunh_avg_sep
  distribution=normal link=identity
 /repeated  SUBJECT=road corrtype=ar(1).
********************************
genlin f_avg_oct  with    tmp_avg_oct s_avg_oct   pre_avg_oct  w_avg_oct sunh_avg_oct
                            /model    tmp_avg_oct s_avg_oct   pre_avg_oct  w_avg_oct sunh_avg_oct
  distribution=normal link=identity
 /repeated  SUBJECT=road corrtype=ar(1).
********************************


genlin f_avg_nov  with    tmp_avg_nov s_avg_nov   pre_avg_nov  w_avg_nov sunh_avg_nov
                            /model    tmp_avg_nov s_avg_nov   pre_avg_nov  w_avg_nov sunh_avg_nov
  distribution=normal link=identity
 /repeated  SUBJECT=road corrtype=ar(1).


**************************************
genlin f_avg_dec  with    tmp_avg_dec s_avg_dec   pre_avg_dec  w_avg_dec sunh_avg_dec
                            /model    tmp_avg_dec s_avg_dec   pre_avg_dec  w_avg_dec sunh_avg_dec
  distribution=normal link=identity
 /repeated  SUBJECT=road corrtype=ar(1).
*********************
```

Figure 3.19 Snapshot of the GEE code in SPSS

3.4 DISCUSSION AND COMPARISON

These two models evaluated the impact of weather variables on cycling infrastructure usage. The results showed that among weather variables, temperature positively impacted cycling infrastructure usage, which was in agreement with the existing literature (Flynn et al., 2012; Saneinejad et al., 2012; Brandenburg et al., 2007; Bergström and Magnusson, 2003). Also, the results indicated that the impact of temperature was not the same on weekend cyclists and weekday cyclists. Weekend cyclists were more affected by temperature. This finding was in line with the literature (Miranda-Moreno et al., 2013; Brandenburg et al., 2007). The results of the impact of temperature on infrastructures usage on weekdays and weekends is shown in Figure 3.20.

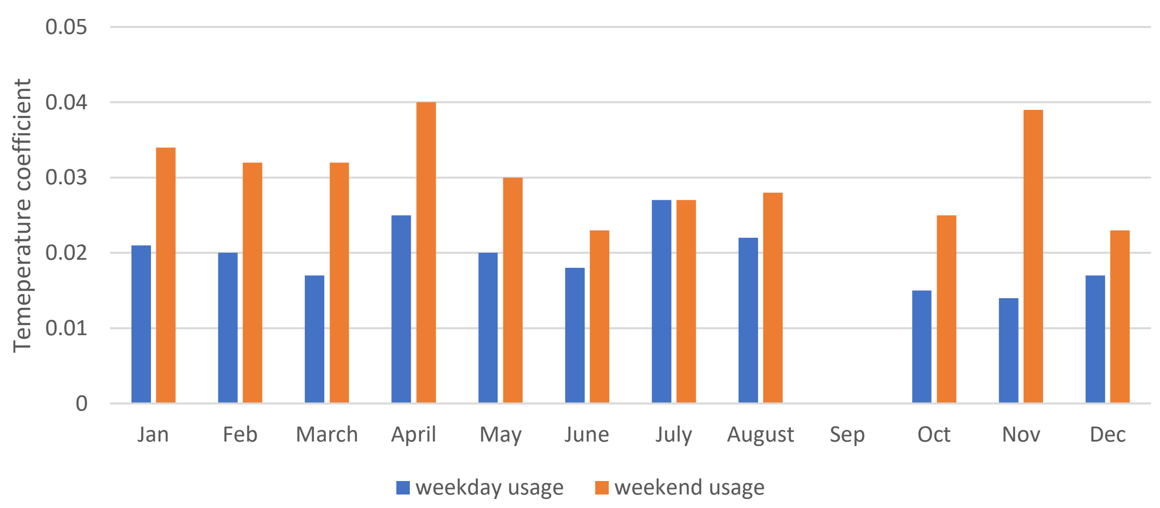


Figure 3.20 Impact of temperature on cycling infrastructure usage on weekdays and weekends

Precipitation is another weather variable that impacted cycling. Precipitation showed a negative impact on cycling infrastructure usage on both weekdays and weekend cyclists. However, the magnitude of the impact of precipitation was larger for weekend cyclists, especially in the summer. The effect of precipitation on infrastructure usage is shown in Figure 3.21.

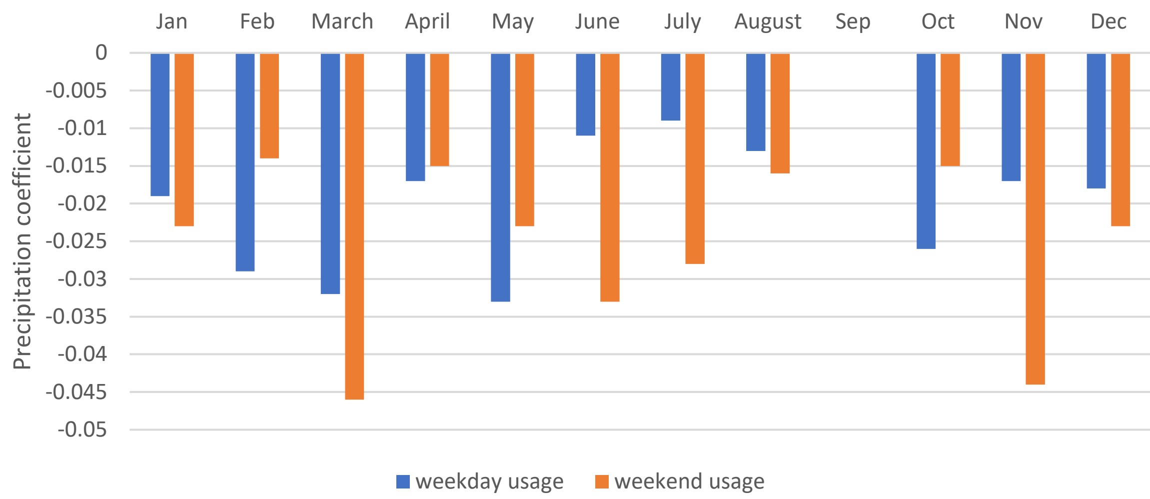


Figure 3.21 Impact of precipitation on cycling infrastructure usage on weekdays and weekends

Also, wind speed and cumulated snow on the ground showed a negative impact on cycling infrastructure usage. Figure 3.22 and Figure 3.23 show the effect of these variables on infrastructure usage.

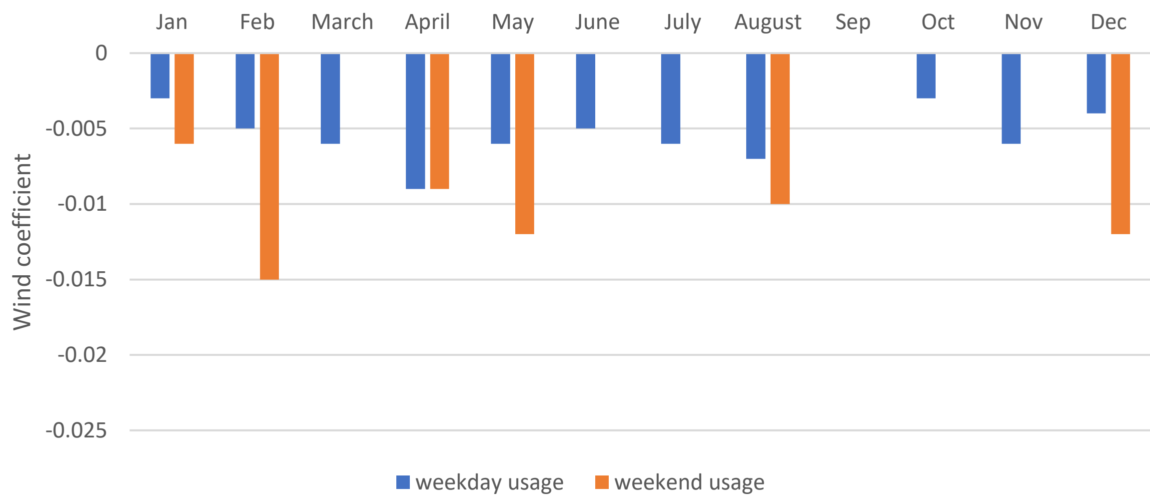


Figure 3.22 Impact of wind on cycling infrastructure usage on weekdays and weekends

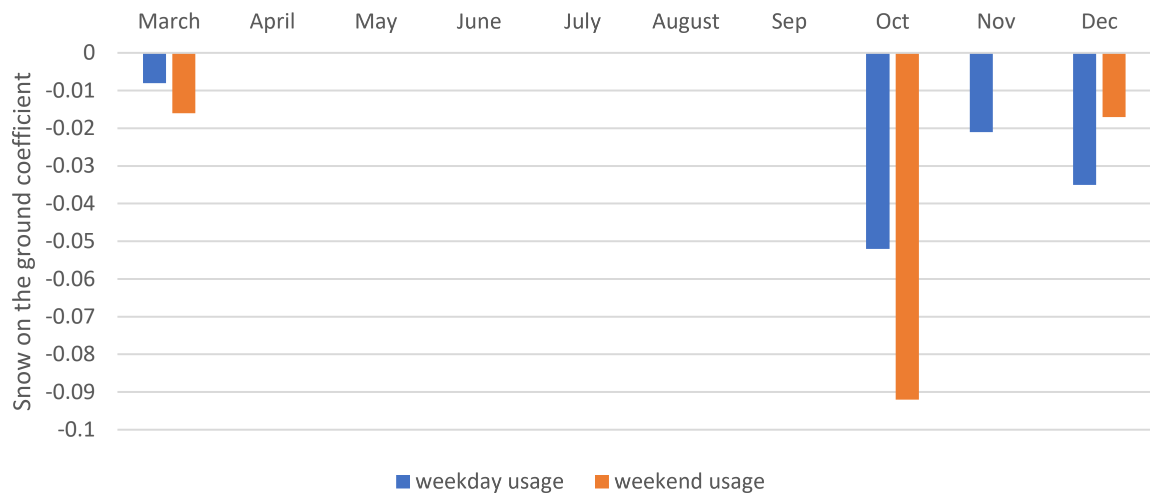


Figure 3.23 Impact of snow on the ground on cycling infrastructure usage on weekdays and weekends

Sunlight hours showed a positive correlation with cycling infrastructure usage. However, the correlation was significant for only a few months of the year. Moreover, as shown in Figure 3.20 to Figure 3.23, the results indicated that the magnitude of the impact of these variables was different for different months of the year.

CHAPTER 4: VALIDATION

This chapter first describes the procedure for selecting the proper model to estimate the weekday and weekend usage. After selecting the model with the best goodness of fit, the model was validated using the portion of data that was kept for this purpose.

4.1 MODEL APPLICATION

As mentioned in the previous chapter, two separate models were developed. The first model was developed based on weekday usage, and the second model was developed based on weekend usage. Both models were applied to determine which model yielded a better estimation result for weekday and weekend usage.

4.1.1 Weekend usage estimation

The first step was to estimate weekend usage. Both models were applied to estimate infrastructure usage during weekends. The usage for each month of the year was estimated using September usage and weather variables (β_1 to β_5). The mean absolute percentage error (MAPE) between the actual and estimated usage was calculated to evaluate the model accuracy. The MAPE is calculated as follows:

$$\text{MAPE:} \frac{1}{N}\sum_{i=1}^{N}\left|\frac{ACF-ECF}{ACF}\right| \tag{4.1}$$

Where *ACF* represents actual cycling frequency, *ECF* represents Estimated cycling frequency and *N* represents the number of observations used for validation, which in this case is the number of months for which the infrastructure sections were observed. Months with an average usage of below 100 cyclists/month were excluded from the model. The small difference between the estimated and actual usage in those months led to a large MAPE value because of the small actual usage. This result increased the total MAPE value unrealistically. For example, for an infrastructure section with actual usage of five cyclists in month and the estimated usage of nine cyclists per month, the difference in model estimation and actual usage is only four cyclists while the estimation error is 80%. Infrastructure usage during weekends was estimated using both models. The model results were compared with the actual usage, and the MAPE value for each model was calculated. Table 4.1 presents the MAPE values of the two models for estimating the weekend usage of different sections of infrastructure.

Table 4.1 MAPE values of estimating weekend usage

Infrastructure section (assigned number)	First model MAPE value	Second model MAPE value
Peace Bridge (1)	0.50	0.26
5 St and 10 Ave SW (2)	0.44	0.20
Memorial Drive and 19 St NW (3)	0.53	0.35
River Walk (4)	0.62	0.33
Bow River Pathway (5)	0.44	0.21
7 St and 3 Ave SW (9)	0.37	0.15
8 Ave and 3 St SW (7)	0.47	0.18
5 St and 5 Ave SW (6)	0.51	0.19
Stephen Avenue (8)	0.33	0.09
12 Ave and 2 St SW (12)	0.41	0.23
Memorial Drive at Prince’s Island (14)	0.56	0.32
Nose Creek at Bow River (17)	0.55	0.33
8 Ave and 8 St SW (18)	0.50	0.25
12 Ave and 3 St SE (16)	0.57	0.23
Lindsay Park (20)	0.50	0.20
9 Ave and 4 St SE (21)	0.67	0.30

As shown in Table 4.1, the second model showed a lower MAPE value compared to the value for the first model for estimating the usage on weekends for all infrastructure sections. Hence, the second model is used to estimate infrastructure usage on weekends.

4.1.2 Weekday usage estimation

The second step was to estimate weekday usage. Both models were applied to infrastructure sections. Usage on weekdays was estimated and compared with the actual usage on weekdays to evaluate the results of the models. MAPE values for both models were calculated. Months with an average usage below 100 cyclists/month were excluded for the same reason they were excluded from the weekend estimation. Table 4.2 shows the results of the two models for estimating weekday usage.

Table 4.2 MAPE values of estimating weekday usage

Infrastructure section (assigned number)	First model MAPE value	Second model MAPE value
Peace Bridge (1)	0.21	0.15
5 St and 10 Ave SW (2)	0.19	0.14
Memorial Drive and 19 St NW (3)	0.29	0.19
River Walk (4)	0.31	0.23
Bow River Pathway (5)	0.29	0.20
7 St and 3 Ave SW (9)	0.16	0.19
8 Ave and 3 St SW (7)	0.16	0.15
5 St and 5 Ave SW (6)	0.21	0.17
Stephen Avenue (8)	0.18	0.14
12 Ave and 2 St SW (12)	0.17	0.14
Memorial Drive at Prince's Island (14)	0.52	0.40
Nose Creek at Bow River (17)	0.45	0.31
8 Ave and 8 St SW (18)	0.14	0.15
12 Ave and 3 St SE (16)	0.38	0.31
Lindsay Park (20)	0.60	0.49
9 Ave and 4 St SE (21)	0.22	0.16

As shown in the results, for some of infrastructure sections, such as at 8 Ave and 8 St SW and 7 St and 3 Ave SW, the first model had a better estimation result with a lower MAPE value compared to the second model, while for other infrastructure sections, the second model fit better. The reason that different models fit better to estimate weekdays usage for different infrastructure sections is assumed to be due to difference in types of users that use these infrastructure sections. The first model fits better for infrastructure sections that located in business areas mainly used by commuter cyclists while for infrastructure sections that located at recreational areas, the second model fits better.

To find more about the type of cyclist using an infrastructure section, cycling infrastructure sections were classified. As mentioned, some of the infrastructure sections were mainly used by commuter cyclists during a year, while others were mainly used by utilitarian cyclists. Further, each infrastructure section had different types of users during weekdays and weekends. Weekend cyclists were mainly utilitarian cyclists, while weekday cyclists were a combination of commuter cyclists and utilitarian cyclists. To discover more about the characteristics of infrastructure users, a user type ratio (UTR) is defined as follows:

$$\text{UTR}= \frac{CW}{CT} \tag{4.2}$$

Where *CW* represents average number of cyclists on weekends and *CT* represents total number of cyclists on weekdays and weekends. The UTR for each infrastructure section shows the ratio of the average usage on weekends compared to the total usage on both weekdays and weekends. The user type ratio for 2016 was calculated for all the infrastructure sections analyzed in this study. The details about infrastructure sections UTR are shown in Table 4.3.

Table 4.3 Infrastructure user type ratio for 2016

Infrastructure section (assigned number)	User type ratio (UTR)
Peace Bridge (1)	0.2048
5 St and 10 Ave SW (2)	0.1395
Memorial Drive and 19 St NW (3)	0.2049
River Walk (4)	0.2758
Bow River Pathway (5)	0.1672
7 St and 3 Ave SW (9)	0.1021
8 Ave and 3 St SW (7)	0.1136
5 St and 5 Ave SW (6)	0.1399
Stephen Avenue (8)	0.1876
12 Ave and 2 St SW (12)	0.1591
Memorial Drive at Prince's Island (14)	0.3085
Nose Creek at Bow River (17)	0.2592
8 Ave and 8 St SW (18)	0.0916
12 Ave and 3 St SE (16)	0.2028
Lindsay Park (20)	0.3290
9 Ave and 4 St SE (21)	0.1527

A high UTR for an infrastructure section represents high usage on weekends compared to total usage. Usage on weekends is mainly attributed to utilitarian cyclists. Hence, infrastructure sections with a high UTR contain a high number of utilitarian cyclists among their users and are mainly used for utilitarian purposes. Table 4.3 shows the UTR for the selected infrastructure sections. As shown in Table 4.3, infrastructure sections, such as Lindsay Park and Memorial Drive at Prince's Island Park, have high UTRs; these two infrastructure sections, which are identified with red boxes in Figure 4.1, are located along the Bow River or in parks. In contrast, infrastructure sections such as 8 Ave and 8 St and 7 St and 3 Ave SW, which have low UTRs, are usually located far away from recreational places. These infrastructure sections are identified with blue boxes in Figure 4.1.

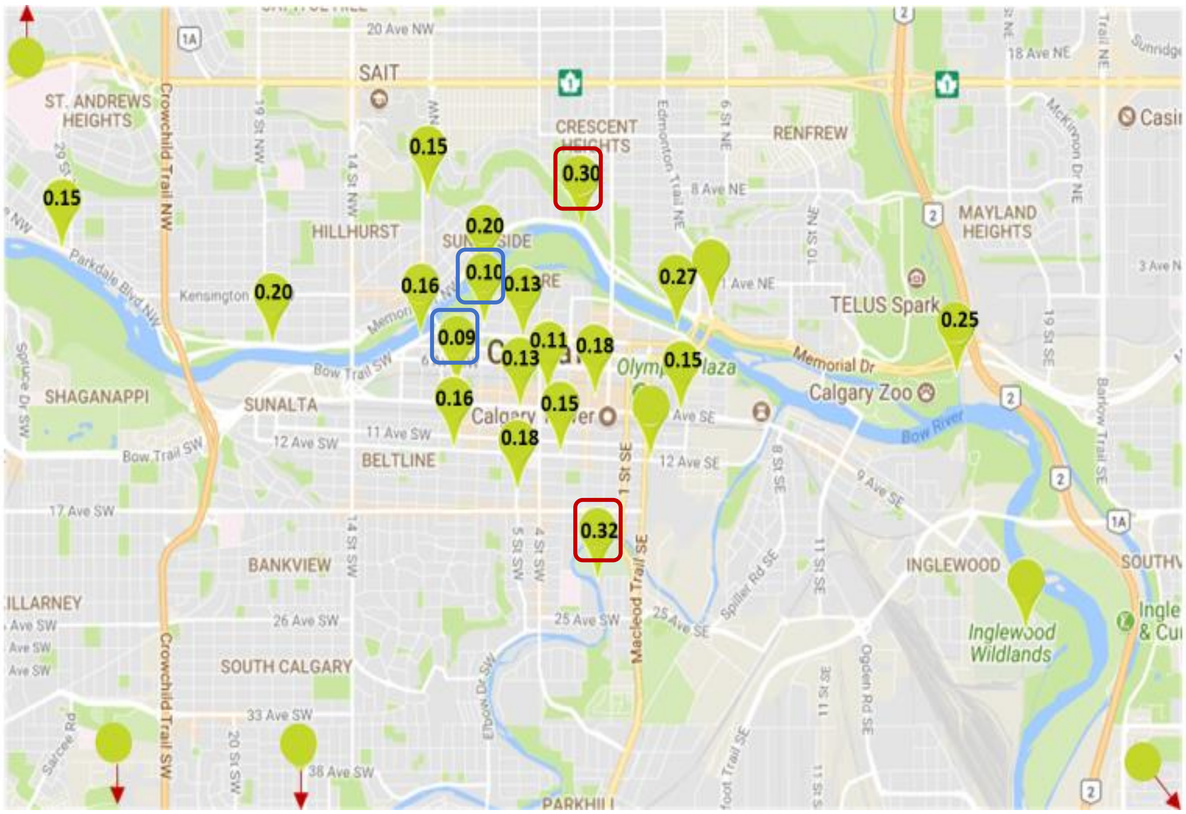


Figure 4.1 User type ratio for cycling infrastructure sections based on their location

Also, looking at hourly trends of infrastructure sections usage, we noticed that usage trends of different infrastructure sections were not the same. Infrastructure sections with mainly commuter users (low UTR) have higher peak usage in the morning and afternoon and lower usage at midday compared to infrastructure sections with utilitarian users (high UTR). Figure 4.2 shows hourly usage trends of four infrastructure sections on weekdays. In contrast, on weekends, infrastructure sections with utilitarian users have higher midday usage as shown in Figure 4.3.

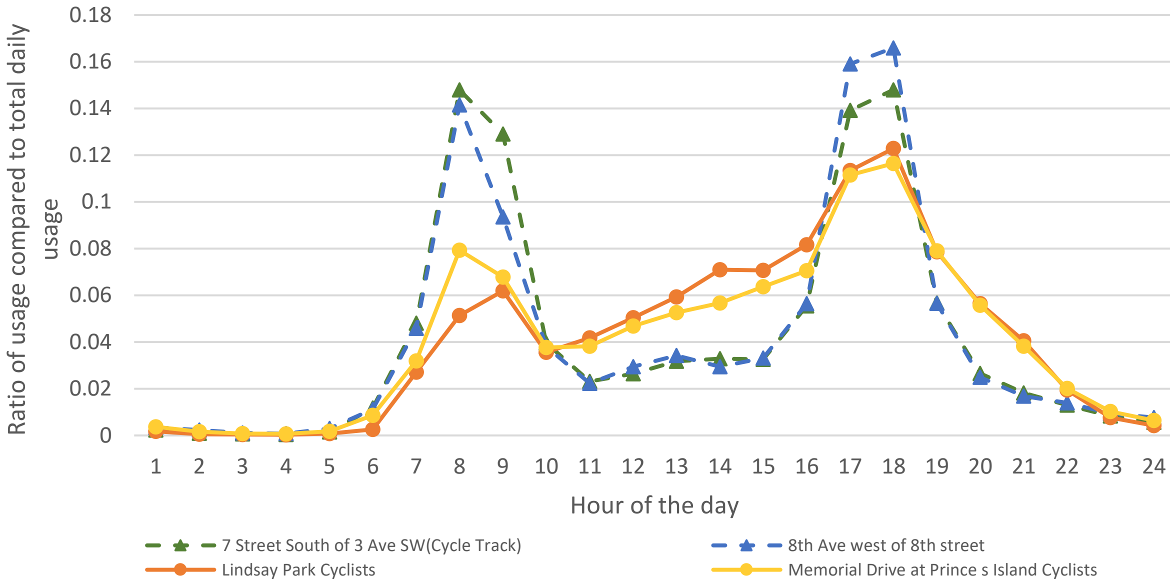


Figure 4.2 Hourly usage patterns for utilitarian and commuter usage on weekdays

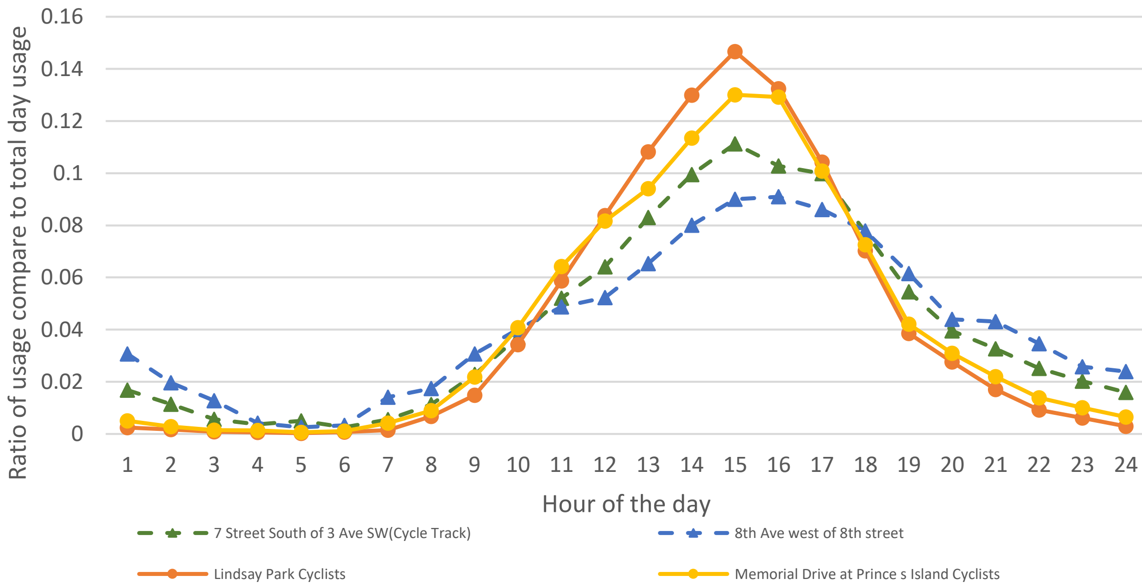


Figure 4.3 Hourly usage patterns for utilitarian and commuter usage on weekends

For the next step, we sorted the infrastructure sections based on the user type ratio (UTR) value. The UTR values were calculated based on infrastructure usage in 2016. Table 4.4 shows the

UTR and MAPE values for infrastructure sections. As shown in Table 4.4, for infrastructure sections with low UTR values, the first model had the best goodness of fit, while for infrastructure sections with high UTR values, the second model showed the best goodness of fit. Further, for infrastructure sections with a UTR of more than 0.11, the second model gave a better fit compared to the first model, and for infrastructure sections with a UTR lower than 0.11, the second model yielded a better result for estimating weekday usage.

Table 4.4 Infrastructure sections UTR and MAPE values

Infrastructure section (assigned number)	UTR	Weekday model MAPE value	Weekend model MAPE value
8 Ave and 8 St SW (18)	0.0916	0.14	0.15
7 St and 3 Ave SW (9)	0.1021	0.16	0.19
8 Ave and 3 St SW (7)	0.1136	0.16	0.15
5 St at 10 Ave SW (2)	0.1395	0.19	0.14
5 St and 5 Ave SW (6)	0.1399	0.21	0.17
9 Ave and 4 St SE (21)	0.1527	0.21	0.16
12 Ave and 2 St SW (12)	0.1591	0.29	0.2
Bow River Pathway (5)	0.1672	0.27	0.23
Stephen Avenue (8)	0.1876	0.21	0.15
12 Ave and 3 St SE (16)	0.2028	0.52	0.40
Peace Bridge (1)	0.2048	0.45	0.31
Memorial Drive and 19 St (3)	0.2049	0.31	0.23
Nose Creek at Bow River (17)	0.2592	0.52	0.40
River Walk (4)	0.2758	0.6	0.49
Memorial Drive Prince's Island (14)	0.3085	0.52	0.40
Lindsay Park (20)	0.3290	0.60	0.49

Since the first model was developed using weekday usage and cyclists who cycle on weekdays are mainly commuter cyclists combined with utilitarian cyclists we named the first model commuter model. The second model was developed based on weekend usage and in general, cyclists who cycle during weekends have utilitarian purposes. They cycle for purposes such as shopping, going to restaurants, or for exercise, all of which fall into the category of utilitarian purposes. Hence, we named the second model utilitarian model.

The results showed that the second model fits better for estimating weekend usage and weekday usage for infrastructure sections with mainly utilitarian users (high UTR) and the first model fits better for estimating weekday usage for infrastructure sections with mainly commuter users (low UTR). Usage can be categorized based on type of users. Commuter usage and utilitarian usage. Two new models were developed based on the new categorize of usage. The first model was developed based on weekdays usage of infrastructure sections with mainly commuter users (commuter usage) and the second model was developed based on weekday usage of infrastructure sections with mainly utilitarian cyclists and weekend usage (utilitarian usage). The two new models were applied to estimate the usage of infrastructure sections and their result was compared with the results of the previous two models to find which one of them estimates the usage more accurately. Table 4.5 and Table 4.6 show the results of the new developed models and the previous developed models for estimating weekend and weekday usage for infrastructure sections.

Table 4.5 Comparison of the estimation of developed models for estimating weekday usage

	New models		Previous models	
Infrastructure section (assigned number)	First model MAPE value	Second model MAPE value	First model MAPE value	Second model MAPE value
8 Ave and 8 St SW (18)	0.24	0.40	0.14	0.15
7 St and 3 Ave SW (9)	0.26	0.37	0.16	0.19
8 Ave and 3 St SW (7)	0.28	0.38	0.16	0.15
5 St at 10 Ave SW (2)	0.37	0.31	0.19	0.14
5 St and 5 Ave SW (6)	0.46	0.31	0.21	0.17
9 Ave and 4 St SE (21)	0.34	0.29	0.21	0.16
12 Ave and 2 St SW (12)	0.36	0.30	0.29	0.20
Bow River Pathway (5)	0.40	0.53	0.27	0.23
Stephen Avenue (8)	0.25	0.44	0.21	0.15
12 Ave and 3 St SE (16)	0.71	0.31	0.52	0.40
Peace Bridge (1)	0.38	0.25	0.45	0.31
Memorial Drive and 19 St (3)	0.46	0.21	0.31	0.23
Nose Creek at Bow River (17)	0.70	0.24	0.52	0.40
River Walk (4)	0.56	0.19	0.6	0.49
Memorial Drive Prince’s Island (14)	0.86	0.24	0.52	0.40
Lindsay Park (20)	1.01	0.31	0.60	0.49

Table 4.6 Comparison of the estimation of developed models for estimating weekend usage

	New models		Previous models	
Infrastructure section (assigned number)	First model MAPE value	Second model MAPE value	First model MAPE value	Second model MAPE value
Peace Bridge (1)	1.47	0.40	0.50	0.26
5 St and 10 Ave SW (2)	0.92	0.25	0.44	0.20
Memorial Drive and 19 St NW (3)	2.02	0.55	0.53	0.35
River Walk (4)	2.38	0.69	0.62	0.33
Bow River Pathway (5)	1.43	0.40	0.44	0.21
7 St and 3 Ave SW (9)	0.91	0.25	0.37	0.15
8 Ave and 3 St SW (7)	0.74	0.31	0.47	0.18
5 St and 5 Ave SW (6)	1.60	0.42	0.51	0.19
Stephen Avenue (8)	0.40	0.31	0.33	0.09
12 Ave and 2 St SW (12)	0.93	0.25	0.41	0.23
Memorial Drive at Prince's Island (14)	2.19	0.61	0.56	0.32
Nose Creek at Bow River (17)	1.99	0.57	0.55	0.33
8 Ave and 8 St SW (18)	0.64	0.25	0.50	0.25
12 Ave and 3 St SE (16)	1.45	0.47	0.57	0.23
Lindsay Park (20)	3.12	0.96	0.50	0.20
9 Ave and 4 St SE (21)	0.29	0.34	0.22	0.16

As it can be seen from the two tables, although the newly developed models had a better estimation accuracy for some infrastructure sections, in general, the previous models had a better estimation results compared to the newly developed models. The previous models estimated the weekday and weekend usage with average MAPE value of 0.24 and 0.23 respectively, while the new models estimated weekday and weekend usage with average MAPE value of 0.27 and 0.44 respectively.

UTR values were calculated based on the total year usage. Hence, for infrastructure sections that were only observed during a limited time period (i.e., 1 month of the year), these UTR values were not available. Since this study aims to estimate usage using one-month observation (September), we calculated the UTR values for different infrastructure sections based on the September usage. Table 4.7 shows the UTR values for the infrastructure sections for the whole year and for September. As shown, the UTR values for September are different compared to the UTR values of the full year. However, the threshold UTR value for selecting the proper model is still 0.11 when UTR values from September are used.

Table 4.7 Infrastructure sections UTR values for the whole year and for September

Infrastructure section (assigned number)	UTR Year	UTR September
8 Ave and 8 St SW (18)	0.0916	0.0901
7 St and 3 Ave SW (9)	0.1021	0.1053
8 Ave and 3 St SW (7)	0.1136	0.1155
5 St at 10 Ave SW (2)	0.1395	0.1214
5 St and 5 Ave SW (6)	0.1399	0.1340
9 Ave and 4 St SE (21)	0.1527	0.1617
12 Ave and 2 St SW (12)	0.1591	0.1343
Bow River Pathway (5)	0.1672	0.1459
Stephen Avenue (8)	0.1876	0.2034
12 Ave and 3 St SE (16)	0.2028	0.2061
Peace Bridge (1)	0.2048	0.1986
Memorial Drive and 19 St (3)	0.2049	0.1943
Nose Creek at Bow River (17)	0.2592	0.2507
River Walk (4)	0.2758	0.2798
Memorial Drive Prince’s Island (14)	0.3085	0.2971
Lindsay Park (20)	0.3290	0.3264

4.1.3 Summary

The cycling infrastructure total usage during a year was separated into weekday and weekend usage. Two separate models were developed. The commuter model was developed based on weekday usage, and the utilitarian model was developed based on weekend usage. The utilitarian model had better estimation results compared to the commuter model to estimate usage during the weekend. Further, the commuter model was better at estimating usage during weekdays for infrastructure sections with a UTR of less than α (011), but the utilitarian model was better for infrastructure sections with a UTR more than α. The α value varies based on the length and time of the available observation. For observation during September and full year observation the α value is equal to 0.11. Figure 4.4 shows the flowchart for estimating cycling infrastructure usage during weekdays and weekends.

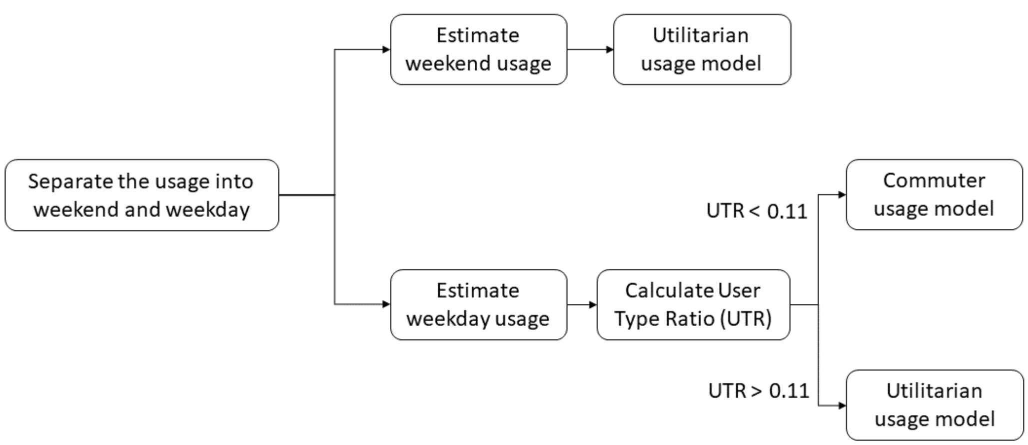


Figure 4.4 Flowchart for estimating infrastructure usage throughout the year

4.2 MODEL VALIDATION

In this section, the estimated model was validated using two groups of infrastructure sections. The first group of infrastructure sections was located near the downtown core area of the city, while the second group was located outside of downtown area. The model was validated for the year 2017 for both groups to determine if the estimation model was capable of estimating the usage for infrastructure sections. Since the model was developed based on data from downtown core area, these two groups of infrastructure sections were selected to find if the model is capable of estimating the usage for any location in the city and to determine that it is not limited to the downtown core area.

4.2.1 Validation for the downtown core area

The model was used to estimate the usage of four known pieces of infrastructure based on September usage. The results were compared with infrastructure actual usage, which was counted by bicycle counters during the year. Four pieces of infrastructure with different locations and characteristics were selected for the model validation to determine the model can estimate the usage for a variety of infrastructure sections. Table 4.8 shows an overview of the characteristics of the selected infrastructure pieces. The four road characteristics are as follows:

1. Parkdale is a cycling pathway located alongside the Bow River in a recreational neighborhood and outside the downtown area.
2. 10 St and 5 Ave SW is a cycling road separated by a lane from the road that vehicles drive on, located outside the downtown area in a residential location.
3. 5 St and 15 Ave SW is a physically separated path for cyclists located in south part of the downtown area.

4. 12 Ave and 8 St SW is a physically separated pathway located in the east part of the downtown area.

Table 4.8 Overview of infrastructure characteristics

Infrastructure section	Type of infrastructure	Location	UTR (year)	Monthly Av. usage (2016)	Neighborhood
Parkdale	Off-road pathways	Outside downtown	0.159227	126870	Recreational
10 St and 5 Ave NW	On-road Sep. Lane	Outside downtown	0.1529	10947	Residential
5 St and 15 Ave SW	On-road Sep. physically	Downtown	0.181443	15491	Commercial
12 Ave and 8 St SW	On-road Sep. physically	Downtown	0.167634	14346	Commercial

Figure 4.5 shows the location and the assigned number (Figure 3.2) of the selected infrastructure sections used for validation. Also, pictures of the infrastructure sections are shown in Figure 4.6.

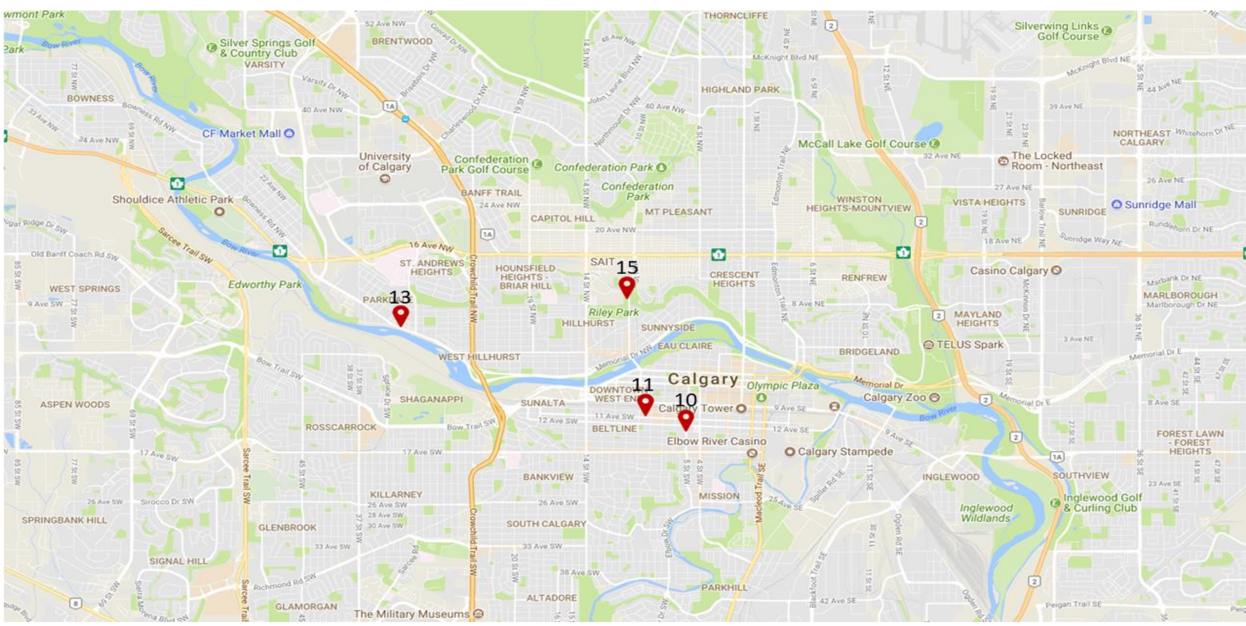


Figure 4.5 Locations of the selected infrastructure sections

a) Parkdale

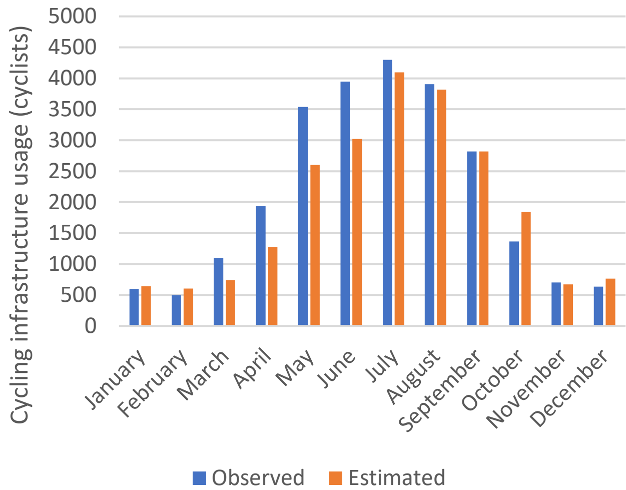
b) 10 St and 5 Ave NW

c) 5 St and 15 Ave SW

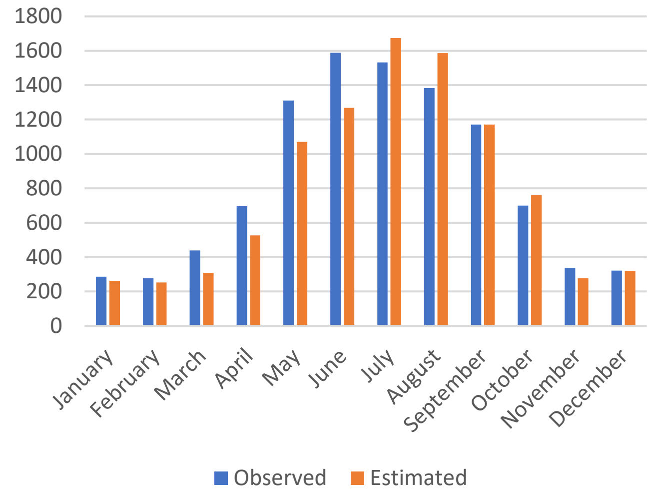
d) 12 Ave and 8 St SW

Figure 4.6 Infrastructure sections pictures

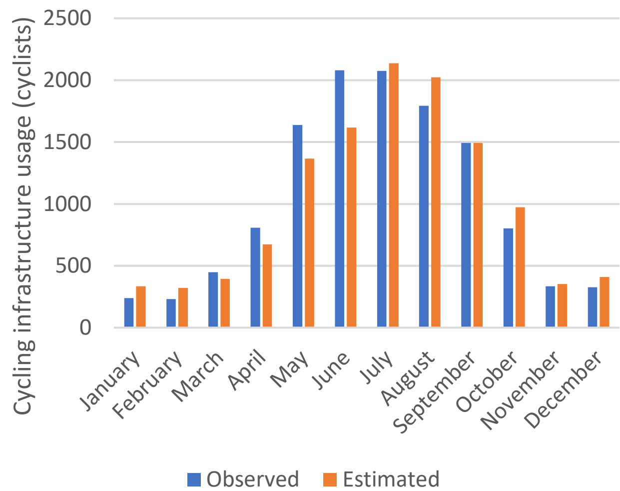
Infrastructure usage on weekends and weekdays was estimated separately based on September weekend and weekday usage. Since the UTR for all selected infrastructure sections was higher than 0.11, the utilitrian model was used to estimate both weekend and weekday usage. The total monthly usage was calculated by adding the weekend and weekday usage together. Estimated and observed usages are shown in Table 4.4.



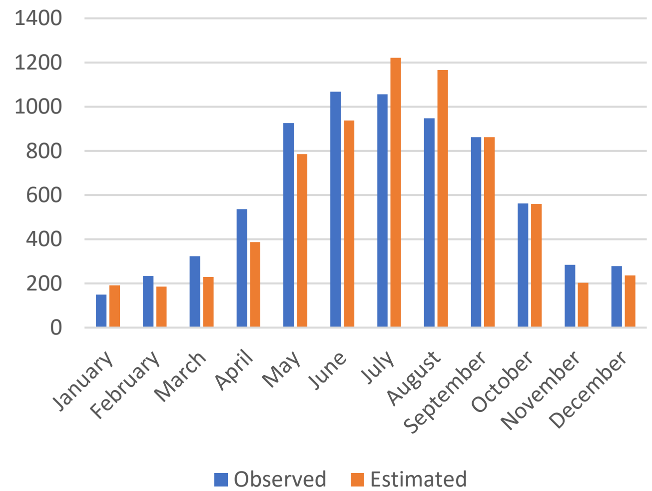
a) Parkdale



b) 12 Ave and 8 St SW



c) 5 St and 15 Ave



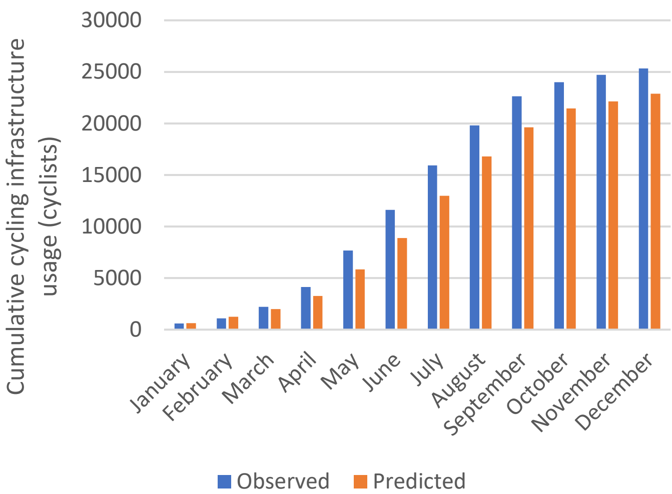
d) 10 St and 5 Ave

Figure 4.7 Comparison of the estimation model results and actual cycling frequency

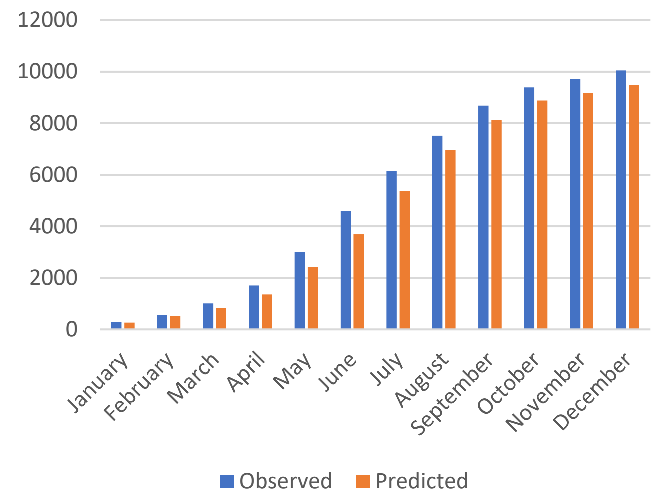
The mean absolute percentage error between the actual and estimated values for cycling frequency was calculated to evaluate the model accuracy for estimating annual usage. Table 4.9 shows the absolute error for each month of the year and the mean absolute error for all the months for each section of infrastructure. In addition, to determining how accurately the model estimates annual usage, we calculated the cumulative usage during the year. Figure 4.8 shows the cumulative usage of the four selected cycling infrastructure sections.

Table 4.9 Model absolute percentage error

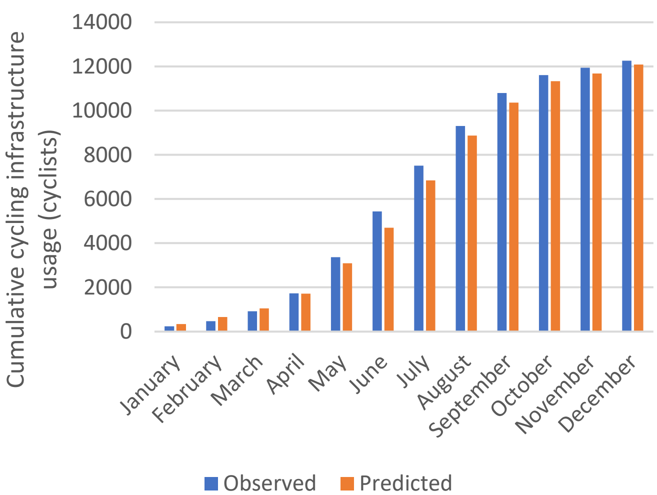
Month	12 Ave and 8 St SW	5 St and 15 Ave SW	Parkdale	10 St and 5 Ave NW
January	0.083427	0.398688	0.061798	0.27814
February	0.091651	0.399152	0.225528	0.201269
March	0.295396	0.121435	0.331794	0.292022
April	0.243193	0.16691	0.341594	0.277919
May	0.182814	0.165659	0.265864	0.151757
June	0.20119	0.222252	0.23407	0.12185
July	0.092577	0.030657	0.045967	0.157206
August	0.146033	0.127318	0.022352	0.230135
September	-	-	-	-
October	0.088748	0.213372	0.351066	0.005154
November	0.178162	0.055123	0.042545	0.288086
December	0.001293	0.252119	0.200311	0.147098
Mean	0.145	0.195	0.192	0.195



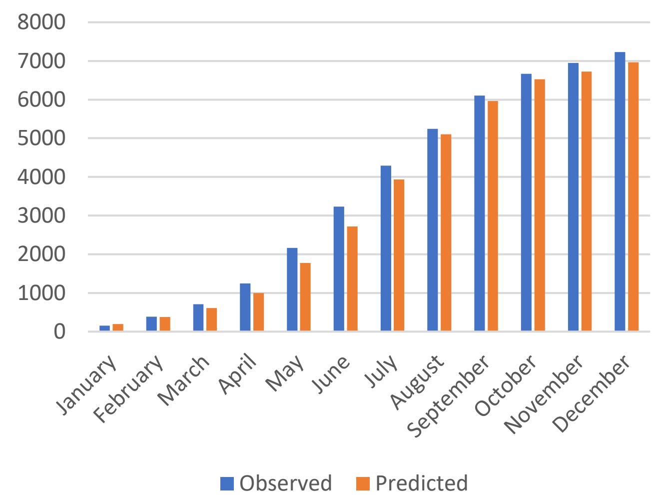
a) Parkdale



b) 12 Ave and 8 St SW



c) 5 St and 15 Ave SW



d) 10 St and 5 Ave NW

Figure 4.8 Comparison of the cumulative estimation model results and actual cycling frequency

The annual usage for each section of infrastructure was calculated by adding up the usage numbers of different months, which were estimated based on the September usage. The difference between the annual estimated usage and actual usage was calculated as follows:

$$Absolute\ Error = \left|\frac{ACF-ECF}{ACF}\right| \tag{4.3}$$

Where *ACF* represents actual cycling frequency, *ECF* represents Estimated cycling frequency. Table 4.10 shows the results of the model for estimating annual usage.

Table 4.10 Estimated and observed annual usages of infrastructure sections

	12 Ave and 8 St SW	5 St and 15 Ave SW	Parkdale	10 St and 5 Ave NW
Actual annual usage	10043.89	12265.9	25344.54	7226.682
Estimated annual usage	9481.938	12094.36	22894.21	6965.951
Estimation error	0.055	0.013	0.096	0.036

4.2.2 Validation for outside of downtown area

The model was validated for the second group of infrastructure sections. These infrastructure sections were located outside of downtown area in the city of Calgary. Data for these infrastructure sections were available only for limited period of the year. Figure 4.9 shows the locations of the selected infrastructure sections. The four selected infrastructure sections and their assigned numbers are as follows:

27. Bowmont Bridge is a cycling pathway located in the northwest part of the city. The pathway is located in a recreational neighborhood.

28. Shouldice Bridge is a cycling pathway located in the northwest part of the city. The pathway is located in a recreational neighborhood.

29. Max Bell Arena is a cycling pathway is located to the east of the core area. The pathway starts from a parking lot and goes into the downtown area.

30. 32 Ave and Campus Dr NW is a street without any cycling facilities located on the north side of the University of Calgary.

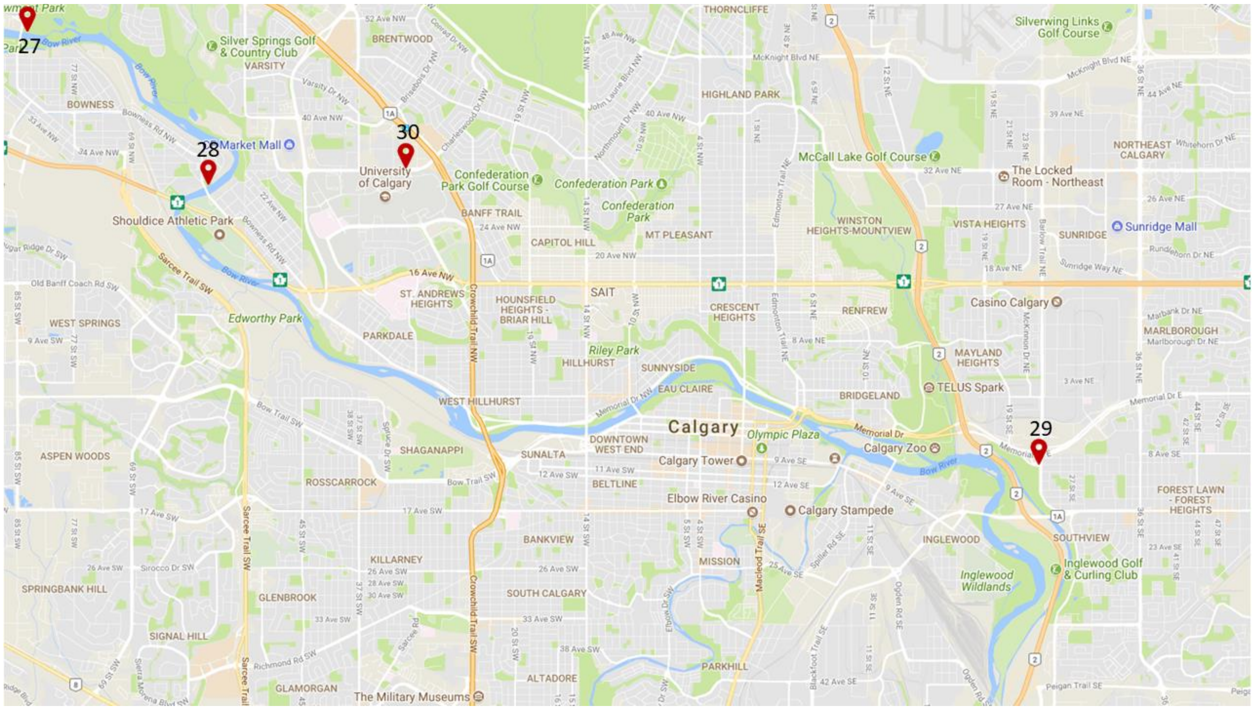


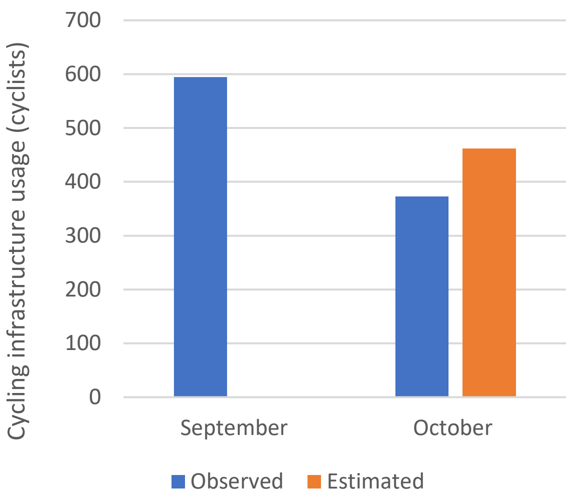
Figure 4.9 Locations of the four infrastructure sections with partial data

Portable counters were installed in September 2017 for the first three infrastructure sections. They counted daily usage in September and October 2017. Due to the counters' limitations, they were removed in early November because they were not able to count bicycles if there was snow on the ground. Observations were conducted during September, October, November, and December for the fourth road. Observations were conducted during peak hours for one week in each month.

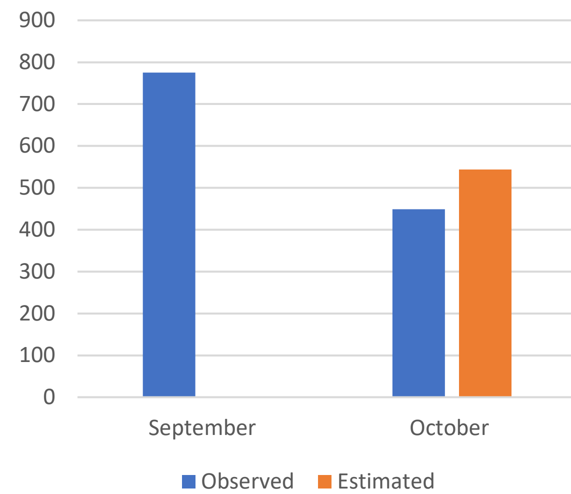
For the first three roads, the utilitarian model was used to estimate the weekend and weekday usage. The UTR ratios for these roads were more than 0.11. However, for the fourth road, the utilitarian model was used to estimate usage on weekends, and the commuter model was used to estimate usage on weekdays because the observations were conducted during peak hours of day, meaning, in general, the cyclists were commuter cyclists. Figure 4. 10 shows the estimated and observed usage and shows the absolute percentage error for the four selected infrastructure sections.

Table 4.11 Model absolute percentage error for roads with partial data

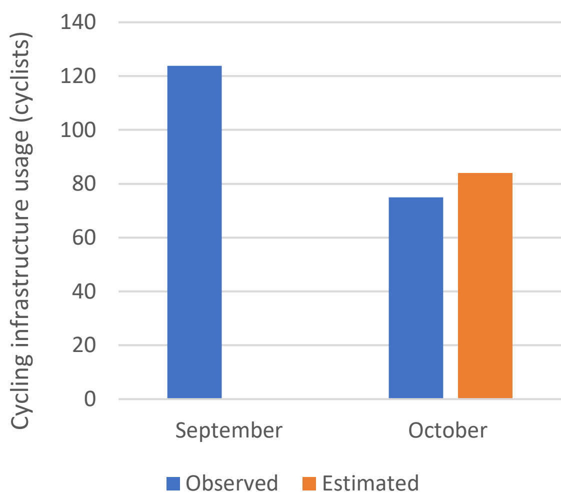
Infrastructure location		Observed usage	Estimated usage	Absolute error
Bowmont Bridge	Sep.	594.33	_	_
	Oct.	373.03	641.58	0.237
Shouldice Bridge	Sep.	775.25	_	_
	Oct.	448.90	543.79	0.211
Max Bell Arena	Sep.	123.88	_	_
	Oct.	74.92	83.86	0.119
32 Ave and Campus Dr NW	Sep.	148.20	_	_
	Oct.	137.10	143.88	0.049
	Nov.	56.80	60.98	0.073
	Dec.	94.20	117.82	0.250



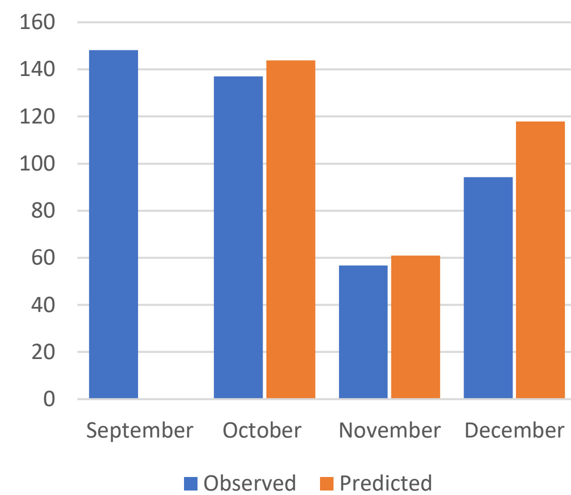
a) Bowmont Bridge



b) Shouldice Bridge



c) Max Bell Arena



d) 32 Ave and Campus Dr NW

Figure 4. 10 Comparison of the estimation model results and actual cycling frequency for roads with partial data

4.2.3 Summary

Two models were developed based on different types of usage. Both models were applied to each infrastructure section to verify which model gave a better result for estimating usage during weekends and weekdays. To estimate weekend usage, the second model yielded a better estimation result. To predict weekday usage, for infrastructure sections with mainly commuter cyclists (UTR<0.11), the first model yielded a better estimation result while for infrastructure section with mainly utilitarian cyclists (UTR>0.11), the second model had a better estimation.

The model was validated with two groups of infrastructure sections across the city of Calgary. The validation results showed that the model was capable of estimating cycling infrastructure usage throughout the year across the city with an average error of 17%. Also, infrastructure annual usage was estimated using the models with an average error of 5%.

4.3 ESTIMATION TOOL

A tool was created in Microsoft Excel that would allow users to use the results of this study to estimate cycling infrastructure usage. The tool requires the average usage observed during weekdays and weekends, the month of the year that the usage was observed and weather data as input data. Users can also choose the type of model that the tool uses to estimate the usage. Three options are available (i.e. utilitarian model, commuter model, auto selection model). The auto selection model chooses the proper model to estimate the weekday and weekend usage based on UTR value. The utilitarian option estimates both weekday and weekend usage with the utilitarian model and commuter option estimates weekday usage with commuter model and weekend usage with utilitarian model.

The weather data can be entered manually, or the user can select from the existing weather data. Weather data for the City of Calgary for the years 2013, 2014, 2015, 2016 and 2017 are available within the tool. The tool estimates the usage of the other months of the year using the observed usage and the two models developed in this study.

Figure 4.11 shows the working area of the tool. Users have to enter the average usage of weekday and weekend (Box C and D) and the month with observed usage (Box A). Users also have to choose between existing weather data or enter their own weather variables (Box B) and choose a model for usage estimation (Box E). After entering the input data, the user can run the tool to estimate the usage.

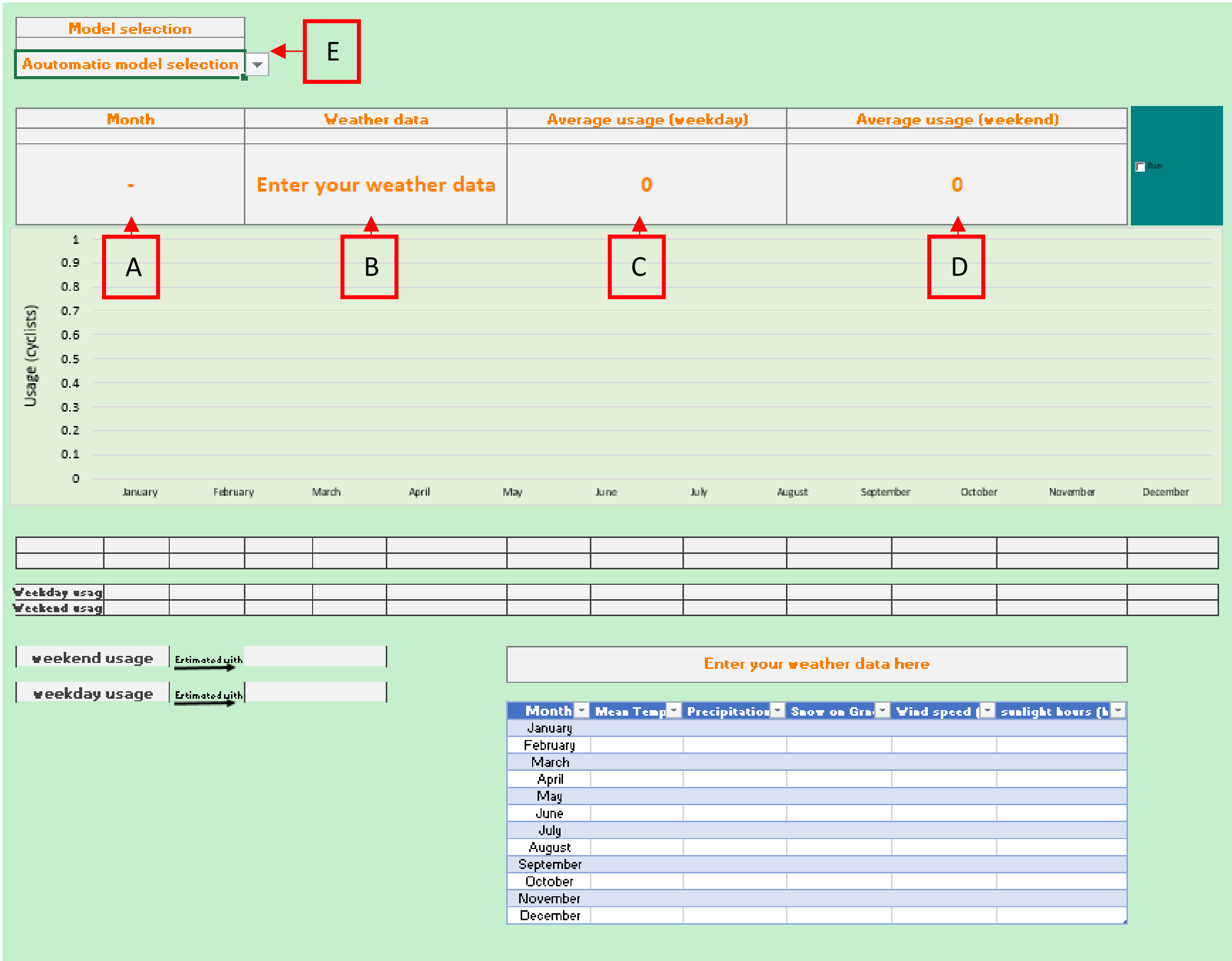


Figure 4.11 Working area of the usage estimation tool

Figure 4.12 shows the analysis section of the tool for estimating the usage. UTR values are calculated based on weekday and weekend average usage (shown within the red box in the Figure). Average weekend usage for different months of the year is estimated using the utilitarian model and weekend observed usage. If the user selects the auto selection model, weekday usage of different months of the year is estimated based on UTR value. If the user selects the commuter model, weekday usage is estimated with the commuter model and if the user selects the utilitarian model, weekday usage is estimated based on the utilitarian model. The weekend usage is estimated with the utilitarian model in both cases.

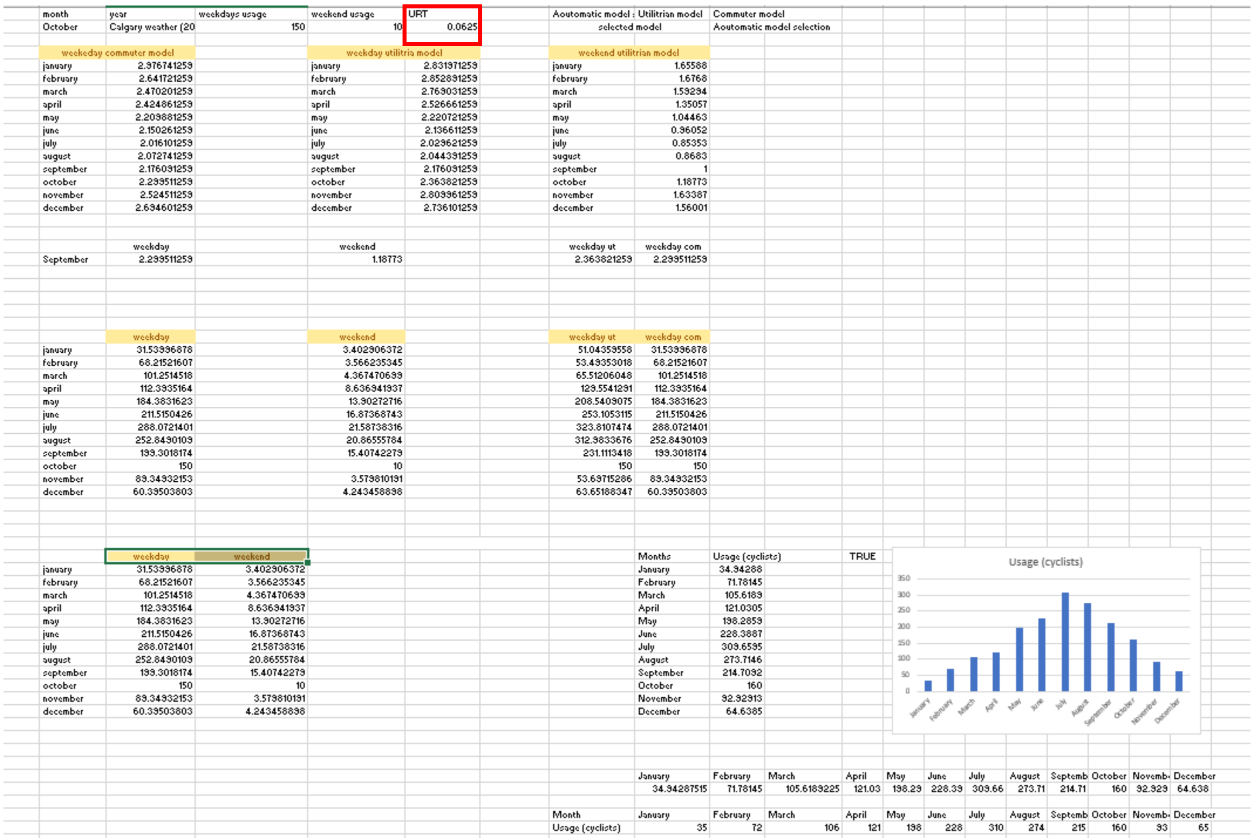


Figure 4.12 Analysis section of the usage estimation tool

Figure 4.13 shows the estimated usage for different months of the year based on observations in October. The automatic model selection is chosen which means that models will be selected based on UTR value. The red box in the figure shows which models were used to estimate the weekend and weekday usage. The tool estimates the weekend and weekday usage for each month separately. The total average usage for each month is calculated by adding the average weekday and weekend usages.

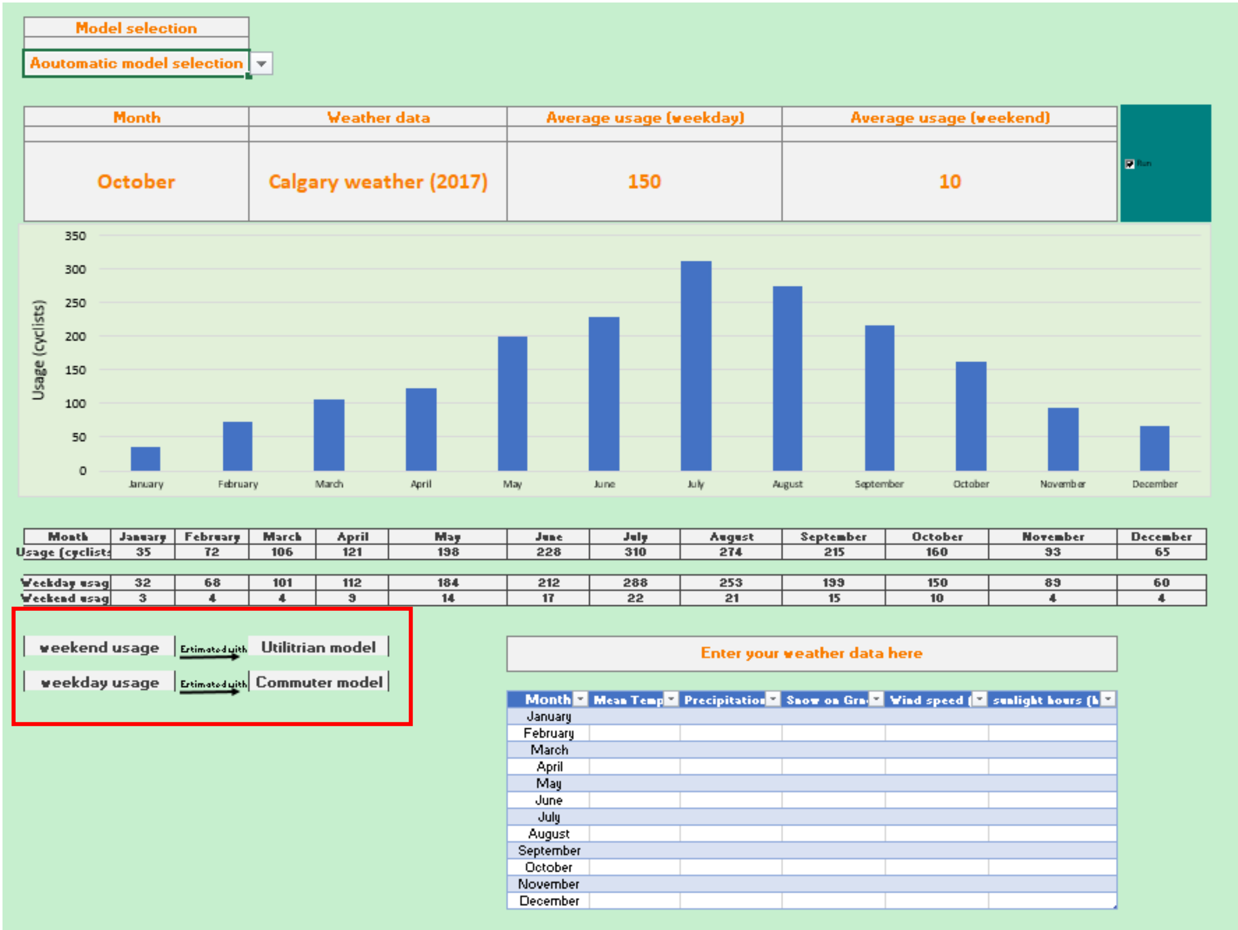


Figure 4.13 Output of the usage estimation tool

CHAPTER 5: CONCLUSION

5.1 SUMMARY

Cycling has gained increasing attention among both researchers and municipalities in the last few decades as a sustainable alternative to car-based transportation. To encourage cycling, municipalities provide better conditions for cyclists by improving cycling infrastructure networks with using infrastructure usage for decision making. While using data for an entire year is more accurate and reliable for decision making, collecting data for an entire network for a year is not economically feasible. Hence, currently, municipalities use partial infrastructure usage data collected for a limited period of time (i.e., 1 month) to make decisions about cycling networks.

The objective of this study is to develop a framework for estimating cycling infrastructure annual usage with short count data. First, a methodology is identified to evaluate the magnitude of impact of weather variables on cycling infrastructure usage and then an estimation model is developed that estimates the cycling infrastructure usage. The developed model is capable of estimating usage for different months of the year, and consequently, the model accurately estimated the annual usage using data from only one month.

An extended review of the literature on cycling was presented in chapter 2. Variables that affected cycling are identified and categorized into five groups: cyclists' characteristics, attitudes toward cycling, cycling infrastructure, built environment and network connectivity, and weather conditions. Then, studies that investigated cycling infrastructure usage trends and potential causes for cycling usage variation were presented. The previous literature investigated cycling usage trends and speculated that the cause of cycling infrastructure usage variation during a year was because of the effect of weather on cyclists. Many studies investigated the impact of variables that affect cycling found that temperature, precipitation, snow on the ground, wind speed and sunlight hours are the weather variables that affect cyclists.

Data used in this study were collected with 26 counters located at different locations in city of Calgary. The assumption about the effect of weather variables on usage variation throughout a year was confirmed by comparing cycling infrastructure usage trends with two weather variables trends (i.e. temperature, precipitation). Cycling infrastructure usage decreased in days with low temperature and high precipitation and increased in days with low precipitation and high temperatures. An estimation model was developed using weather variables to estimate cycling infrastructure usage. Users during weekday and weekend were different. Hence, two separate models were developed based on weekday and weekend usage (first model based on weekday usage and second model based on weekend usage). Impact of five weather variables were examined in this study namely temperature, precipitation, snow on the ground, wind speed, and sunlight hours. Data used in this study were collected at repeated times during a year for different infrastructure sections, as a result it was correlated over time and unbalanced for different infrastructure sections. Since data were correlated and unbalanced, generalized estimating equations (GEE) was chosen to evaluate the impact of weather variables on usage as GEE is an estimation approach used for correlated and unbalanced data.

The two developed models were examined, the results of both models were compared with observed usage of the infrastructure sections used for model development to determine which model more accurately estimated infrastructure usage. To estimate weekend usage, the second model (based on weekend usage) showed the best goodness of fit for all the tested infrastructure sections. However, to estimate weekday usage for infrastructure mainly used by commuter cyclists, the first model (based on weekday usage) fit better, while for infrastructure used mainly by recreational cyclists, the second model fit better. The user type ratio (UTR) was defined to determine how much infrastructure was used by different types of cyclists. The UTR showed the percentage of weekend usage compared to total usage. A high UTR indicated that usage was attributed to recreational cyclists, while a low UTR represented usage that was attributed to commuter cyclists. The UTR ratio was calculated based on September usage of all infrastructure types. The results showed that to estimate weekday usage for infrastructure sections with a UTR of less than 0.11, the first model should be used, while for infrastructure types with a UTR value of higher than 0.11, the second model fitted better.

The developed models were validated using two groups of infrastructures, grouped according to their infrastructure characteristics, that covered all varieties of infrastructure conditions. The first group was located in the downtown core area, while the second group was located outside the core area. The usage during 11 months of the year was estimated based on the September usage and was compared with the available observed usage. The model was able to estimate the usage for different months of the year with an average error of 15% and estimate the annual usage with an average error of 5% using observations from only one month.

5.2 RESEARCH CONTRIBUTIONS AND FINDINGS

The contributions of this study are related to the two sub-objectives defined in Chapter 1.

To evaluate the impact of weather variables on cycling infrastructure usage generalized estimating equations (GEE) was selected as an estimation approach. Cycling infrastructure usage data from the city of Calgary was used. Data was collected at repeated times in different locations (correlated and unbalanced). To analyse correlated unbalanced data, use of GEE was proposed in this study. The validation showed how accurately weather coefficients evaluated using GEE can estimate the usage. Among the variables that impacted cycling infrastructure usage, it was found that weather variables were accountable for the usage variation during a year since other variables were constant during a year for a specific infrastructure. The effect of five weather variables namely, mean temperature, precipitation, snow on the ground, wind speed, and sunlight hours, on cycling infrastructure usage was evaluated. The magnitude of the variables' effect on cycling infrastructure usage was evaluated for different months of the year and for different types of cyclists.

An estimation model was developed by defining a selection criterion for estimation usage of infrastructure sections based on their type of user. The model estimated the usage for weekends and weekdays separately for each infrastructure section since the types of cyclists were different during weekdays and weekends.

The model estimated the usage for different months of the year based on observations from one month. Two types of errors were calculated to examine the estimation accuracy. The first type of error represented the average error for estimating the usage variation during the year, and the second error represented the error for estimating the annual usage. These two types of errors showed how accurately the model could estimate the variation in usage and the annual usage of infrastructure using one month of observation data.

The results of this study will help planners for decision making on funding allocation about Calgary's cycling infrastructure network. Planners and decision makers can use the model to identify optimum locations to spend their money and to allocate funding between different infrastructure sections based on the usage.

5.3 DISCUSSION

Effect of weather variables on usage

The results of this study showed the effect of different weather variables on cycling infrastructure usage. Among the weather variables, temperature and sunlight hours showed a positive impact on cycling infrastructure usage, while precipitation, collected snow on the ground, and wind speed showed a negative impact on cycling infrastructure usage. Also, the results indicated that the magnitude of the impact of these variables is not the same for different months of the year. The impact of the variables with a negative impact, such as collected snow on the ground and precipitation, was generally higher during cold months of the year compared with warm months. Since, snow on the ground showed a high negative impact on cycling during cold months of the year, clearing snow in the cold months could be one of the efforts to encourage cycling in cold months. The study showed that within weather variables, temperature and precipitation are the main variables that impact cycling since they have the highest magnitude of impact on usage and their impact was significant for all months.

The impact of the weather variables was evaluated separately for weekday and weekend usage because weekends and weekdays have different types of cyclists. While weekday usage is generally attributed to commuter cyclists, usage on weekends is attributed to utilitarian cyclists. As the results of the model showed, the impact of weather variables was generally higher for weekend usage compared to weekday usage. The higher magnitude of the impact of weather variables on weekend usage indicated that utilitarian cyclists were more affected by weather conditions than commuter cyclists were.

The accuracy of the model

The validation showed that when estimating usage for different months of the year, the model had an error value between 14% and 19% with an average error of 17% for estimating the usage of different months of the year. This error represented the accuracy of the model in terms of estimating usage variation for different months of the year. When estimating the annual usage of infrastructure, the model had an error value between 1% and 9% with an average error value of 5 %. The higher average error of estimating usage for different months of the year was mainly attributed to the months with a low number of cyclists. Since the error was calculated as a difference between model estimation and actual usage divided by actual usage, the small difference between the model result and actual usage led to a high error in months with small actual usage.

Comparing the results of this study with previous studies showed that this study estimates the annual usage with higher accuracy. With using one-month observations, El-Esawey (2014) estimated annual cycling infrastructure usage with an average error of 11%. Nordback et al (2015) estimated annual usage with an average error of 15%. While this study estimates the annual usage with average error of 5%. The accuracy of estimating usage variation could not be compared with previous studies since none of the previous studies estimated the usage variation during a year. Since, the model presented in this study estimates the cycling usage with evaluating the impact of weather variables on cyclists, it could be used for any year with available weather data and it is not limited to a specific year.

Applicability of the Model

The estimation model developed in this study allows the usage throughout a year to be estimated based on the effect of weather conditions on cyclists. In order to develop the model, weather data from the city of Calgary was used. Since cities around the world have different types of climates, a climate classification system was used to distinguish cities and discuss about model applicability. There are several classification methods to classify climate. One of the well-known classification methods is the Köppen climate classification which has been used widely. This classification system was first introduced by a climatologist (Wladimir Köppen) in 1884. Köppen classifies the weather based on temperature and precipitation. The results of this research showed that among the weather variables tested in this study, temperature and precipitation were the two main variables that impact cycling infrastructure usage as they showed the highest impact on usage and their impact was significant for all months. Since the Köppen classifies climate based on temperature and precipitation, this classification system was used for discussing model applicability. According to Köppen, climate can be classified into 5 groups: tropical, arid, temperate, cold, and polar. The descriptions for each climate type are as follows:

1. Tropical: average temperature of 18 °C or higher for every month of the year. This type of climate has a significant amount of precipitation.
2. Arid: deficient precipitation throughout the year with a range of temperatures.
3. Temperate: average temperature between 0 °C and 18 °C for the coldest months of the year. This type of climate has at least one month with an average temperature above 10 °C.
4. Cold: at least one month with an average temperature below 0 °C and at least one month with an average temperature above 10 °C.
5. Polar: every month of the year has an average temperature below 10 °C.

According to the Köppen climate classification, Calgary is a city with a cold climate. Cold climate can also be classified into more details groups. There are 12 different types of cold climate (i.e. Dfa, Dfb, Dfc, Dfd, Dwa, Dwb, Dwc, Dwd, Dsa, Dsb, Dsc, Dsd). The descriptions for the first four cold climate types are as follows:

1. Dfa: this type of climate is humid with hot summer. The coldest month has an average temperature of below 0 °C. The warmest month has an average temperature above 22 °C.
2. Dfb: this type of climate is humid with a warm summer. The coldest month has an average temperature of below 0 °C. There are at least four months with an average temperature above 10°C.
3. Dfc: this type of climate has cold long winters. The coldest month has an average temperature of below 0 °C. Between 1 to 3 months have an average temperature above 10°C.
4. Dfd: this type of climate is extremely cold climate with the coldest month of the year with an average temperature below -38 °C. Between 1 to 3 months have an average temperature above 10°C.

According to the Köppen climate classification, Calgary is a city with a Dfb climate type. The effect of weather conditions on cyclists is not the same for people who live in different climate types. For example, people living in cold cities are acclimatized to cold weather compared to people who live in warmer cities; hence, the effect of temperature on cyclists is not the same. As a result, since the model presented in this study used data from a city with a cold climate, the model can be used to estimate cycling infrastructure usage for cities with a similar climate to Calgary. Cold climate type classified into three main categories (i.e. Df, Dw, Ds). These three categories have different precipitation levels. Since the model is developed based on Dfb climate type, it could be used for only Df climate type. For Df climate type, the model can be used for Dfa, Dfb, Dfc due to similarities between these climate types. However, since Dfd climate type has extremely cold winter the model should be tested with data from this climate type to find out if the could be used for this climate type. Figure 5.1 shows the world map according to the Köppen climate classification (Peel et al. 2007). As shown, in general, cities with a similar latitude as Calgary, such as most of North America and northern parts of Europe and Asia, have the same climate as Calgary according to the Köppen climate classification.

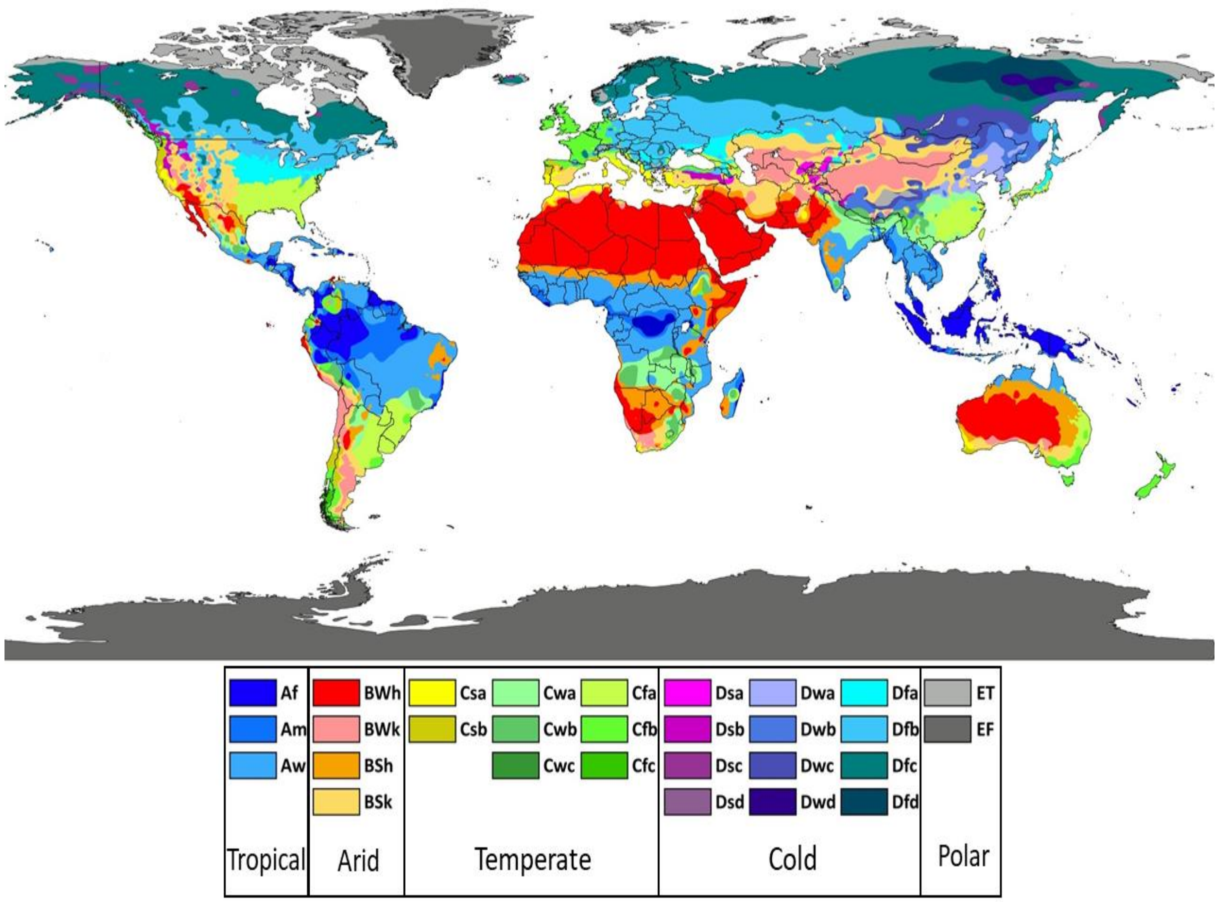


Figure 5.1 World map of climate classification (Peel et al. 2007)

5.4 LIMITATIONS OF THE MODEL

The followings are the limitations of this study.

Sample size

This study used data from 26 counters in the City of Calgary, and data was collected for less than two years. Due to a small sample size, especially for the utilitarian model, the effect of some weather variables for some months could not be evaluated. Developing the model using a larger sample size will improve the coefficients of the model.

Variable measurement

This study used weather variables that were measured at the Calgary International Airport. The amount of snow on the ground was different at the locations of the counters compared to the amount of snow at the Calgary International Airport. However, this data from the airport was the only available data on the amount of snow on the ground.

Cultural impact

This study did not investigate impact of cultural and psychological factors on cycling. Investigating the applicability of the model for other locations with similar climates, the impact of cultural factors on cycling should be considered.

Long-term estimations

Using the model for long term estimation, the impact of demographic changes should be included. For instance, if there is a demographic change (e.g. change in population), the model required another short count.

5.5 RECOMMENDATIONS FOR FUTURE WORK

Future research can be grouped into the following categories:

Model improvement

- Use long-term data to get more reliable and significant coefficients.

- Input the model with accurate values of snow on the ground for each infrastructure type to improve model accuracy.

Model development and application

- Develop models for different types of climates using data from other climate types.
- Develop models based on observations from different months of a year and investigate threshold values of UTR for different months of observation.

REFERENCES

Amiri, M., Sadeghpour, F. (2014). Cycling characteristics in cities with cold weather Total green house gas emissions in Canada 199 0-2010. *Sustainable Cities and Society*, *14*, 397–403.

Aultman-Hall, L., Kaltenecker, M. G. (1999). Toronto bicycle commuter safety rates. *Accident Analysis and Prevention*, *31*(6), 675–686.

Bell, S., Morse, S. (2012). *Sustainability indicators: measuring the immeasurable?*. Routledge

Bergström, A., Magnusson, R. (2003). Potential of transferring car trips to bicycle during winter. *Transportation Research Part A: Policy and Practice*, *37*(8), 649-666.

Bernhoft, I. M., Carstensen, G. (2008). Preferences and behaviour of pedestrians and cyclists by age and gender. *Transportation Research Part F: Traffic Psychology and Behaviour*, *11*(2), 83–95.

Brandenburg, C., Matzarakis, A., Arnberger, A. (2004). The effects of weather on frequencies of use by commuting and recreation bicyclists. *Advances in tourism climatology*, *12*, 189-197.

Brandenburg, C., Matzarakis, A., Arnberger, A. (2007). Weather and cycling – a first approach to the effects of weather conditions on cycling, *67*, 61–67.

Buehler, R. (2012). Determinants of bicycle commuting in the Washington, DC region: The role of bicycle parking, cyclist showers, and free car parking at work. *Transportation Research Part D: Transport and Environment*, *17*(7), 525–531.

Caulfield, B., Brick, E., McCarthy, O. T. (2012). Determining bicycle infrastructure preferences - A case study of Dublin. *Transportation Research Part D: Transport and Environment*, *17*(5), 413–417.

Cavill, N., Davis, A. (2007). Cycling and health-what's the evidence?.

Cox, D. R. (1983). Some remarks on overdispersion. *Biometrika*, *70*(1), 269-274.

Daley, M., Rissel, C. (2011). Perspectives and images of cycling as a barrier or facilitator of cycling. *Transport Policy*, *18*(1), 211–216.

Deenihan, G., Caulfield, B. (2014). Estimating the health economic benefits of cycling. *Journal of Transport* & *Health*, *1*(2), 141-149.

Dickinson, J. E., Kingham, S., Copsey, S., Hougie, D. J. P. (2003). Employer travel plans, cycling and gender: will travel plan measures improve the outlook for cycling to work in the UK?. *Transportation research part D: transport and environment*, *8*(1), 53-67.

Dill, J. (2009). Bicycling for transportation and health: the role of infrastructure. *Journal of public health policy*, *30*(1), S95-S110.

Dill, J., Carr, T. (2003). Bicycle Commuting and Facilities in Major U.S. Cities: If You Build Them, Commuters Will Use Them. *Transportation Research Record*, *1828*(1), 116–123.

Donnell, E., Karwa, V., Sathyanarayanan, S. (2009). Analysis of effects of pavement marking retroreflectivity on traffic crash frequency on highways in North Carolina: application of artificial neural networks and generalized estimating equations. *Transportation Research Record: Journal of the Transportation Research Board*, (2103), 50-60.

Esawey, M. (2014). Estimation of annual average daily bicycle traffic with adjustment factors. *Transportation Research Record: Journal of the Transportation Research Board*, (2443), 106-114.

El Esawey, M., Lim, C., Sayed, T., Mosa, A. I. (2013). Development of daily adjustment factors for bicycle traffic. *Journal of Transportation Engineering*, *139*(8), 859-871.

El Esawey, M., Lim, C., & Sayed, T. (2015). Development of a cycling data model: City of Vancouver case study. *Canadian Journal of Civil Engineering*, *42*(12), 1000-1010.

El Esawey, M., Mosa, A. I. (2015). Determination and application of standard K factors for bicycle traffic. *Transportation Research Record: Journal of the Transportation Research Board*, (2527), 58-68.

Flynn, B. S., Dana, G. S., Sears, J., Aultman-hall, L. (2012). Weather factor impacts on commuting to work by bicycle. *Preventive Medicine*, *54*(2), 122–124.

Garrard, J., Rose, G., Lo, S. K. (2008). Promoting transportation cycling for women: The role of bicycle infrastructure. *Preventive Medicine*, *46*(1), 55–59.

Garrett-Peltier, H. (2011). Pedestrian and bicycle infrastructure: A national study of employment impacts. *Amherst, MA: Political Economy Research Institute*.

Gatersleben, B., Appleton, K. M. (2007). Contemplating cycling to work: Attitudes and perceptions in different stages of change. *Transportation Research Part A: Policy and Practice*, *41*(4), 302–312.

Heinen, E., Maat, K., Van Wee, B. (2011). The role of attitudes toward characteristics of bicycle commuting on the choice to cycle to work over various distances. *Transportation Research Part D: Transport and Environment*, *16*(2), 102–109.

Heinen, E., Van Wee, B., & Maat, K. (2010). Commuting by bicycle: an overview of the literature. *Transport reviews*, *30*(1), 59-96.

Hutchings, C. B., Knight, S., Reading, J. C. (2003). The use of generalized estimating equations in the analysis of motor vehicle crash data. *Accident Analysis & Prevention*, *35*(1), 3-8.

Hunt, J. D., & Abraham, J. E. (2007). Influences on bicycle use. *Transportation*, 34(4), 453-470.

Jacobsen, P. L. (2003). Safety in numbers: more walkers and bicyclists, safer walking and bicycling. *Injury prevention*, 9(3), 205-209.

Krizek, K. J., Roland, R. W. (2005). What is at the end of the road? Understanding discontinuities of on-street bicycle lanes in urban settings. *Transportation Research Part D: Transport and Environment*, *10*(1), 55–68.

Lindsey, G., Chen, J., Hankey, S. (2013). Adjustment Factors for Estimating Miles Traveled by Non-motorized Traffic, *6351*, 2–22.

Li, Z., Wang, W., Liu, P., Ragland, D. R. (2012). Physical environments influencing bicyclists' perception of comfort on separated and on-street bicycle facilities. *Transportation Research Part D: Transport and Environment*, *17*(3), 256–261.

Liang, K. Y., Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13-22.

Larsen, J., El-Geneidy, A. (2011). A travel behavior analysis of urban cycling facilities in Montr??al, Canada. *Transportation Research Part D: Transport and Environment*, *16*(2), 172–177.

Lord, D., Persaud, B. (2000). Accident prediction models with and without trend: application of the generalized estimating equations procedure. *Transportation Research Record: Journal of the Transportation Research Board*, (1717), 102-108.

Ma, M., Yan, X., Abdel-Aty, M., Huang, H., Wang, X. (2010). Safety analysis of urban arterials under mixed-traffic patterns in Beijing. *Transportation Research Record: Journal of the Transportation Research Board*, (2193), 105-115.

Maring, W., Van Schagen, I. (1990). Age dependence of attitudes and knowledge in cyclists. *Accident Analysis & Prevention*, *22*(2), 127-136.

Martens, K. (2004). The bicycle as a feedering mode: experiences from three European countries, *9*, 281–294.

Meng, M., Zhang, J., Wong, Y. D., & Au, P. H. (2016). Effect of weather conditions and weather forecast on cycling travel behavior in Singapore. *International journal of sustainable transportation*, *10*(9), 773-780.

Miranda-Moreno, L., Nosal, T., Schneider, R., Proulx, F. (2013). Classification of bicycle traffic patterns in five North American Cities. *Transportation Research Record: Journal of the Transportation Research Board*, (2339), 68-79.

Moritz, W. (1997). Survey of North American Bicycle Commuters: Design and Aggregate Results. *Transportation Research Record*, *1578*(1), 91–101.

Nankervis, M. (1999). The effect of weather and climate on bicycle commuting. *Transportation Research Part A: Policy and Practice*, *33*(6), 417–431.

Nordback, K., Marshall, W., Janson, B., & Stolz, E. (2015). Estimating annual average daily bicyclists: Error and accuracy. *Transportation Research Record: Journal of the Transportation Research Board*, (2339), 90-97.

Osama, A., Sayed, T., Bigazzi, A. Y. (2017). Models for estimating zone-level bike kilometers traveled using bike network, land use, and road facility variables. *Transportation Research Part A: Policy and Practice*, *96*, 14–28.

Osberg, J. S., Stiles, S. C., Asare, O. K. (1998). Bicycle safety behavior in Paris and Boston. *Accident Analysis and Prevention*, *30*(5), 679–687.

Parkin, J., Meyers, C. (2010). The effect of cycle lanes on the proximity between motor traffic and cycle traffic. *Accident Analysis and Prevention*, *42*(1), 159–165.

Parkin, J., Wardman, M., Page, M. (2008). Estimation of the determinants of bicycle mode share for the journey to work using census data. *Transportation*, *35*(1), 93–109.

Peel, M. C., Finlayson, B. L., McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and earth system sciences discussions*, *4*(2), 439-473.

Plaut, P. O. (2005). Non-motorized commuting in the US. *Transportation Research Part D: Transport and Environment*, *10*(5), 347–356.

Poch, M., Mannering, F. (1996). Negative binomial analysis of intersection-accident frequencies. *Journal of transportation engineering*, *122*(2), 105-113.

Publique Toronto, S. (2012). Road to Health: Improving Walking and Cycling in Toronto.

Pucher, J., Buehler, R. (2008). Making cycling irresistible: lessons from the Netherlands, Denmark and Germany. *Transport reviews*, *28*(4), 495-528.

Pucher, J., Dill, J., Handy, S. (2010). Infrastructure, programs, and policies to increase bicycling: an international review. *Preventive medicine*, *50,* S106-S125.

Rodgers, G. B. (1995). Bicyclist deaths and fatality risk patterns. *Accident Analysis and Prevention*, *27*(2), 215–223.

Rodgers, G. B. (1997). Factors associated with the crash risk of adult bicyclists. *Journal of Safety Research*, *28*(4), 233-241.

Sadeghpour, F., Isaac, S., & Amiri, M. (2015). Winter cycling in very cold climate-a case study in Calgary. *Journal of Construction Project Management and Innovation*, *5*(2), 1238-1265.

Saneinejad, S., Roorda, M. J., Kennedy, C. (2012). Modelling the impact of weather conditions on active transportation travel behaviour. *Transportation Research Part D: Transport and Environment*, *17*(2), 129–137.

Sisson, S. B., Tudor-Locke, C. (2008). Comparison of cyclists' and motorists' utilitarian physical activity at an urban university. *Preventive Medicine*, *46*(1), 77–79.

Spencer, P., Watts, R., Vivanco, L., Flynn, B. (2013). The effect of environmental factors on bicycle commuters in Vermont: Influences of a northern climate. *Journal of Transport Geography*, *31*, 11–17.

Stutts, J. C., Hunter, W. W. (1999). Motor vehicle and roadway factors in pedestrian and bicyclist injuries: An examination based on emergency department data. *Accident Analysis and Prevention*, *31*(5), 505–514.

Titze, S., Stronegger, W. J., Janschitz, S., Oja, P. (2008). Association of built-environment, social-environment and personal factors with bicycling as a mode of transportation among Austrian city dwellers. *Preventive Medicine*, *47*(3), 252–259.

Twaddle, H., Hall, F., Bracic, B. (2011). Latent Bicycle Commuting Demand and Effects of Gender on Commuter Cycling and Accident Rates. *Transportation Research Record: Journal of the Transportation Research Board*, *2190*(1), 28–36.

Vandenbulcke, G., Thomas, I., de Geus, B., Degraeuwe, B., Torfs, R., Meeusen, R., Int Panis, L. (2009). Mapping bicycle use and the risk of accidents for commuters who cycle to work in Belgium. *Transport Policy*, *16*(2), 77–87.

Walker, I. (2007). Drivers overtaking bicyclists: Objective data on the effects of riding position, helmet use, vehicle type and apparent gender. *Accident Analysis and Prevention*, *39*(2), 417–425.

Wang, X., Abdel-Aty, M. (2008). Modeling left-turn crash occurrence at signalized intersections by conflicting patterns. *Accident Analysis & Prevention*, *40*(1), 76-88.

Waters, N. (2016). Tobler's First Law of Geography. International Encyclopedia of Geography: People, the Earth, *Environment and Technology*, 1-15.

Wegman, F., Zhang, F., Dijkstra, A. (2012). How to make more cycling good for road safety? *Accident Analysis and Prevention*, *44*(1), 19–29.

Zhao, J., Wang, J., Xing, Z., Luan, X., Jiang, Y. (2018). Weather and cycling: Mining big data to have an in-depth understanding of the association of weather variability with cycling on an off-road trail and an on-road bike lane. *Transportation Research Part A: Policy and Practice*, *111*, 119-135.