Route Choice Modeling of Cyclists in Toronto

by

Kathryn Grond

A thesis submitted in conformity with the requirements
for the degree of Master of Applied Science
Civil Engineering
University of Toronto

Route Choice Modeling of Cyclists in Toronto

Kathryn Grond
Master of Applied Science

Civil Engineering

University of Toronto

2016

Abstract

This thesis builds on a growing body of research that seeks to understand how the physical and environmental attributes of the road network influence cyclists' route choice. Better understanding of the costs and benefits of physical infrastructure can be used by planners to prioritize investment in cycling facilities. The thesis uses a high quality GPS dataset of bike trips recorded between August 23 and September 23 2015, using the City of Toronto's cycling app. Trip route characteristics are obtained by matching the GPS traces to a detailed GIS network dataset of road attributes. A path-size multinomial logit model is used to assess the utility of cycling facilities and the costs of road features, such as high traffic volumes, steep hills and turns at busy intersections. The study also examines the route variation based on attitudes and demographic characteristics of cyclists.

Acknowledgments

I would first like to thank my supervisor Eric J. Miller, for giving me the freedom to explore a topic for which I am passionate. Thank you for providing guidance at weekly meetings and the support I needed to complete the project. Thank you to professor Marianne Hatzopoulou, the second reader of my thesis, for the suggestions and questions that strengthened the project.

I am very indebted to professor Darren Scott and Ron Dalumpines from McMaster University for sharing their map matching code, and to especially to Ron for your time helping with the tool. I am also very thankful to Andrew Clarry for his work on developing tools to count turns and matching road networks, as well as his support with the python coding of the map matching and choice set generation.

I am thankful to the City of Toronto for sharing their cycling app data and to the UofT Map and Data Library for support with ArcGIS. I would also like to thank Alan Tonks for the scholarship in support of my studies.

Lastly, I would like to thank my parents for their tremendous support, love and encouragement to pursue my Master's degree, and to my husband Robert for his patience, compassion and camaraderie as we both finished our dissertations.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Project Overview

This project seeks to better understand the physical factors and personal characteristics that influence cycling route choice in the City of Toronto. The project analyses route trip data collected with a smartphone app, and uses a discrete choice model to identify road attributes with significant costs and benefits to cyclists. The variation in preference for road attributes based on trip purpose and cyclist demographics and attitudes are also estimated. An understanding of the utilities and trade-offs made by cyclists is invaluable information for future research and planning of cycling networks.

The objectives of this project are to:

- Measure the benefits of cycling infrastructure in route choice, as well as the costs of barriers such as streetcar tracks, steep hills and heavy traffic. Cycling infrastructure is a significant factor in cycling route choice, and the trade-off with respect to other road attributes for cycling in Toronto is compared to cycling route choice studies in other cities.
- Investigate the impact of intersections and turn maneuvers, such as left turns at busy arterial streets and turns at traffic lights. Intersections are the most common location of cycling accidents, and yet they often neglected in the design and planning of cycling infrastructure. A detailed examination of influence of turns and intersections characteristics on cycling route choice has only been studied once in Portland by Broach et al. (2012).
- Compare the variation in preference for trip attributes based on user characteristics. Studies have found that the utility route attributes varies based on trip purpose, time of day, cyclist gender and frequency of cycling. Understanding the taste heterogeneity of cyclists can help planners develop more inclusive and equitable cycling infrastructure.

1.2 Project Motivation

Active transportation is an integral travel mode for sustainable transportation. Increasing the share of cycling and walking benefits individuals by promoting physical activity and offering more choices for travel, and active transportation also benefits cities and the environment by reducing congestion and pollution from fossil fuel based transportation. Cycling for utilitarian trips is an attractive, albeit minor, travel mode in Toronto. Cycling can often be faster and more reliable than transit or auto modes in the downtown where travel distances are shorter and there is higher traffic congestion and delays. In the 2011 Transportation of Tomorrow Survey, the cycling mode share for the City of Toronto was 1.72%, and was as high as 9.3% in some downtown wards (Data Management Group, 2013). On arterial roads in the downtown with bike lanes such as College St., the volume of cyclists is almost equal to the volumes of cars at peak travel times (Cycle Toronto, 2016). Surveys of both cyclists and non-cyclists show that separated infrastructure provides cyclists with an increased sense of safety, and studies have shown increases of cycling volume on roads after the installation of new cycling infrastructure. There was a 300% and 200% increase in travel on Sherbourne St. and Richmond St. respectively after the installation of separated cycle tracks (City of Toronto, 2014; Toronto Star, 2015). Yet the efficacy of adding cycling infrastructure to encourage more cycling trips is still not fully understood.

City transportation engineers and planners are installing more cycling infrastructure in an effort to attract new cyclists and to provide safer travel conditions for existing cyclists. The expansion and investment in active transportation is also supported by the City of Toronto's complete streets policy that promotes the construction of safe road facilities for all road users (City of Toronto, 2014). In June 2016, the City of Toronto approved a new 10-year cycling plan that will add 280 kilometres of new cycling infrastructure, doubling the size of the cycling network. Several of the plan's new bike lanes are proposed along major corridors, where contested reallocation of space will require planners to quantify the cost and benefits to all road users. Better evidence to support the investment in cycling infrastructure is required because changes in road design are often opposed by local residents and businesses who find that these investments come at a cost of reduced parking and increased vehicular congestion. Regional transportation models are not currently capable of assessing the impact of cycling infrastructure on cycling mode or route choice, and this in part due to the lack of detailed data collection for cycling trips.

The cycling data used in this study was collected as part of the cycling network plan's analysis to understand the existing travel patterns of cycling on the road network. A cycling route choice model can provide a better understanding of the costs and benefits of road characteristics of the paths travelled by cyclists. This study's route choice model and the City's route choice data can in future be used to develop new approaches and tools to integrate active transportation into regional travel demand models in Toronto. The utilities of road attributes from the route choice model can be used to generate a cycling trip assignment model, which can be used for a scenario analysis of planned improvements to the cycling network. By better understanding which obstacles pose the highest cost and which infrastructure can counter those costs can help planners identify location and designs that will provide an effective increase in accessibility for cyclists.

1.3 Project Approach

This project makes use of a new and valuable GPS dataset of cycling trip routes, collected by app commissioned by the City of Toronto. The data show where cyclists are riding, along with key demographics and attitudinal information about the trip. The app's GPS data is analysed using a route choice model to understand the value that cyclists place on cycling facilities and the costs of obstacles such as steep hills and busy intersections. The project's route choice model involves the following steps:

- Build a network data set of important characteristics for cycling;
- Convert the GPS traces into trips and match them to the network and develop a series of characteristics about the trips as well as the users;
- Create a choice set of feasible alternative routes;
- Use discrete choice modeling to estimate the relative utility of each road attribute on route choice.

Chapter 2

Literature Review

This chapter reviews the literature on cycling route choice, including proposed route choice modeling methods and the findings of previous cycling route choice studies. The first section provides an overview of the factors that have been found to influence cycling mode choice and route choice, as well as the factors that influence route formation and knowledge. The findings of other route choice studies on the impact of road network attributes and user characteristics are also summarized. A review of route choice modeling methods examines the various methods for data collection, generating and evaluating choice sets, as well as discrete choice modelling frameworks.

2.1 Factors that Influence Cycling Route Choice

2.1.1 Factors Valued by Cyclists: Effort, Safety and Pleasantness

Route choice theory for other modes of transportation focuses primarily on maximizing utility and minimizing travel costs such as travel time and fares. However, cycling route choice includes more factors than travel time and cost. Stated preference (SP) surveys show that travel costs of a cycling trip can be understood not just as travel time but in terms of effort, including distance and elevation. Furthermore, cyclists will often choose an alternative route to the shortest distance path that has features that increase perceived safety. Revealed preference (RP) studies of observed cycling routes can be used to measure the trade-off that cyclists make between distance and routes with features such as bike lanes and low traffic speeds, as well as avoidance of busy intersections. A third group of factors that contribute to route choice is route pleasantness. Route pleasantness factors include smoothness of pavement, the presence of trees, and land-use types. Such factors are often selected as important factors by some cyclists in stated preference surveys. It is more difficult to define quantitative measures of these factors and they have not often been included in RP studies. Table 2.1 summarizes the factors that have been found either in SP or RP studies to possibly influence cycling route choice.

Table 2.1 Factors of Cycling Route Choice Studies

Built Environment Features	Road Characteristics	Trip Characteristics	User Characteristics
A. Street trees and shade	F. Bike facility	O. Trip length	X. Age
B. Adjacent land use	G. Elevation	P. Travel time	Y. Gender
C. Population density	H. Road class	Q. Average speed	Z. Income
D. Proximity to CBD	I. Traffic volume	R. Left and right turns	AA. Cycling experience
E. Number of auto destination	J. Number and width of travel lanes	T. Bike lane continuity	BB. Comfort cycling
	K. Speed limit	U. Trip purpose	CC. Cycling frequency
	L. On-street parking	V. Time of day	DD. Winter cyclist
	M. Streetcar tracks	W. Wrong-way travel	EE. History of accident
	N. Intersection control (traffic lights, stop signs)		

Route Choice Study Authors	Type of Study	Study Area	Cycling Factors Included
Hood et al (2011)	Revealed Preference	San Francisco, CA (United States)	B, F, G, I, J,K, N, O, R, U, W, X, Y, CC
Broach et al (2012)	Revealed Preference	Portland, OR (United States)	F, G, I, N, O, R, U, X, Y, Z, AA, BB, CC, DD
Menghini et al (2010)	Revealed Preference	Zurich (Switzerland)	F, G, N, O, Q
Casello et al (2014)	Revealed Preference	Waterloo, ON (Canada)	F, G, I , O, Y, BB
Khatri (2015)	Revealed Preference	Phoenix, AZ (United States)	F, G, I, N, O, R, V, W
Pereira Segadhila et al (2014)	Revealed Preference & Stated Preference	Sao Carlos, SP (Brazil)	A, F, H, I, J, K, L, N, O, P, Q, U, V, Y
Sener et al (2009)	Stated Preference	Texas (United States)	F, G, I, J, K, L, N, O, U, X, Y, AA
Larsen & El-Geneidy (2011)	Revealed Preference & Stated Preference	Montreal, QC (Canada)	D, E, F, N, O, U, X, Y, BB, CC, EE
Winters et al (2010)	Revealed Preference & Stated Preference	Vancouver, BC (Canada)	A, B, C, F, G, N, O, X, Y, Z, CC

2.1.2 Summary of Cycling Route Choice Studies

All studies reviewed find that travel distance has an important and significant cost for cycling route choice. Even so, most cyclists will deviate from the shortest path for alternatives that are perceived to be safer or more attractive. The following section summarizes findings from previous RP and SP cycling route choice studies.

Cycling Facilities

Most stated preference and revealed preference studies find a significant preference for bike lanes and trails, but their utility varies. A study of cyclists in Portland finds the benefit of bike lanes approximately offset the impact of high traffic volumes, and is comparable to a quiet street (Broach et al., 2012). The study also finds that cyclists would travel an extra 10-18% and 16-26% for bike boulevards and off-road trails respectively, with the lower range for commuting cyclists (Broach et al., 2012). This indicates that commuter cyclists place a higher value on lower travel time and are less willing to detour to travel on a bike facility. One exception is a study conducted in Guelph, Ontario, by Aultman-Hall et al. (1997), who found that bike routes do not have a significant impact on route choice; however, the lack of on-road cycling infrastructure in the study area at the time could explain the insignificant preference the few available off-road trails.

Hills

The difference in quantifying road grade, either as a total rise over the route, or by a percentage of route on steep grades, makes the comparison of results difficult. SP and RP studies in Zurich, Portland, San Francisco, Waterloo, and Texas find positive road grades to have a significant negative coefficient in the route choice model, but less so than travel time and cycling facilities (Menghini et al., 2010; Broach et al., 2012; Hood et al., 2011; Casello and Usyukov, 2014; Khatri, 2015). A regression analysis of observed vs. shortest path trips in Vancouver by Winters et al. (2012) finds that observed routes do not have significantly less hills than the shortest route, but this may be due to observed routes being compared exclusively to the shortest path, and the fact that a low grade alternative is not included in the analysis.

Traffic Volumes

An SP route choice model by Sener et al. (2009) found that avoiding high-traffic streets was the second highest rated factor after minimizing travel time. Broach et al (2012) found that cyclists would detour 137% to avoid streets with 10,000-20,000 vehicles per day and Khatri (2015) found a negative coefficient for traffic volume. While some studies found that avoiding high traffic volumes on streets is a significant factor, others found that traffic volume attributes have positive coefficients and are excluded from their final models (Casello and Usyukov, 2014; Hood et al., 2011). Broach et al. (2012) also found that factors such as traffic volume, which was of high importance to cyclists in previous SP surveys, was not significant in the route choice model.

Road Class

Road classifications such as arterial, collector or local roads are less often included in route choice studies. This may be due to the indirect influence that road class has on cycling behaviour, and that road class is correlated with traffic speeds and volume. A study of bike trips in Vancouver by Winters et al. (2011) reports that bike trips are distributed on average by 33% on local streets, 31% on arterials, 21% on collectors, and 14% off-street trails.

Turns

Turns on a route add a small delay in travel time, as well as a mental cost for remembering each turn along the route. Intersections are also the source of most accidents and conflicts with vehicles, so one would expect cyclists to avoid busy intersections, which are correlated with busy, arterial streets. Minimizing turns has a smaller impact on route choice than bike lanes and traffic volume, but studies do consistently find that turn frequency has a negative utility in route choice models (Hood et al., 2011; Broach et al, 2012; Khatri, 2015).

Intersections

The impact of the types of intersection is less clear. Signal-controlled intersections can provide a safer - and often faster - crossing of busy streets. However, signalized intersections are more likely to add a delay to the trip by stopping at a red light or in slowing down to navigate conflicts with right-turning cars. Broach et al. (2012) find that cyclists add between 2-3.6% of their trip distance to avoid one additional left turn or crossing per mile at a signalized intersection, whereas Khatri (2015) find that bike-share cyclists in Phoenix derive a positive benefit and detour between 5-8% for an additional traffic light. The SP survey by Sener et al. (2009) found that avoiding a high frequency of stoplights and traffic lights was an important factor for cyclists, which is more important than avoiding steep hills, but less important that traffic volume. Menghini et al. (2010) also found that traffic lights have a negative utility.

2.1.3 Moderating Variables on Cycling Preference: Trip Purpose, Time of Day and User Characteristics

Both the theoretical and observed significance of the factors affecting route choice are still being studied, and may vary based on demographics or cycling experience, as well as across cities with different built forms and transportation cultures. Studies have identified taste variations based on reported socio-demographic characteristics. One study by Sener et al. (2009) uses a panel mixed logit on SP data and finds that there is a significant effect of unobserved personal preferences in the route choice model. A study in Melbourne, Australia by Garrard et al. (2008) finds that women prefer cycling facilities that are separated from traffic, and Hood et al. (2011) find that women and commuters avoid steep hills more than men or non-commuters. Tilahun et al. (2007) find that women are even more likely than men to choose facilities perceived as safer. Larsen and El-Geneidy (2011), however, do not find any impact of gender or age on the type of cycling facilities.

Not all cyclists will sacrifice route directness for safety in the same amount, and the trade-offs may vary for the same person under different circumstances, such as trip purpose or time of day. It has been found that inexperienced cyclists have a stronger preference for cycling infrastructure compared to experienced cyclists, and they will make a longer detour to use such facilities (Larsen & El-Geneidy, 2011; Hood et al., 2011). However, a study in Zurich, Switzerland finds that faster cyclists, which may indicate that they are more experienced, prefer bicycle facilities and marked routes more than slower cyclists, who might be less familiar with the bike network (Menghini et al., 2010). Broach et al. (2012) find that commuters are less likely to add travel distance to their route to increase attributes such as bike lanes or to reduce turns, but they are more sensitive to traffic volumes than non-commuters. This would make sense as roads with high traffic volumes will most likely be congested and unpleasant during rush hour, when most people commute. A study of bike-share users in Phoenix finds that cyclists travelling during peak hours are only willing to detour half as much from the shortest path compared to off-peak time periods (Khatri, 2015). The study also finds that casual subscribers are more sensitive to the number of right and left turns compared to regular subscribers.

2.2 Route Choice Modeling Methods

2.2.1 Approaches to Route Choice Analysis

Studies comparing aggregate rates of cycling and levels of infrastructure can provide an overview of general trends in mode share, but are often insufficiently fine-grained to identify causal relationships. Studying individual preferences using SP in surveys or RP of observed trips can provide more insight into the impact of physical infrastructure on cycling route choice. SP surveys can shed light on the preferences and trade-offs that both cyclists and non-cyclists might make between various types of cycling routes and infrastructure (Sener et al., 2009). This is especially useful for determining the preference for features that are not currently present on the road, such as a new bike lane design.

However, depending on the formulation of the survey questionnaire, and the many response biases endemic to the survey format, the preferences declared in the surveys may not correspond to actual behaviour. Respondents may express a high preference for separated facilities if they think it will impact the planning of new bike lanes, but this stated preference might differ from actual behaviour, where the directness of route and elevation are more important than road separation (Winters et al., 2010). It is helpful, therefore, to pair SP surveys with observed behaviour data to test the correlation between observe behaviour and the stated preferences.

The benefits of RP studies are that route attributes can be measured and compared to the shortest path to evaluate the relative value of those attributes. Some studies use a regression analysis to find significant factors of the actual route compared to the shortest path (Pereira Segadilha & da Penha Sanches, 2014; Larsen & El-Geneidy, 2011; Winters et al., 2010). Other studies use a multinomial discrete choice model to estimate the relative utility of each attribute (Menghini et al., 2010; Broach et al., 2012; Hood et al., 2011; Casello and Usyukov, 2014; Khatri, 2015). A limitation of RP studies compared to SP surveys is that the full set of alternatives considered is not observed. Because these studies cannot observe the decision process, route familiarity, and the alternative paths considered, an alternative choice set must be generated by the researcher.

2.2.2 Data Collection Methods for Revealed Preference Data

Recent developments in GPS devices and smartphone apps that collect detailed spatial data of route paths have provided researchers and planners with a rich dataset to study actual route paths taken. RP cycling route choice studies use GPS-based trip data collected by stand-alone GPS devices (Pereira Segadilha & da Penha Sanches, 2014; Broach et al., 2012; Casello and Usyukov, 2014) or a smartphone app to record trip path locations (Hood et al., 2011; Menghini et al., 2010; Khatri, 2015). The use of GPS can provide researchers with a spatially and temporally accurate representation of a trip path. However, interference from tall buildings, overpasses and ravines can interfere with the accuracy of the GPS signal. Less accurate GPS data can make matching GPS traces to the network with a map-matching algorithm more difficult. Data collected with smartphone apps that rely on users to start and end the recording of trips increase the change of incomplete recorded paths that lead to an inaccurate recording of trip origins and destinations and trip duration. Furthermore, trips are only recorded when the user remembers to use the device, so we may not have a complete travel record over the course of the day, leading to further reporting bias in the sample.

There are several other types of smartphone apps that record recreational trips or passive travel survey apps that can be used for research purposes. For example, the popular recreational running and cycling smartphone app *Strava* has been used to study cycling travel patterns in cities such as Seattle, Washington (Strava Metro, 2015). Passive smartphone apps, such as the Waterfront TO App continuously collect locational data for the whole day using wi-fi and cell tower data, minimizing battery and data use. Such an app can be used for multiple days, which can help reduce the bias in selecting and recording trips, but this method relies on trip and mode detection algorithms to identify trips and activities. Passive data collection apps have the potential to provide a geographically and demographically representative survey sample is they are incorporated into regional travel surveys.

Menghini et al. (2010) used a privately collected dataset using continuously recording GPS devices and thus had to extract trips and mode from the GPS points recorded throughout the day. Because they used a third party application, however, the researchers did not have access to data about the trip purpose or user characteristics. Observed routes can also be recreated by users on hand-drawn maps (Aultman-Hall, 1996) or on web-based platforms (Winters et al., 2010; Larsen and El-Geneidy, 2011), but these responses may be inaccurately reported.

2.2.3 Choice Set Generation

Choice set generation is a crucial component of route choice modeling. For each origin-destination (OD) pairs, alternatives routes to the chosen route are required for choice modeling. Because the number of routes in the universal choice set is very large and unknown, a subset of feasible alternative routes must be selected for the model estimation. Once the choice set is generated, it is filtered to remove identical and highly overlapping routes, as they would not be considered as separate routes by the traveller and, in any case, their inclusion would violate the independence of irrelevant alternatives (IIA) property of multinomial models. The goal of these methods is to generate a complete and unbiased set of routes that leads to a correct estimation of cyclist's preference for attributes. The composition of the choice set is not trivial, and has been found to affect the model parameters significantly (Prato and Bekhor 2007; Broach et al., 2010).

A crucial assumption in route choice modelling is that the model contains all routes considered by the traveller (Bekhor et al., 2006). It is assumed that travellers have complete information about travel costs for all routes and they choose the route that maximizes benefits and minimizes travel cost. This assumption, however, is not often possible in reality; people have imperfect knowledge of network costs and usually only identify a few possible paths (Prato, 2009). How different groups of people develop knowledge of possible routes and how they perceive the differences of these routes should also be taken into consideration when developing the choice set. Studies have hypothesised that the frequency of cycling might influence the knowledge of the route qualities (Menghini et al., 2010). Commuting cyclists have been found to prefer shorter routes with less hills (Broach et al., 2012 and Hood et al., 2011) and another study found that faster cyclists, thought to be more experienced, use more bike lanes than slower cyclists (Menghini et al., 2010

2.2.4 Types of Choice Set Generation Methods

K-shortest path search

The simplest approach to generating a feasible alternative route to the chosen path uses the Dijkstra's shortest path algorithm that finds the minimal cost path in a network. Repeated path search algorithms are the most common technique. K-shortest path methods include link elimination and link penalty. They use a deterministic approach that iteratively searches for routes using Dijkstra's cost minimization algorithm, where successive, shortest path searches vary the network costs, either by increasing the link cost by a pre-set increment or by eliminating one or all links of previous shortest path searches. There are many variations of these techniques, which are reviewed by Prato (2009).

A stochastic approach to K-shortest path search uses simulation to define the link cost by drawing from a specified random distribution of costs. This variability is thought to account for travelers' imperfect knowledge of route costs. Doubly stochastic methods account for the heterogeneity of user preferences (Neilsen, 2000; Bovy & Fiorenzo-Catalano, 2007).

Labelling

K-shortest paths can only minimize one generalized cost function, but travellers may have different preferences for route attributes. Surveys and previous route choice models have shown that cyclists have significant taste heterogeneity for route attributes, and these variations will have implications for cycling policies. The labelling technique proposed by Ben-Akiva et al. (1984) generates a choice set by creating separate generalized costs for each attribute, called a label, which is then minimized in a shortest path search. The effectiveness of this method is particularly affected by the researcher's specification of the label's cost, which can lead to poor model parameter estimates.

Broach et al. (2010) propose a calibration method for labeling choice set generation that produces a series of routes where the cost of the label is incrementally increased to a maximum value. The maximum value is calibrated so that the distribution of the label's attribute generated in the choice set is well-matched to the distribution of the attribute in the observed choice set. Systematically varying the distribution of each model attribute in the choice set helps to reduce the bias of researcher specification in generating the choice set (Broach et al., 2010).

Constrained Enumeration

Constrained enumeration methods use a branch and bound path search algorithm that uses various decision rules of route formation, such as avoiding backtracking, avoiding congested roads and junctions, or repeatedly leaving and returning to a road link (Prato & Bekhor 2006). Starting at the trip origin, the algorithm steps through each link, selects branching links from the end node of the previously chosen link, and then uses various logical constraints to select links until the destination is reached. All paths that satisfy each threshold are considered feasible routes. This technique produces an exhaustive choice set and has high levels of observed route replication (Prato & Bekhor 2006), but the computational speed of the algorithm increases exponentially, which can become challenging over large networks. There is also a lack of knowledge about the effect of the variation of threshold on the quality of the choice set (Prato, 2009).

2.2.5 Evaluation Criteria for Choice Set Generation

Each method used to generate a choice set has merits and pitfalls. It is agreed that choice sets should aim to include only relevant and feasible alternatives, and that they have some theoretical similarity to the route set considerations of travellers. Unwanted alternatives in choice set can be characterised in three ways:

- *Overlapping*: routes that are too similar and cannot be considered as separate alternative routes;
- *Unattractive*: routes that contains more negative and less positive attributes, reverse of the trade-off of road attributes found in SP surveys;
- *Inefficient:* routes that have a lower amount of positive attributes compared to routes in the choice set with similar same negatives attributes (Erghott et al. (2012).

The degree of route replication of the observed route and partial overlap of the observed route are considered to be important route choice set criteria. Other criteria for choice set generation include: choice set size, sufficient and unbiased distribution of costs and attributes relative to observed dataset, reasonableness of the model's parameter estimation, computational efficiency, and the accuracy of predicting the chosen route for a separate validation set. The choice set's replication of the observed route indicates whether the generation technique captures the observed behaviour and whether its routes are feasible and part of the considered set by the traveler. The number of routes in the choice set is a poor criterion, as a small number of routes may not contain all feasible routes, while a large choice set may contain too many routes that are irrelevant or overly similar.

The variation of attributes in the choice set will impact the sign, value and significance of parameter estimates, and this provides the impetus for the calibrated labeling technique proposed by Broach et al., which is also the method used in this study. When testing the choice set generation techniques for our cycling study, the number of alternatives and the distribution of attributes in the choice set alternatives was found to vary the model goodness of fit and parameter estimation markedly. This is discussed further in Chapter 4. Often, the reasonableness of parameter estimates increased with the number of well-distributed alternative routes in the choice set, but this reduced the model's goodness of fit statistic. Models with poor alternatives had high goodness of fit measures, but the parameters had incorrect signs and values.

A few empirical studies have compared the results of the choice set on the model estimation results (Bovy 2009, Bekhor et al. 2006, Broach et al., 2010). These studies compared the choice set's reproduction of the observed route path, and in some cases, the results of the choice model's parameter coefficients. However, further research is needed to develop metrics that can evaluate whether the choice set generation method produces a complete and unbiased set of the considered choice set (Broach et al., 2010). Furthermore, most comparative choice set studies are tested on auto travel, which has a much simpler network cost, whereas the multiplicity of generalized cost factors and the heterogeneity in the preference for each trip attribute means that the choice set generation for cyclists requires a different set of criteria.

2.2.6 Choice Model Estimation

Route choice can be modelled using the multinomial logit model, provided that a correction factor is used to account for the correlation of the error distribution cause by the overlap of paths in the choice set. The C-logit and Path Size logit (PSL) model provide two correction terms to the MNL formulation for route choice problems. The PSL model was first proposed by Ben-Akiva et al. (1984), and is the most commonly used model formulation because the utility formulation is simpler and it can easily be estimated using freely available software with relative computation efficiency. A comparative study by Bekhor et al. (2006) shows that PSL and CNL models have similar goodness of fit measures, but the latter has a more complex model formulation and takes significantly longer time to estimate.

Other types of discrete choice models that relax the IIA condition, such a mixed logit, probit, and GEV models have been considered for this problem. The probit model has been proposed for use in route choice, but it requires simulation and a complex covariance structure to be specified for a verge large choice set and is challenging to use in practical size problems (Prato, 2009). Generalized Extreme Value models such as the Nested and Cross-Nested Logit model have also been proposed, but have not yet been successfully used for large route choice studies, as the estimation is too complex and unwieldy (Frejinger and Blerlaire, 2007).

Chapter 3

Study Data

3.1 Description of Cycling Data

This study uses data collected from a smartphone app that allows cyclist to record detailed cycling trips. The Toronto Cycling smartphone app was commissioned by the City of Toronto and developed by Brisk Synergies. The app can be downloaded for free from the Apple and Android app stores[1]. The GPS trip recorder is manually activated by the app user, and at the end of each trip users are prompted to add a trip purpose and notes about infrastructure, weather or safety concerns. Cyclist can complete an optional demographic profile and cycling aptitude survey. The data collected from this app is used by city planners to map travel patterns of cycling in Toronto, which provides a much richer account of cycling than statistics from the TTS or street bike counts at a small number of locations.

The full GPS dataset contains trips recorded over 32 days, including 5713 trips and 554 unique users. The dataset provided by the City for this study includes GPS points, trip data and user survey responses for trips recorded between August 23 - September 23, 2015. Data points for each trip are trimmed by 30 seconds on either end of the trip for user privacy, and all user information is anonymized. The tables of GPS data points recorded include the following information:

- Trip ID and User ID
- Time recorded, in the UTC time zone
- Longitude, latitude, altitude
- Horizontal and vertical accuracy of the GPS point measurement
- Travel speed, recorded by the phone
- Trip purpose
- User responses to survey on demographics and cycling experience

---

[1] City of Toronto. Toronto Cycling App. Retrieved August 2016: http://www1.toronto.ca/wps/portal/contentonly?vgnextoid=5c555cb1e7506410VgnVCM10000071d60f89RCRD&vgnextchannel=6f65970aa08c1410VgnVCM10000071d60f89RCRD&appInstanceName=default

3.2 Description of the Network Dataset

The second dataset for route choice modelling is a high resolution digital road network map that can be used in GIS software packages such as ArcGIS Desktop. These network attributes are used to describe in more details the characteristics of each cycling trip and the attributes are also used to generate a choice set of alternative routes.

The road network's attributes come from several data sources, primarily the City of Toronto's Open Data website, including the road centreline file, a list of signalized intersections, and transit route maps from Toronto's General Transit Feed Specification (GTFS) data. The UofT Library provided digital elevation model files to estimate the gradient change over network links, and simulated traffic volumes for major and minor arterials are obtained from the GTAModel V4.0 regional travel demand model. The following section describes each road attribute data set included in this study's road network.

3.2.1 Road Centreline and Intersections

The study uses a detailed road network shapefile created by the City of Toronto, which contains all roads and laneways, as well as many footpaths and trails in public parks. The network has 38,791 nodes and 53,765 links with a total length of 6,446.3 km. Each road link contains a series of fields describing the link's attributes, such as the road name, road classification, direction of travel, and cycling facility type.

Each link in the road centerline file contains the ID of the two nodes that define the beginning and end of the link. The nodes are labelled 'TO' and 'FROM', and they correspond to the direction of digitization of the link. The nodes are included as a separate point shapefiles, but this file does not have detailed information about intersection controls or geometry. A table of signalized intersection controls is joined to the network's node file based on the node ID index. Road class and cycling facility types are aggregated into three road and four bike classes for use in the model; the names of aggregated classes are in **Error! eference source not found.** and Table 3.2. Collector roads and other minor road types such as lanes and access roads are labelled as 'local' roads. These roads typically have fewer lanes, travel speeds, and traffic volumes. Figure 3.1 shows the location of signalized intersections as well as an modified road classification. Figure 3.2 maps the location of cycling facilities and typical cycling facilities in downtown Toronto are shown in Figure 3.3.

Table 3.1 Road Classification

Description	Road Class	Total length in the network (km)
Major Arterial	1	743.7
Major Arterial Ramp	1	
Minor Arterial	2	410.6
Minor Arterial Ramp	2	
Collector	3	5,293.0
Collector Ramp	3	
Local	3	
Other	3	
Laneway	3	
Pending	3	
Trail	3	
Walkway	3	
Access Road	3	

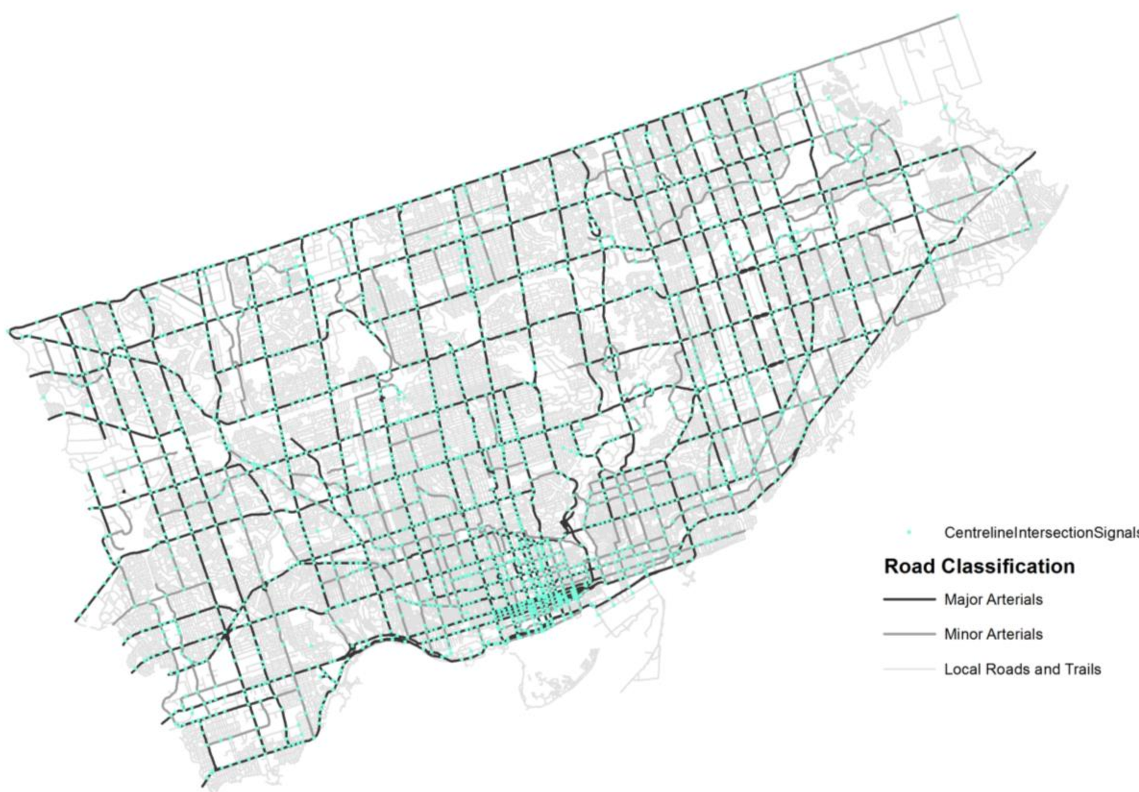


Figure 3.1 Road Classification and Signalized Intersections in Road Centreline Network

Table 3.2 Cycling Facility Classification

Facility Description	Bike class	Total length in the network (km)
Suggested On-Street Routes	4	141.6
Suggested On-Street Connections	4	71.5
Signed Routes	4	126.0
Sharrows	4	11.7
Park Roads Cycling Connections	3	28.2
Minor Multi-use Pathway	3	3.8
Major Multi-use Pathway	3	186.9
Informal Dirt Footpath	3	4.4
Cycle Tracks	2	5.5
Contra-Flow Bike Lanes	1	5.2
Bike Lanes	1	111.6
None	0	5,607.6

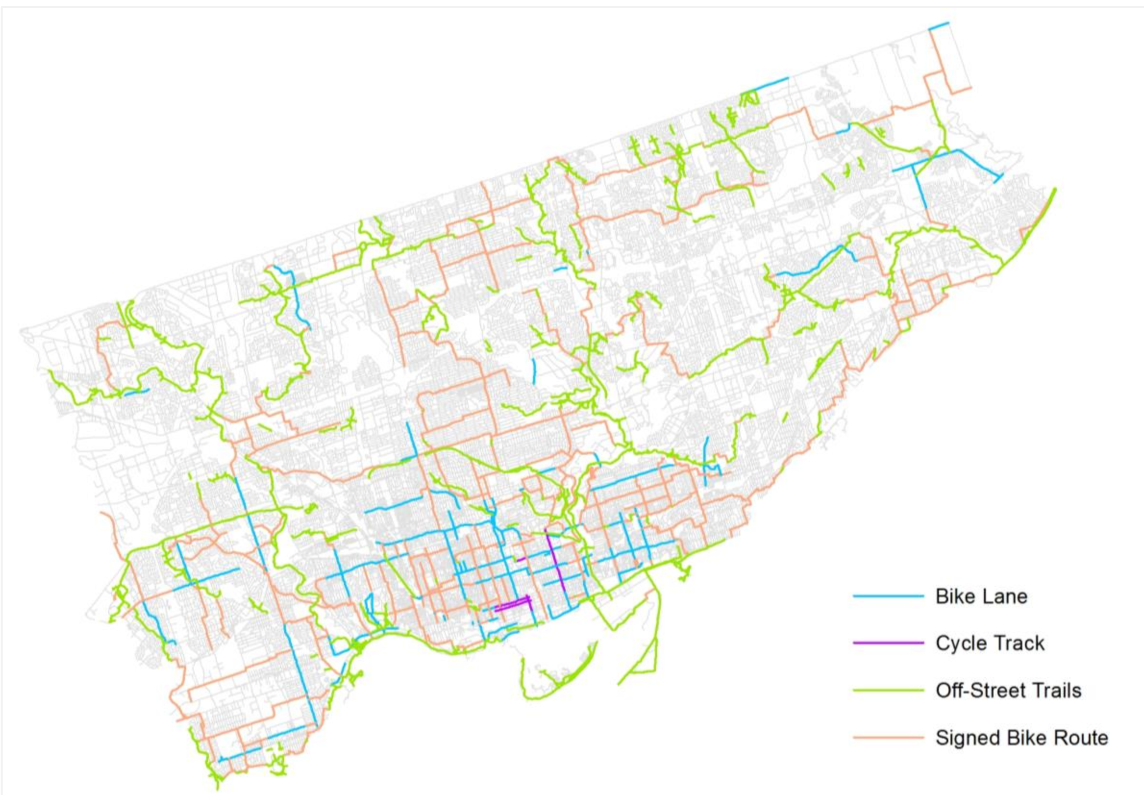


Figure 3.2 Cycling Facilities

Figure 3.3 Images of Typical Bike Facilities[2]

Bike Lanes

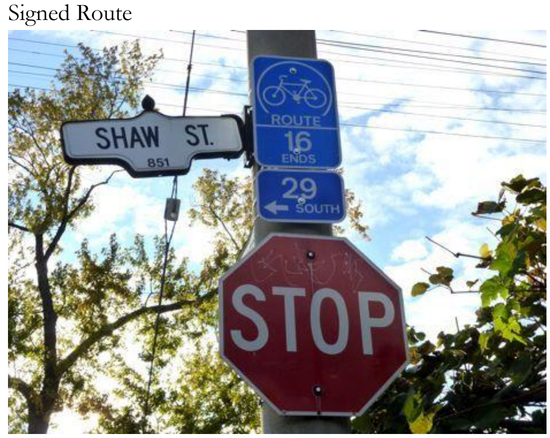
Off-Road Trail

Cycle Tracks

Sharrows

Contra Flow Lane

Signed Route



[2] Photo references, clockwise from Top left: maps.google.com; Waterfront Toronto, newblueedge.ca; Martin Reis, flickr.com; Lloyd Alter, treehugger.com; Corbin Smith, torontoist.com; minimumgrid.ca.

3.2.2 Additional Network Data

Streetcar Tracks

Streetcar tracks are a significant barrier for cycling in downtown Toronto, and they have been found to be a direct factor in 32% of all cycling injuries in Toronto (Teschke et al., 2016). The location of tracks can be found from a shapefile of streetcar routes in the GTFS files provided by the City. There are several unused streetcars tracks left in roadways not included in the GTFS files. These are manually coded into the network, using a map posted by Transit Toronto as a reference (Fomin, 2012). The location of streetcar tracks are shown in Figure 3.4.

Figure 3.2 Location of Streetcar Tracks

Road Grade

The grade of each road link is calculated as a change in elevation between the lilnk's nodes. This simple method should work for Toronto's typography, which is quite flat in the areas with the highest bike share. The average link length is 150m, but longer road segments may have elevation changes not captured in this simple approach. The elevation of each node is spatially joined from the 2005 digital elevation model (DEM) raster files for the GTA, published by the Ontario Ministry of Natural Resources and obtained from the UofT Data Library. Figure 3.3 shows the gradient of links in the direction of digitization. Roads with a grade of 4% or more are due to over/underpasses or are found around the city's ravines and river valleys. The elevations of nodes of certain roads and trails are incorrect if a bridge passes over them, as only the higher elevation point is recorded in the DEM. These node elevations are corrected manually, based on the elevations of surrounding nodes. Link gradients are capped at ±10%, because a higher gradient is most likely an error that was not captured in the manual editing.

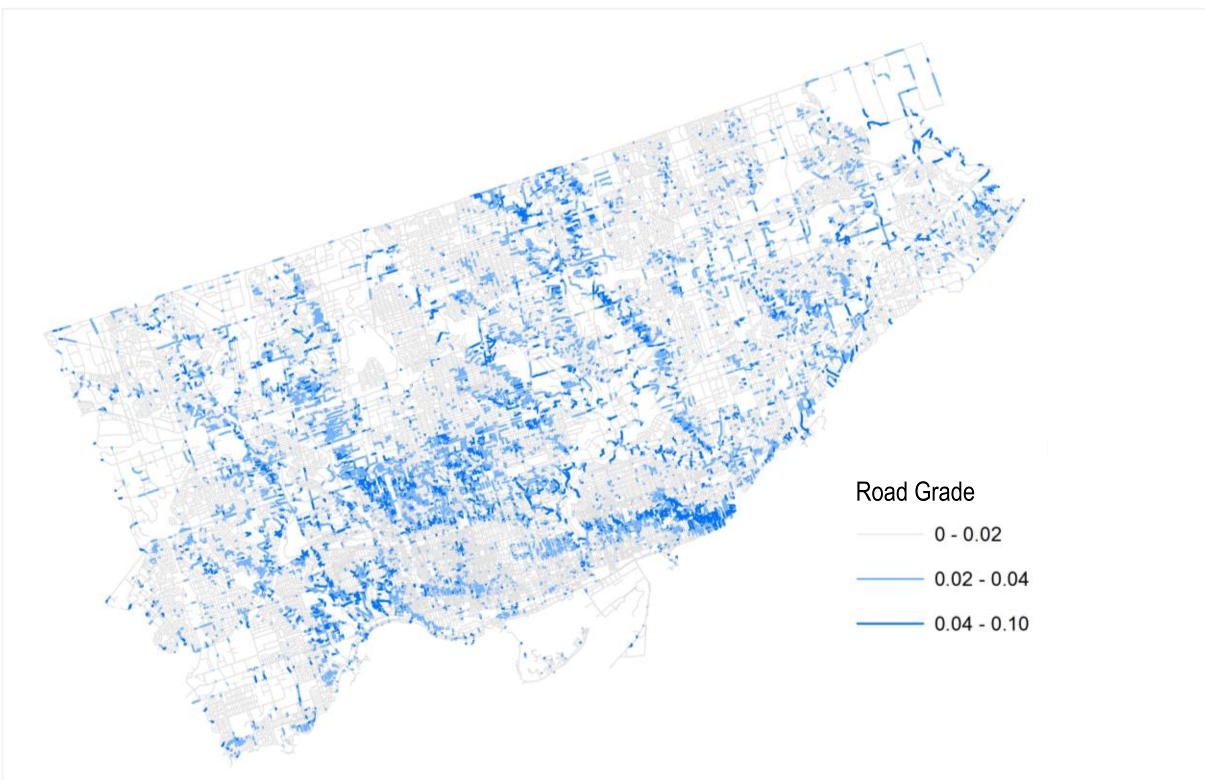


Figure 3.3 Road Grade

Traffic Volume

Comprehensive traffic volume counts of the road network in Toronto are not available, so a regional traffic assignment model is used to estimate auto volumes. UofT's GTAModel V4.0 is used to run a regional auto assignment for one day (Travel Modelling Group, 2015). The model is calibrated to 2011 Transportation Tomorrow Survey data. Traffic assignments are performed using the EMME network modelling package. The road network used by the traffic assignment does not include every road. Rather, it is comprised of a network of arterial roads and some collector roads, as well as centroid connector links. Trips start and end from the centroid of the traffic analysis zones and are linked to the network using centroid connector links, but these do not correspond to roads on the network and so are removed for this analysis.

The output of the assignment includes a table with peak hourly auto volumes of four time periods: Morning [6:00-9:00 (3hr)], Midday [9:00-15:00 (6hr)], Afternoon [15:00-19:00 (4hr)] and Evening [19:00-24:00 (5hr)]. The model does not generate volumes for the hours of 0-6am. The traffic volume for local roads not included in the EMME simulation are estimated at a rate of 1 car/min, or 1080 in total for the day (18hrs).

Because the traffic volume is estimated on a different network from the road centreline, a custom script was developed by an undergraduate research assistant, Andrew Clarry, using ArcGIS tools to transfer the attributes of one network to the other. The two main challenges in transferring attributes from the EMME network to the road centreline are that the centreline file has many more nodes and edges for the same length of road and that the EMME network links are doubled, one link for each direction of travel.

The process of matching the two networks primarily uses the mapping tool 'Transfer Attributes'. The tool is capable of matching network features with varying geometries, and in this case it is used to transfer the EMME link IDs to the centreline road network shapefile. Manual editing is required to check that all EMME links are correctly assigned to the centreline network – sometimes the tool incorrectly assigns an arterial to a parallel laneway in the centreline file. When there are two links in the EMME file (for each direction), the tool only transfers one ID. A python script is then used to determine if the

EMME ID link is with or against the direction of digitization of the centreline file. The script checks the angle of the EMME link using a straight line between the start and end nodes, and calculates the difference with the angle of the centerline's direction of digitization. If the difference of the angle is less than 90 degrees, then their directions match, and the EMME link ID is added to a field called EMME_MATCH. The script then looks for the opposite direction link in the EMME data set with the same two nodes and adds the ID to the field EMME_CONTRA. Traffic volumes for each EMME link are joined to the network from a table using the EMME link IDs. Traffic volumes on the network are shown in Figure 3.6.

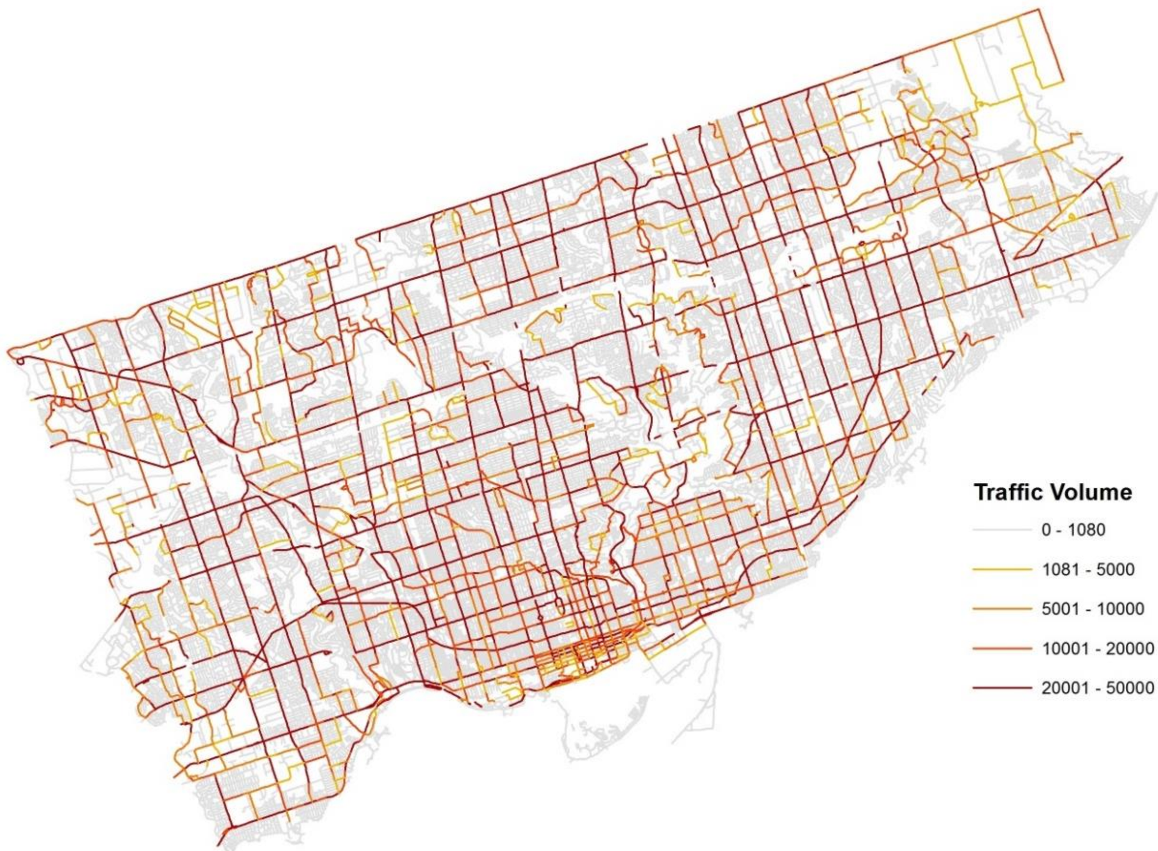


Figure 3.4 Map of Traffic Volume

3.2.3 Network Attribute as Impedances in ArcGIS Network Analyst

Once all attributes are amalgamated into the road centreline file, the shapefile is converted into a network dataset for use with the ArcGIS network analyst (NA) extension, which solves many routing problems, including a shortest path search. The network dataset creation wizard is used to specify the network attributes, such as costs and restrictions, as well as global turn delays and connectivity. The network itself can be defined with many costs (network attributes), but the shortest path search tool called 'Solve', which uses Dijkstra's algorithm, can only use one attribute, called the impedance value, at a time. The network dataset must be created within a file geodatabase, and not from a shapefile directly, for it to be correctly located by a python script.

3.2.4 Detailed Turns Classification

This study is particularly interested in the effect of intersections on route choice behaviour, and so it is important to get detailed information about the types of turns occurring within the route. The output of the route solver tool in ArcGIS is not sufficiently detailed; it only provides a feature class with a single polyline trace of the trip. This feature class can be exported into a shapefile, which can then be split by the road junctions and then joined with the features of the links. However, this does not tell us the direction of travel or the types of turns used[3]. The Network Analyst tool called "Copy Traversed Source Feature" is more useful because it generates the same shortest path as the solve tool, but it generates a solved route with the individual edges travelled along the route.

A custom python script to identify the direction of travel along those edges and the types of turns at each junction was created by Andrew Clarry. The counting turns script is used in

[3] The solve tool can be used to count the number of turns with some road classification information, if a network impedance is used as an accumulator in the solve tool and is specified with a global turn delay of 1 unit for the specific type of turn. However, the road classification and types of turns possible in the global turn delay are restrictive and it is not possible to extract more detailed information about the turns, such as the traffic volume of the cross road or presence of streetcar tracks. Other methods, such as extracting the turn type from the direction file, are often inaccurate because turns can be grouped together if they occur within a certain distance of each other.

conjunction with the map matching and choice set generation scripts. The Copy Traversed Source Feature tool is used to generate the shortest path, and then the attributes of the network links are joined to the edge feature class of the solved route using the shared Object ID. The count turns tool creates a list of the edge and nodes features and a few of the route attributes, such as link distance, road type, presence of streetcar tracks and traffic volume. The sequence of the edges is determined by sorting each link based on its cumulative distance value. To calculate the turn manoeuvre at each intersection, the tool steps through each node and calculates the geometry of the two edges associated with the node. Because edges can be curved, the turn is determined by calculating the angle of the end of the first link and beginning of the second. Once the turn type is determined (straight, left, right, reverse), the intersection features are determined. All other links connected to the node are found, including their relevant attributes. The script then counts the number of turn manoeuvre types that are of interest, such as left turns at intersection with a traffic signal, right turns at an intersection with high cross traffics, etc.

Chapter 4

Study Methods

This chapter details the study's method for cleaning and filtering the trip dataset, matching the GPS data to the road network, generating a choice set and then estimating the route choice model parameters. In brief, the process first converts the tabular GPS data into shapefiles for GIS analysis, and trips that fail a set of criteria for data quality and trip feasibility are removed. Then, the GPS traces for each trip are aligned with the study's road network in a process called map matching, using custom python scripts and network analysis tools from ArcGIS. After map matching, a set of alternative routes that will comprise the choice set for a path-size model is generated for each trips' origin and destination. The choice set uses a calibrated label repeated shortest path search that minimizes general cost formulated for route attributes, such as presence of bike lanes or avoiding hills. Finally, a python script is used to remove identical and overlapping routes in the choice set, and to generate a series of statistics about each route that can be used as variables for model estimation.

4.1 Data Cleaning

This data set is filtered for valid trips in an initial data cleaning process. Not all routes recorded by the app can be used in the model. Trips for exercise and leisure purposes are excluded from this study because they often have circular routes, and the trade-offs made by cyclists in terms of distance and directness of route are very different to those for utilitarian cyclists, which would reduce the accuracy of the model estimation.

Trips that are too short or too long in duration are also excluded, as they are most likely mis-reported by the user. A short trip under two minutes will most likely be a captive trip with a short distance that will not yield useful route preference information. A trip over two hours is most likely a recreational trip or the user has forgotten to stop recording the trip. From the initial set of 5713 trips, 358 trips are removed because they are too long or too short, leaving 5356 trips. The trip data tables are then converted to point shapefiles using a python script and two ArcGIS tools – 'Display XY points' and 'Convert from layer to shapefile'. Once the points are generated, further data cleaning is required.

In some cases, there are errors in the GPS traces that prevent a match with the network in the map matching process. Large gaps between points will fail in the map matching algorithm, because a continuous 50-metre buffer is generated around each point. If a gap of a 100 metres or more is present, the buffer will no longer be continuous. These routes are therefore removed to save processing time. A python script is used to identify trips with gaps over 100 metres between two successive GPS point, or trips with an origin or destination within 500m. This filtering process eliminates 1458 of the 5356 trips (85% had a gap in GPS points, and 15% had OD pairs within 500m). There are some instances where there is a clear trace of the route taken, but a few GPS points are erroneously located far from the path taken. These trips could be useful in the analysis, but would require an algorithm that identifies erroneous points, which was beyond the scope of this project.

4.2 Map matching the Observed Route to Network

Once the GPS traces are converted to point shapefiles and unwanted trips are removed, the points are converted into path along the network, so that each link travelled along the network and its attributes can be summarized and used as variables in the route choice model. A growing field of map matching refers to algorithms that align GPS points onto a reference road network, approximating the most likely path that the GPS trace of points represent (Schuessler & Axhausen, 2009).

The simplest method, the geometric procedure, locates the closest roads to each GPS points. The final path is created by adding all selected road links in closest proximity to the GPS points. However, when the GPS points approach a cross-street and are closer in distance to it than the road travelled, the cross-streets is incorrectly added to the route, and looks like the person briefly turned onto the street before turning back. Topographical methods use algorithms to avoid these errors, using various rules to recreate the route path. Most topological methods use a geometric approach to find an initial link or point on the network, and then sequentially build the rest of the route by choosing from a set of candidate links.

The method used for this study was developed by Ron Dalumpines and Darren Scott at McMaster University (Dalumpines & Scott, 2015). It utilizes a few simple tools in ArcGIS, rather than complex topographical algorithms. The approach first creates a buffer around the GPS points to define a boundary within which travel is most likely to occur. The first and last points of the trip are then used as the start and end location for a shortest path search within the buffer area.

The size of the buffer strongly influences the accuracy of the trip, as well as the success rate of the algorithm. If the buffer is smaller than the spatial accuracy of the GPS traces, then the road network will may fall outside of the buffer area, but if the buffer is too large, then several alternative routes within the buffer area may be possible, and the shortest path algorithm may find a different path that the one chosen. A buffer of 100 metres or more has a much higher likely hood of incorrectly retracing the route path so the final buffer size of 50 metres was chosen. The distance between most streets is larger than 50 metres, which minimizes the number of alternative routes possible. The highest accuracy in path reproduction was more important than including a high percentage of trips because the sample was quite large and the analysis is interested in small details of route selection. A sample route with a 50-metre buffer is shown in Figure 4.1. The blue dots represent GPS points and the pink line represents the solved route.

Figure 4.1 Example of a Map Matched Route with 50 meter Buffer

Using a 50-metre buffer, the map matching process successfully identified a route for 54% of trips, leaving 2009 trips in total and 350 unique users, with 1709 trips for the model estimation and 300 for verification. The high proportion of failed routes during the map matching is due to limitations of the method as well as errors in the GPS data. Most points recorded have a reported horizontal accuracy of 5-10 metres, but a few points have a higher spatial inaccuracy of 30 metres or more. The map matching algorithm will not find a good route if the trace of GPS point deviates more than 50 metres from the network at just one location, either due to signal inaccuracy or if the person takes a shortcut through a link not included in the centreline network map, such as a parking lot. Tall buildings in the downtown or trails in ravine parks can create areas with GPS signal distortion and, these locations may be under represented in the model estimation. The map matching method could be improved by adding common shortcuts to the network, and by iterating the map matching procedure to increase the buffer size until a route is found. However, iterating the map matching process would significantly increase computing times.

4.3 Choice Set Generation

4.3.1 Simple and Calibrated Route Labelling

Two choice sets were generated using the labeling technique and the difference between these two choice set and the impact on the model estimation are discussed in this section. The first choice set is generated using a simple labelling approach. The labeled approach provides a number of routes with varying sensitivities to route attributes, which reflects variation in preference for route attributes found in SP surveys. The attributes included as labels are chosen are listed in Table 4.1. Each route is generated using a generalized cost function that either reduces or increases the cost of the link if the label attribute is present on the link.

The second choice set is a modification of the labeling technique proposed by Broach et al. (2010). The calibrated labelling method generates several routes per attribute by incrementally varying the generalized cost associated with the label. The increment and range of the variation of the generalized cost function is defined so that the distribution of the routes generated for the label matched the distribution of the attribute in the observed route dataset. Calibrating labels to the observed data helps to generate a more comprehensive and balanced distribution of attributes in the choice set which is important for the proper estimation of parameters in the choice model (Broach et al., 2010).

The labelling technique can be easily programmed using GIS shortest path tool and a python script, however the larger number of route searches per OD pair ArcGIS did considerably lengthen the computational time, especially for the second choice set. CS2 generated 37 alternative routes and this took significantly longer to run. The average time for one OD pair choice in CS2 set was approximately 12 minutes, compared to 5 minutes per OD pair for the CS1. Other studies have used python based network analysis packages, such as the NetworkX python extension used by Hood et al. (2011), which may explain the dramatically faster computation time for generating the choice set, solved in hours instead of weeks. It would be worth comparing a choice set generation using the two programs.

The values used to calculate each generalized cost parameter of labels in both choice sets are shown in Table 4.1 and Table 4.2. The unit for impedances was initially chosen as metres, but was converted to minutes in order to include global turn delays which must have a unit of minutes or seconds. Travel time on network links is based on shape length divided by the average speed of all trips in the dataset (4.32 m/s).

Table 4.1 Choice Set 1: Labels and Impedances

	Label	Impedance Calculation
A1	Shortest path by time	Shape length /average travel speed (4.32 m/s)
A2	Prefer bike lanes	Reduce time by 60% for bike lanes only
A3	Prefer all bike facilities	Reduce time by 60% for cycle tracks, 50% for bike lanes, 40% for off-road trails, and 20% for signed routes and sharrows
A5	Avoid steep hills	If slope >1%: minutes*[1+(SLOPE*10)] , 10% slope has an increase of 100% travel time
A7	Avoid streetcar tracks	Travel time increases 50% is streetcar track present, 25% if there is also a bike lane.
A8	Minimize right and left turns	Left turns: 180 sec, right turns 120 sec
A9	Minimize signalized intersections	Add 30 seconds for each intersection (point barrier with scaled cost)
A10	Minimize traffic	If road traffic above 10,000/day, add 50% travel time
A11	Prefer arterials	If road is local, add 50% travel time
A12	Prefer local streets	If road is major arterial, add 50% travel time

Table 4.2 Choice Set 2: Labels and Impedances

	Label	Impedance Calculation
A1	Shortest path	Shape length (metres)
A2-4	Bike lane	If bike lane, minutes *(0.4, 0.6, 0.8)
A5-7	Cycle tracks	If cycle tracks, minutes *(0.4, 0.6, 0.8)
A8-10	Off-road trail	If off-road trail, minutes *(0.4, 0.6, 0.8)
A11-13	Bike route	If bike route, minutes *(0.4, 0.6, 0.8)
A14-16	Total bike facilities	If any bike facility, minutes *(0.4, 0.6, 0.8)
A17-19	Local road	If local road, minutes *(0.4, 0.6, 0.8)
A20-22	Minor arterial	If minor arterial road, minutes *(0.4, 0.6, 0.8)
A23-25	Major arterial	If major arterial road, minutes *(0.4, 0.6, 0.8)
A26-29	Traffic	If traffic higher than 10,000/day, minutes*(0.4, 0.6, 0.8)
A29-31	Slope	If slope >1%: minutes*[1+(SLOPE*Beta)] , where Beta: 50, 150, 200
A32-34	Turns	Global turn delay: left (60, 90, 120 sec) & right (30, 60, 90 sec)
A35-36	Left	Global turn delay: left (60, 120 ses)
A37	Right	Global turn delay: right (30 sec)

The impedance of positive grades is calculated by linearly increasing the travel time cost relative to the gradient. An analysis of cycling speed and road gradient found that the reduction in speed for positive gradients was about -50% at a 10% grade, and that this effect was greater that the time savings for a negative gradient, which are only about +30% at a 10% downhill (Parkin and Rotheram, 2010). However, when the time savings of downhill are included in the impedance, the labelled route does significantly avoid large hills compared to other routes in the choice set. Therefore, the network's impedance for the slope label only includes a higher cost for travelling up hills.

Traffic volume of a road link is measured in two ways to account for differences in cyclist's perception of traffic volume. The first measure calculates the percentage of the route on roads within a range of total daily traffic volume, and the second measure is an estimate the cumulative number of vehicles encountered along the route. Total daily traffic volume on a road segment is calculated by adding the time period's average hourly volume multiplied by the hours for each time period. Total daily travel in each direction is combined to simplify analysis. The model will account for high traffic based on the percentage of the route travelled on road segments with two daily traffic volume thresholds: 10000+ or 20000+

vehicles per day. The second measure is an accumulation of traffic exposure along the route that counts the expected number of the number of vehicles encountered along the route. The measure is calculated by multiplying the traffic volume rate of the road link by the travel time of the link, and then adding the traffic volume for all links.

4.3.2 Two Choice Set Generation Techniques

The amount by which a generalized cost function is modified means the route generated for each label will deviate more or less from the shortest path. This variation influences the composition of the chosen route, and creates bias in the model. If only unattractive routes are included in the choice set, the estimated model may have a better fit, but the model's attribute parameters will not accurately estimate the utility of attributes on route choice. In order to estimate better attribute parameters, it is important to generate a set of feasible and attractive alternatives for consideration in the choice set, yet it is not always easy to verify. The variation of the attributes from the first choice set (CS1), shown in **Error! Reference ource not found.**, are not dramatically larger or smaller compared to the observed routes or CS2. Yet, the parameters estimated using this model are significantly different from CS2 – most notably, distance has a positive utility coefficient.

To improve the results, labels that generate inefficient and unattractive routes are removed from the choice set to improve the distribution of attributes in the choice set, which results in more realistic model parameters. In CS1, Routes 2 and 4 are removed because they generated less attractive average attribute compared to Route 3 using the same label. Routes 6, 7, 9, 10, and 11 were removed because they were inefficient: they are longer than the shortest route, but did not increase positive attributes of the route. The reduced choice set generated model parameters with expected sign and level of significance. However, the level of researcher bias in filtering un-attractive and in-efficient routes from the choice set was not satisfactory. Broach et al. (2010) also found poor model parameters with the labelling approach and proposed calibrating the labelling approach to align the distribution of attributes in the choice set with the observed routes. The route choice generation approach of this study adapts the calibration principle outlined by Broach et al. (2010). The labels of the second choice set (CS2) are summarised in Table 4.2, and the average of attributes from CS2 are compared to CS1 in Table 4.3.

Table 4.3 Attributes of Routes in CS1 and CS2

	Observed Path	Shortest Path	Alternatives in CS1 (Full Set)	Alternatives in CS1 (Selected Set)	Alternatives in CS2
# Unique alternatives	1709	1491	13796	7517	22567
Average Length	6497	6182	6814	6980	7223
Detour	10%	0%	8%	10%	9%
All bike facilities	53%	33%	43%	43%	45%
Bike lane	28%	17%	23%	26%	20%
Bike route	16%	12%	13%	11%	18%
Cycle track	5%	3%	4%	4%	3%
Bike trail	8%	5%	7%	5%	8%
Local street	37%	35%	41%	38%	38%
Minor arterial	30%	22%	27%	31%	25%
Major arterial	34%	44%	32%	31%	37%
Minor arterial, no bike lane	8%	11%	9%	10%	9%
Major arterial, no bike lane	23%	34%	23%	22%	26%
Streetcar tracks	18%	22%	16%	16%	19%
HILL 1-3%	13%	14%	14%	14%	14%
HILL +3%	3%	3%	3%	3%	3%
HILL 2-4%	4%	5%	5%	5%	5%
HILL 4-6%	1%	1%	1%	1%	1%
HILL +6%	0%	1%	1%	1%	1%

Table 4.3 (Con't) Attributes of Routes in CS1 and CS2

	Observed Path	Shortest Path	Alternatives in CS1 (Full Set)	Alternatives in CS1 (Selected Set)	Alternatives in CS2
Traffic over 5,000 veh/day	58%	61%	52%	54%	57%
Traffic over 10,000 veh/day	49%	54%	45%	46%	50%
Traffic over 20,000 veh/day	30%	36%	28%	28%	32%
Total traffic volume	34.5	40.4	33.2	33.2	37.3
Traffic lights	13.9	14.3	13.9	14.6	15.6
Turns per km	1.8	2.3	2.5	2.3	2.1
Right turns	5.3	5.8	7.6	7.7	6.9
Left turns	5.1	5.9	7.3	7.0	6.6
Left Turn, Unsignalized, Across Arterial	0.9	1.2	1.4	1.3	1.3
Right Turn, Unsignalized, Across Arterial	1.7	2.0	2.4	2.5	2.2
Straight, Unsignalized, Across Arterial	1.6	2.2	2.2	2.3	2.5

4.4 Route Choice Model

The path size logit (PSL) model is estimated using a freely available logit estimation software called Biogeme (Bierlaire, 2003). Version 2.2 was used for estimation and version 2.5 was used to test panel data. The model formulation used for this study comes from Ben-Akiva and Bierlaire (1999):

$$P_k = \frac{\exp(V_k + \beta_{PS} * lnPS_k)}{\sum_{l \in C} \exp(V_l + \beta_{PS} * lnPS_l)}$$

where P_k is the probability of choosing path k; V_k is the vector of utilities for path k (β* attribute); PS_k is the path size factor (below); and l is an alternative route in the full choice set C. The Path-Size factor is calculated by a weighted sum of each link's proportion of total route distance multiplied by a factor of overlap, which is the inverse of the frequency of the link in the choice set. The formula used for Path Size factor is specified as:

$$PS_k = \sum_{a \in \Gamma_k} \frac{L_a}{L_k} \frac{1}{\sum_{l \in C} \delta_{al}}$$

where L_a is the length of link a, which is part of route k, L_k is the total length of that route, and δ_{al} is a binary that is 1 if link a is in another route l in the choice set C.

4.4.1 Removing Identical Alternatives and Preparing for Choice Modeling

A route analysis tool uses the python pandas package to summarize the attributes of the route in the choice set and formats the data for BIOGEME. Attributes tables of each route shapefile is exported to a csv format are imported as data frame objects. This tool calculates the proportion of attributes for each route, such as percent of route on a bike lane or the traffic accumulated along the route. The tool also identifies and removes duplicate routes, which is important to satisfy the IIA property of the MNL model. Each route is compared to all other routes to identify common links, and the distance of these overlapping links is counted. If the route overlaps another route in the choice set by 90 percent or more, they are considered too similar and one of the two routes is removed from the choice set.

4.4.2 Modeling Repeat Users and Interaction of User Characteristics

The heterogeneity of attribute preference based on trip purpose, time of day and user characteristics are modeled by adding interaction terms. The interaction of a dummy variable for a trip or user characteristics and a particular route attribute is an iterative process and combinations are chosen to reflect interactions found in previous SP and RP surveys.

The inclusion of repeat observations from one user, called panel data, can potentially cause a bias in the parameter generation. The panel data effect of the dataset was modeled in BIOGEME by adding the following term to the utility formulation:

$$\beta_1 * x_1 + \ldots + \text{zero [SIGMA*one]}$$

where $\beta_1 * x_1$ are the road attribute parameters, and *zero* and *SIGMA* are parameters to be estimated, with means fixed at 0 and 1 respectively, and *one* is a constant fixed at a value of 1. The parameters *zero* and *SIGMA* are not significant in a test of 500 trips, and so the panel data was not included in the model estimation using the full dataset.

Chapter 5

Results and Discussion

This chapter will review the trip data and user characteristics, choice sets generated and the results of the route choice model estimation. The results of the model estimation and the model verification are compared to the results of previous route choice studies.

5.1 Description of Cycling App Data Set

5.1.1 User Description

The cycling app survey about demographics and cycling attitudes is used to identify significant differences of route preference with respect to various user groups. A full summary of survey responses are found in Table 5.1 and Table 5.2. The app users that responded to the survey are most likely to be between the ages of 25-49, female, and have a household income of over $100,000. App users are also relatively experienced and comfortable cyclists: 70% of users feel comfortable riding in most roadway conditions and 26% do not.

The survey response rate was varied; questions about age, income, cycling comfort and winter cycling have a high response rate around 90-95%, but gender, cycling frequency and rider history had response rates of (30-55%). It is unclear why the half the app users in the sample did not respond to gender, and it is interesting to note that 66% of those who responded are female, which is the inverse of the gender proportion in the TTS. An analysis of cycling speed shows that those who did not report gender are faster cyclists. It is possible that the non-responders are male, but it is not clear why they would not report their gender. Whether women in the sample are more likely to respond to the gender question, or if the sample of users contains more women is also unclear. For either reason, it may be that women more vulnerable and are more driven to participate in an infrastructure planning exercise and want to communicate to City staff that they choose paths that they find safer.

Trips made by female respondents have longer detours from the shortest path and they travel on a higher proportion of bikes lanes and cycle tracks.

It is also interesting to note that while two thirds of trips are reported as commuting in Table 3.4, 21% report cycling twice a week or more and 31% of users report a cycling frequency less than once a month. It is possible that the 46% that did not respond to the question of the cycling frequency are indeed more frequent cyclists who commute, which if added to those who report cycling frequently, would correspond the two thirds proportion of cycling trips.

Table 5.1 Demographic of Cycling App Trips and of 2011 TTS Cycling Trips

User Demographics	Trips in estimation dataset		2011 TTS
Gender			
“Male"	284	67%	17%
"Female"	618	33%	36%
N/A	807		47%
Age			
"under 18"	0	6%	0%
"18-24"	47	11%	3%
"25-34"	647	22%	38%
"35-49"	619	34%	36%
"50-64"	306	23%	18%
"65+"	40	4%	2%
N/A	50		3%
Income			
“under $20,000"	55	N/A	3%
"$20-39,000"	97	N/A	6%
"$40-59,000"	141	N/A	8%
"$60-79,000"	180	N/A	11%
"$80-99,000"	139	N/A	8%
"$100 or greater"	816	N/A	48%
N/A	281	N/A	16%

Table 5.2 Experience and Attitudes of Cyclists in Estimation Dataset

User Demographics	Responses by trip	
Cycling Frequency		
"Less than once a month"	537	31%
"Several times a month"	80	5%
"Several times per week"	320	19%
"Daily"	33	2%
N/A	739	43%
Cycling Comfort		
"Not comfortable sharing roadways with vehicles."	137	8%
"Only comfortable sharing roadways in clearly designated cycling facilities."	305	18%
"Comfortable cycling in most roadway conditions."	1179	69%
N/A	88	5%
Rider history		
"Since childhood"	437	26%
"Several years"	161	9%
"One year or less"	46	3%
"Just trying it out/ just started"	6	0%
N/A	1059	62%
Winter		
"not a winter cyclist"	898	53%
"a winter cyclist"	699	41%
N/A	112	7%

5.1.2 Trip Description

The average trip distance is 6.4 km and commuting is the most common trip purpose, with 67% of the 2009 trips included in the data set for model estimation. The longer trips tend to be for commuting - trips over 15km are all commute-related – see Figure 5.1 for a distribution of trip lengths and see Table 5.3 for trip purpose distribution. Cycling trips in the 2011 TTS have shorter distances and a smaller percentage of commute trips. Trips recorded in the app have an average duration of 26 minutes and average speed of 12.2 km/hr, and the most common travel period is between 6-9am, during the morning peak, shown in Figure 5.2. The average number of trips recorded per user is 5.5, and is shown in Figure 5.3.

Table 5.3 Trip Purpose

Commute	1348	67%
Errand	50	2%
Home	102	5%
Other	149	7%
School	40	2%
Shopping	97	5%
Social	142	7%
Work-Related	80	4%

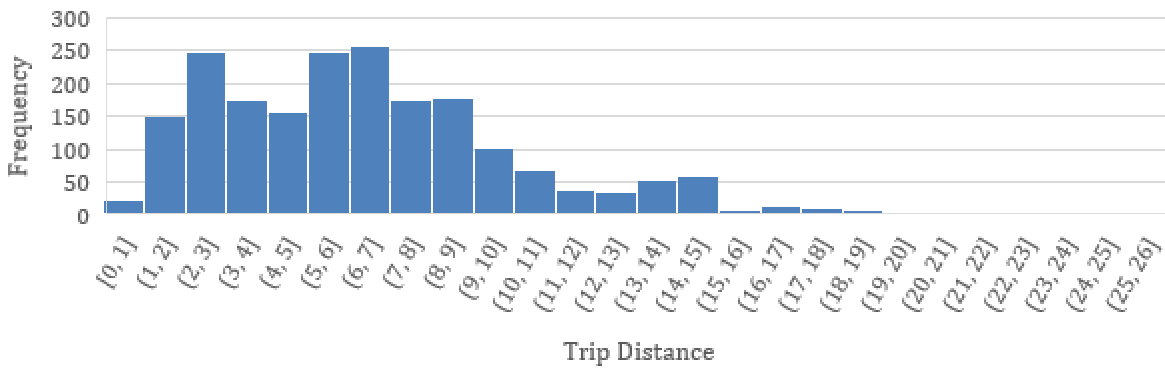


Figure 5.1 Distribution of Trip Distance, in kilometers

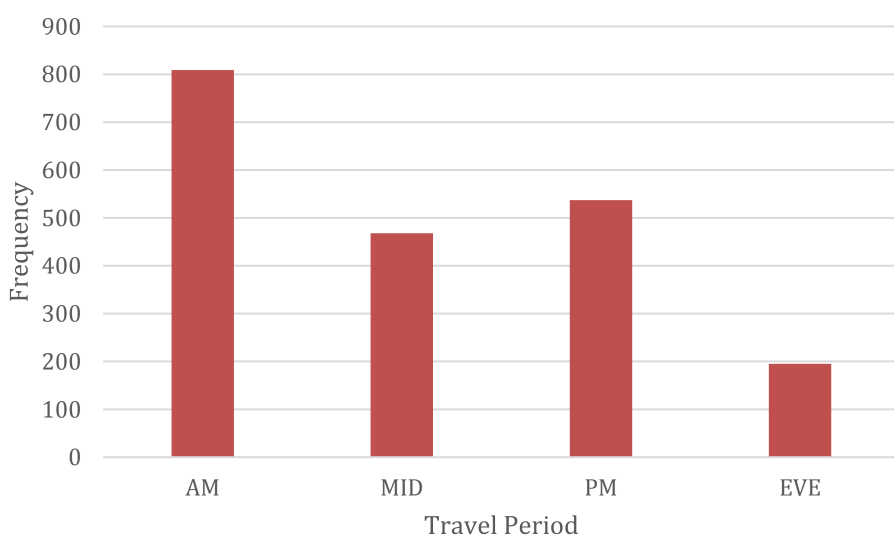


Figure 5.2 Trip Start Time By Traffic Period
AM (6-9am), MID (9am-3pm), PM (3-7pm), EVE (7pm-12am)

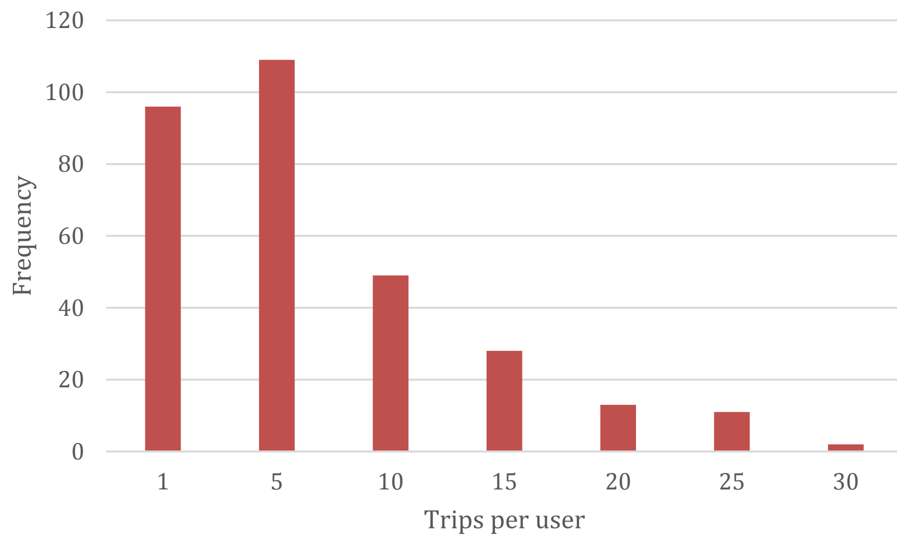


Figure 5.3 Trips per User

5.1.3 Trip Attribute Description

The GPS points in the data set are converted into route paths with the attributes of the road network in a map matching process described in Chapter 4. The map matching process generates the shortest path from the trip's origin and destination, and a comparison of the observed route to the shortest path attributes provides an initial understanding of the trade-offs made between distance and other route attributes. Table 5.4 shows the average, mean and standard deviation of the network attributes of the observed and shortest paths. The average deviation from the shortest route by the observed routes is 10%, which is consistent with other cycling studies (Broach et al., 2012; Winters et al., 2010). On average, 22% of observed trips are the same length as the shortest path. Observed trips use significantly more bike facilities and local roads than the shortest path, spend less time on arterials roads with high traffic, have slightly fewer turns of all types and spend less time on hilly road segments.

Table 5.4 Proportion of Attributes for Observed and Shortest Trip Route

Unique trips	Observed 2132			Shortest 1863		
	Average	Median	Std Dev.	Average	Median	Std Dev.
Length (m)	6473.7	6032.8	3880.2	5851.7	5452.5	3345.2
Detour from Shortest Path	10%	6%	14%	-	-	-
Total Bike Facilities	53%	57%	27%	33%	29%	25%
Bike Lane	28%	24%	24%	17%	10%	19%
Signed Route	16%	11%	16%	12%	5%	15%
Cycle Tracks	5%	1%	9%	3%	1%	5%
Off Road Trail	8%	1%	17%	5%	1%	10%
Hill Gradient 1-3%	12.7%	9.0%	11.5%	13.6%	10.4%	11.7%
Hill Gradient >3%	2.7%	1.4%	3.6%	3.3%	1.7%	4.3%
Street Car Tracks	21%	9%	27%	27%	14%	30%
Local Road	37%	32%	27%	33%	30%	23%
Minor Arterial	29%	26%	25%	22%	16%	23%
Major Arterial	34%	27%	28%	44%	42%	28%
Traffic Volume	35	31	22	41	39	23
Traffic over 5,000 veh./day	57%	60%	26%	62%	65%	23%
Traffic over 10,000 veh./day	49%	49%	25%	55%	57%	23%
Traffic over 20,000 veh./day	30%	26%	22%	38%	36%	22%
# Traffic Light	13.7	13.0	8.3	13.7	12.0	8.8
Right Turns per km	0.9	0.8	0.6	1.1	0.9	0.8
Left Turns per km	0.9	0.8	0.5	1.1	0.9	0.8
Turns per km	1.8	1.6	1.0	2.1	1.8	1.5
Right Turns	5.4	5.0	3.5	5.3	5.0	3.6
Left Turns	5.1	4.0	3.6	5.3	5.0	3.6
Left Turns at Arterial Road, Unsignalized	0.9	1.0	1.0	1.1	1.0	1.1
Right Turns at Arterial Road, Unsignalized	1.7	1.0	1.4	1.8	1.0	1.6
Crossing Arterial Road, Unsignalized	1.7	1.0	2.0	2.0	1.0	2.3

5.1.4 Heat Maps of Travel on the Network

To better visualize the geographic scope of the data set, a network heat map is generated to show the count of trips on each road segment. The network heat map in

Figure 5.4 shows that popular routes overlap strongly with bike facilities – for example Bloor St. Viaduct and Sherbourne St. bike lane (1), Shaw St. contra flow lane (2), Davenport Rd. and Russel Hill Rd. (3), Dundas St. E. bike lane (4), College St W. (5) and the Martin Goodman Trail (6).

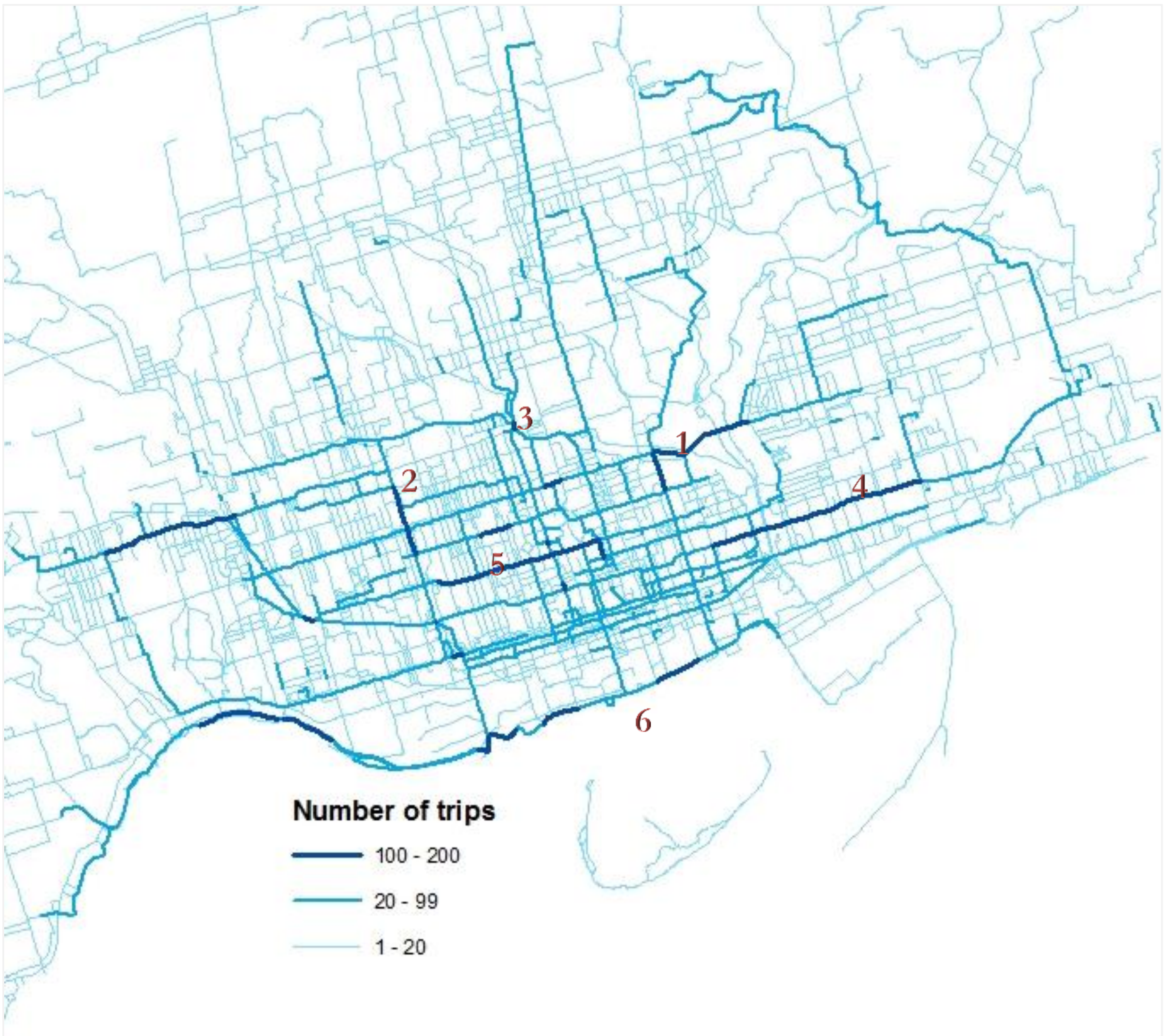


Figure 5.4 Heat Map of Trips Included in the Study

Figure 5.5 is a variation of the network heat map in Figure 5.4, but roads without cycling facilities but high numbers of cycling trips are shown in orange to contrast with street with bike lanes in blue. The thicker the line weight, the more trips are on that road link. Thicker orange lines show areas where cyclists are frequently using streets without cycling infrastructure, and could be compared to 'desire lines' that pedestrian paths carve out on grassy lawns that highlight a mismatch between infrastructure and behaviour.

Figure 5.5 Heat Map of Trip with and without Cycling Facilities

5.2 Evaluation of the Choice Set Generation

The study's two choice sets are compared in this section to better understand the variation in choice set composition due to the generation methods and parameters. The calibrated choice set CS2 generated better model parameters and is used for the estimation of Models 1 and 2. Previous studies of choice sets for route choice modeling have evaluated the quality of choice set by the extent to which: they reproduce the observed route path, the distribution of attributes that corresponds with the observed behaviour, and the quality of parameters coefficients in the model estimation.

5.2.1 Choice Set Size

After the removal of duplicate trips, the average choice set size is 5.3 alternatives (out of 6) for CS1, and 14.2 alternatives (out of 37) for CS2 – see Figure 5.6. The variation of choice set size by length of the observed route for CS2 is shown in Figure 5.7. The choice set size is reduced as the trip length decreases, as the likelihood of route duplicates increases.

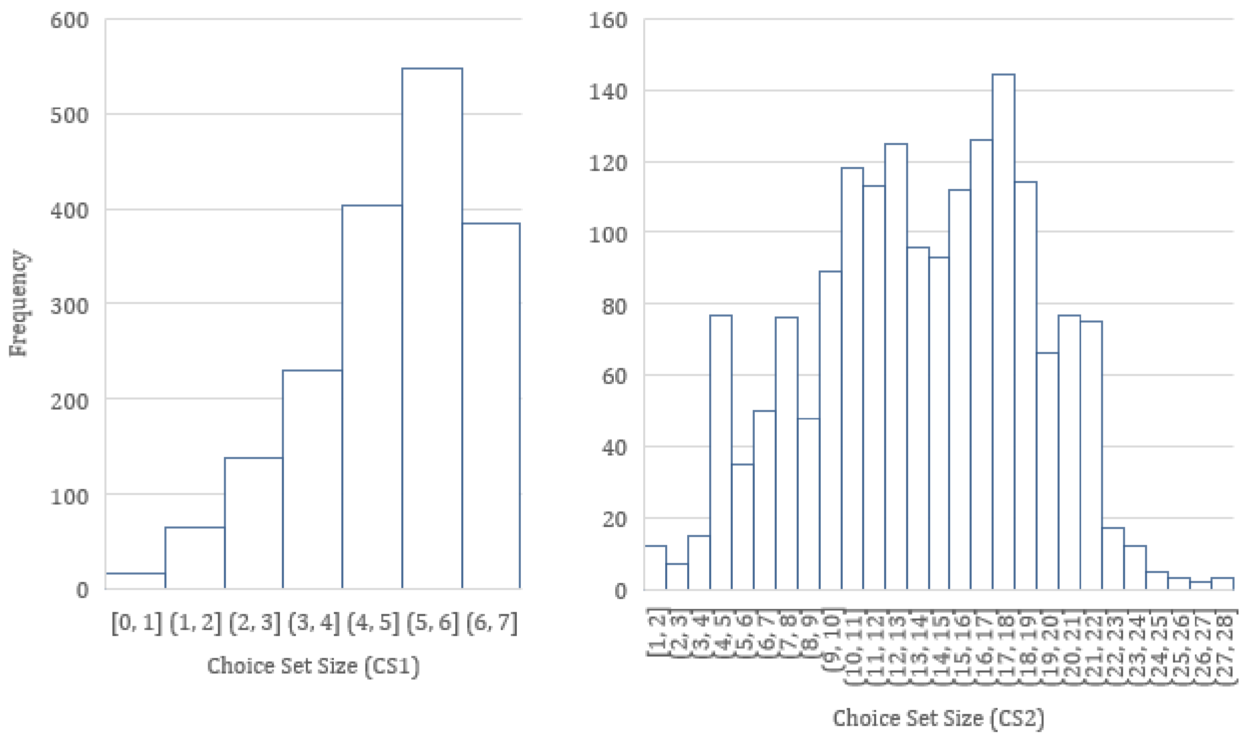


Figure 5.6 Distribution of Choice Set Size, CS 1 & 2

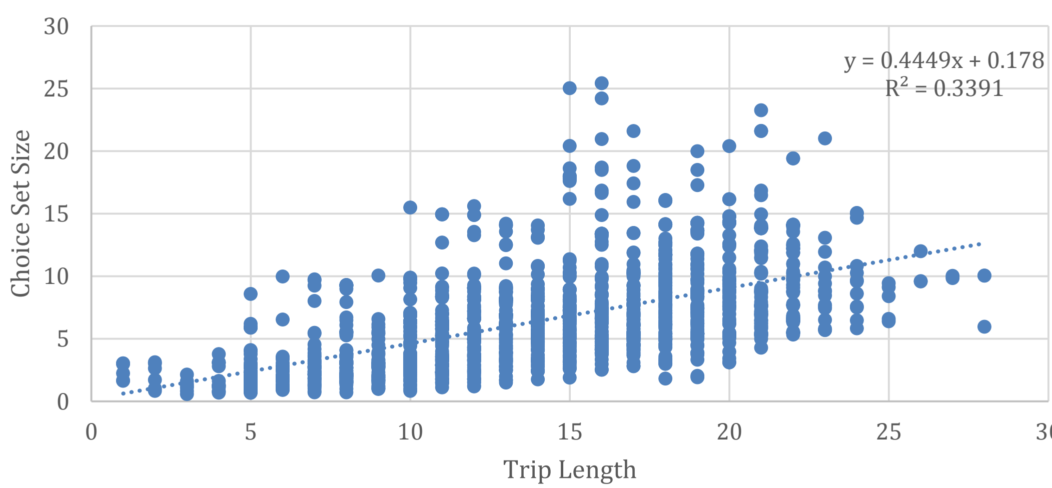


Figure 5.7 Size of Choice Set for each OD Pair by Trip Length (CS2)

5.2.2 Observed Route Path Reproduction

A choice set for a route choice model should be able to reproduce the observed route path. The cumulative distribution of the reproduction of the observed path by CS1 and CS2 are found in Table 5.5. CS2 is slightly better able to completely reproduce the observed route, with 21% exact replication compared to 18% in the simple labeled route. The calibrated labeling choice set by Broach et al. (2010) was able to completely reproduce the observed route in 22.5% of cases, and by 29.4% and 42.3% for the 90 and 80 percent thresholds.

Table 5.5 Overlap of Observed Routes in Choice Sets CS1 and CS2

Choice Set 1			Choice Set 2		
Overlap threshold	**Frequency**	**Cumulative**	**Overlap Threshold**	**Frequency**	**Cumulative**
100%	18%	**18%**	100%	21%	**21%**
90-99%	7%	**25%**	90-99%	11%	**32%**
80-89%	13%	**38%**	80-89%	15%	**47%**
70-79%	15%	**53%**	70-79%	12%	**59%**
60-69%	12%	66%	60-69%	13%	72%
50-59%	12%	78%	50-59%	10%	82%
40-49%	8%	86%	40-49%	9%	91%
30-39%	5%	91%	30-39%	5%	95%
20-29%	5%	95%	20-29%	3%	98%
10-19%	3%	98%	10-19%	1%	100%
0-9%	1%	100%	0-9%	0%	100%

The alternative routes in CS2 with the best overlap the observed route in the choice set are shown in Table 5.6. Alternative A15, which reduced the travel cost for all roads with bike infrastructure by 40%, best approximates the observed route 17.6% of the time.

Table 5.6 Alternative that Best Reproduce Observed Route

Alt	Frequency of Alternative being higher Overlap with observed		Average length of the alternative when max overlap
A15	300	17.6%	8291
A1	231	13.5%	3814
A2	147	8.6%	5082
A22	115	6.7%	6407
A16	103	6.0%	7168

As can be expected, the amount of overlap of the choice set with the observed route increases for shorter trips, because the number of overlapping alternatives in the choice set increases as the distance decreases, as shown in Figure 5.8.

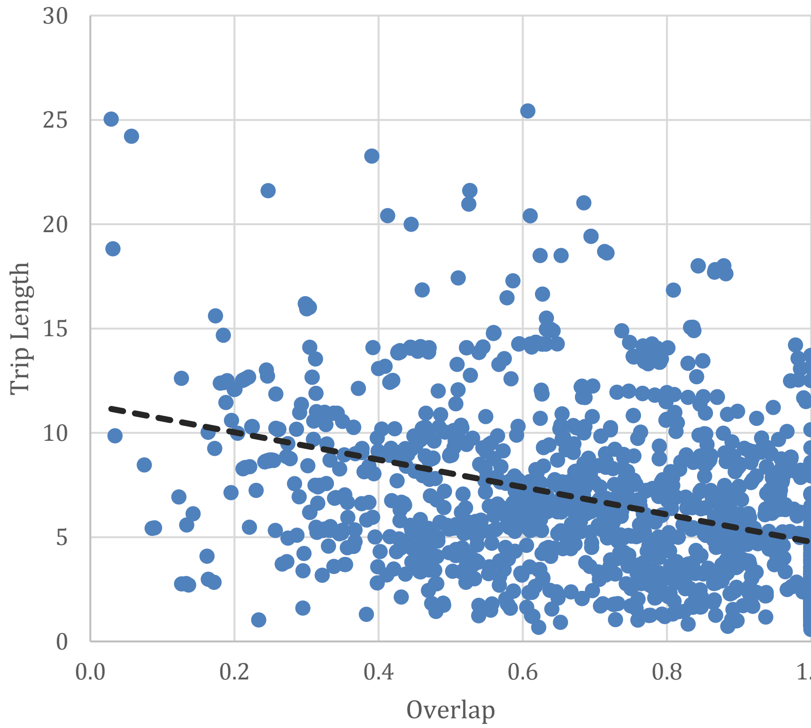


Figure 5.8 Proportion of Observed Route Overlap of Choice Set, by Trip Distance

5.3 Parameter Estimation of Network Attributes

Combinations of parameters are iteratively estimated to identify the model with the best set of attributes and fit statistics. A first model was estimated using only road attributes, and a second model added a few interaction terms of road attributes with trip and demographics variables. The marginal rate of substitution (MRS) measures the trade-off of one unit of one distance for another attribute. The ratio of marginal utilities of the ln of distance and other variables is calculated using the following equation from Broach et al. (2012):

$$Equivalent\ \%\Delta\ distance = \left(Exp\left(\Delta attribute * \frac{\beta\_\text{attribute}}{\beta\_\ln(distance)}\right) - 1\right) * 100$$

5.3.1 Parameter Estimation for Model 1

The estimated parameter coefficients and marginal rates of substitution for Model 1 are shown in Table 5.6 and the marginal rates of substitution are summarized in Table 5.7.

Bike Facilities

The results of Model 1 show that cyclists prefer separated cycle tracks more that painted bike lanes. Cyclist would travel 75% more distance to travel on a cycle track and 50% more distance to travel on a bike lane then a road without a cycling infrastructure. This result supports the findings of stated preference surveys that cyclists prefer separation from traffic. To the author's knowledge, this is the first study to measure the difference in preference between a separated or painted bike lane. Broach et al. included separated bike facilities on bridges, but did not compare painted bike lanes and separated cycle tracks. The MRS value may seem high, but is comparable to other studies. Hood et al. (2011) found an MRS for bike lanes of 49%, and Broach et al. (2012) show that a bridge with separated bike lanes have MRS of 45%.

Off-street bike trails have a smaller importance for cyclists, who will only detour 18% for an off-road trail. This could be due to the fact that there aren't many feasible off-road trails for most OD pairs, and many of these trails are in ravines and have significant hills, which might increase the perceived effort for the route. Bike routes such as signed routes along local streets and sharrows along arterials do not provide a designated space for cyclists, and this minimal amount of infrastructure is not significant in the model. This result contrasts with the findings of Hood et al. (2011), who found that cyclist prefer off-street trails more than bike lanes, which have an average detour of 57% and 49%, respectively. Hood et al. (2011) also found that cyclist preferred bike routes, which are shared with vehicular traffic, compared to bike lanes, with a detour of 92%. The contrasting results are most likely due to the difference configuration of the cycling infrastructure networks in the two cities.

Hills

Cyclists in Toronto will significantly detour to avoid steep hills, which are rare in the City's downtown where most cycling trips were recorded. Cyclists will detour 4.5 times the distance for a path with flat terrain to avoid a road with a grade of 2-4%. This means that a cyclist will detour 459 metres on an alternative path with a lower grade to avoid a 100 metres road segment with a grade of 2-4%. The model estimates that cyclists will detour over 5 kilometres to avoid a 100 m road segment with a grade of 4-6%. The very high detours for steep slopes in the model could be due to the fact that it is highly infeasible in Toronto to have an alternative route for the same OD that can avoid a steep hill, as the variation in topography is mostly due to linear features such as ravines and highways that are hard to avoid if the origin and destination of the trip are on either side of the feature. Broach et al. (2012) found a significant detour distance for steep hills, but the MRS values are much lower. Non-commuters in Portland avoid roads with 2-4% grade by adding 72.3% distance, and roads with 4-6% by adding 2.9 times the distance. Hood et al. (2011) measure the impact of hills using a different metric that counts total rise along the route. They found that cyclists would add 590 metres to their route to avoid a total rise of 10 metres.

Road Class and Traffic Volume

There is a correlation between road class and traffic volume, so the two sets of attributes are modelled separately to see which has more of an impact. There are benefits to using either attribute: road class better represents the road's design and function in the network, but traffic volume better represent conditions of the road that negatively affect cyclists, such as higher speed traffic or congestion. The comparison of models estimated with the two attributes finds that traffic volume better performed in the model than the road class attribute because the later lead to insignificant parameters for bike trails and several turn types.

Model 1 shows that cyclists will add 38% distance to avoid streets with 20,000 cars per day or more. Broach et al. found that cyclist would avoid street with 20,000-30,000 cars per day *without bike lanes* by adding 137% distance. They conclude that the addition of bike lanes to high traffic streets effectively offset the negative impact of traffic volume. Model 1 of this study finds that the benefits of bike lanes outweigh the negative impact of high traffic streets of roads. However, this trade-off varies for different types of cyclists in Model 2.

Table 5.7 Parameter Estimation Choice Set 2, Model 1

Parameter	Coefficient	Std. err	t-test
ln(Length)	-3.54	0.427	-8.31
Bike Lane	2.42	0.159	15.24
Cycle Track	4.89	0.395	12.38
Bike Trail	0.719	0.228	3.15
Hill 2-4%	-6.09	1.03	-5.91
Hill 4-6%	-14.1	2.28	-6.19
Traffic over 20,000/day	-1.13	0.174	-6.48
Turns per km	-0.201	0.034	-5.92
Left Turn with Light	-0.0943	0.0306	-3.08
Left Turn, Unsignalized, Across Arterial	-0.361	0.0338	-10.65
Right Turn with Light	-0.24	0.0321	-7.5
Right Turn, Unsignalized, Across Arterial	-0.055	0.0249	-2.21
Straight, Unsignalized, Across Arterial	-0.156	0.0201	-7.79
ln(PS)	0.968	0.0814	11.9
N observations	1709		
LL(0)	-4389.13		
LL	-3835.16		
Likelihood Ratio Test	1107.952		
Rho-squared	0.126		
Adjusted Rho-squared	0.123		

Table 5.8 Marginal Rate of Substitution of Length by Attribute, Model 1

Attribute	Distance Value (% dist)
Bike Lane	-0.50
Cycle Track	-0.75
Bike Trail	-0.18
Hill 2-4%	4.59
Hill 4-6%	52.68
Traffic over 20,000/day	0.38
Turns per km	0.06
Left Turn with Light	0.03
Left Turn, Unsignalized, Across Arterial	0.11
Right Turn with Light	0.07

Right Turn, Unsignalized, Across Arterial	0.02
Straight, Unsignalized, Across Arterial	0.05

Streetcar Tracks

Travel along roads with streetcar tracks does not have a significant parameter, not does left turns with streetcar tracks. This is an unexpected result, considering the risk of injury caused by streetcar tracks. It may not be possible for cyclists to avoid streetcars tracks due to the particularity of the road network layout in Toronto. There are not many local streets that run in the east-west direction for more than a few block, and most east-west arterials in Toronto are narrow and have streetcar tracks.

Signalized Intersections

The frequency of signalized intersections along a route is not significant in the model and neither is the number of total intersections traversed. However, the impact of signalized intersection is significant once the turn maneuver and road conditions are taken into account.

Turns

The inclusion of turn variables in the model generates an important increase in model fit, and provides an interesting insight into the challenges faced by cyclists when making turns. Turns add a cost to a route by adding a travel time delay as well as a mental cost for remembering the location and sequence of turns. A decrease of 1 turn per kilometre is equivalent to a detour of 6% of the total route, which is small but significant, as it shows that cyclists choose routes that have less turns. Left and right turns all have negative utilities, but these vary based on the intersection type and road class.

The model finds that variables for left and right turns with and without signalized intersections at arterial roads have significant parameters. Left turns at busy streets can be challenging for cyclists. Indeed, the model shows that there is a 11% detour cost for each left turn on busy streets if they do not have a traffic signal, whereas the detour cost drops to 3% if there is a traffic signal. The opposite is true for right turns. Signalized intersections have a higher detour for right turns (7%) than non-signalized intersections along arterials (2%). This may be due to the increased conflict with right turning cars and pedestrians crossing the street, which are more likely to happen at a signalized intersection. The model shows that travelling across a busy arterial road without a traffic signal has a lower cost than a left turn. This may due to the fact that it is a less frequent maneuver, or that the benefits of continuing along the local road outweigh the cost of crossing the road, and an alternative crossing that avoid arterial road is not possible.

5.3.2 Model 2: Interaction Terms of Demographics and Trip Purpose

The response rate for the user survey varies considerably by attribute, and especially for variables such as gender, and so the results of Model 2 are not representative of the cycling population in Toronto. However, it is interesting to see the impact of the interaction terms of those who did respond. Table 5.8 shows the number of responses for the interaction terms tested. Model 2 includes the final interaction terms of trip and demographic variables and the model estimation results and marginal rates of substitution are found in Table 5.9 and Table 5.10.

Table 5.9 Sample Size of Interaction Term Variables

Variable	N
Commute	**1709**
1 - Yes	1178
0 - No	531
Peak	**1709**
1 - On	1151
0 - Off	558
Gender	**902**
Male	284
Female	618
Cycling Frequency	**970**
Monthly	617
Daily and Weekly	353
Comfort riding on roads	**1621**
Not comfortable	442
Can ride on roads	1179

Table 5.10 Parameter Estimation Choice Set 2, Model 2 – with Interactions

Parameter	Coefficient	Std. err	t-test
ln(Length)	-3.49	0.43	-8.12
Bike Lane	1.87	0.188	9.97
Bike Lane * Female	1.23	0.265	4.63
Cycle Track	4.94	0.396	12.47
Bike Trail	0.589	0.231	2.55
Hill 2-4%	-4.19	1.26	-3.33
Hill 2-4% * Female	-5.81	2.2	-2.64
Hill 4-6%	-14.8	2.3	-6.44
Turns per km	-0.219	0.0341	-6.42
Traffic over 20,000/day	-2.54	0.287	-8.84
Traffic over 20,000/day * Comfortable	1.92	0.302	6.36
Left Turn with Light	-0.0841	0.0307	-2.74
Left Turn, Unsignalized, Across Arterial	-0.352	0.034	-10.36
Right Turn with Light	-0.241	0.0321	-7.51
Right Turn, Unsignalized, Across Arterial	-0.0529	0.025	-2.12
Straight, Unsignalized, Across Arterial	-0.151	0.0203	-7.45
ln(PSF)	0.976	0.0822	11.87
N observations	1709		
LL(0)	-4389.133		
LL	-3802.668		
Likelihood Ratio Test	1172.93		
Rho-squared	0.134		
Adjusted Rho-squared	0.13		

Table 5.11 Marginal Rate of Substitution of Length by Attribute, Model 2

Attribute	Distance Value (% dist)
Bike Lane	-0.41
Bike Lane * Female	-0.30
Cycle Track	-0.76
Bike Trail	-0.16
Hill 2-4%	2.32
Hill 2-4% * Female	4.28
Hill 4-6%	68.46
Turns per km	0.06
Traffic over 20,000/day	1.07
Traffic over 20,000/day * Comfortable	-0.42
Left Turn with Light	0.02
Left Turn, Unsignalized, Across Arterial	0.11
Right Turn with Light	0.07

Right Turn, Unsignalized, Across Arterial	0.02
Straight, Unsignalized, Across Arterial	0.04

Interaction terms of distance with commuting vs non commuting as well as peak vs off-peak trips are insignificant. This was surprising because studies by Broach et al. (2012) and Hood et al. (2010) both found that the commute interaction term was significant. It may be that cyclists in Toronto do not vary their route choice behaviour significantly based on trip purpose or by time of day, or it may be due to the loss of many afternoon peak trips in the map matching process. A larger sample of off-peak and non-commute trips is required to identify if travel time and purpose have a significant interaction. There is also considerable overlap of commute trips and peak trips as would be expected - 75% of commute trips occurred during peak periods – and this may explain in part why they were both insignificant interaction variables.

When compared to men and non–respondents, women have a lower sensitivity to travel distance and turn frequency. They are willing to travel 30% further for bike lanes and 200% further to avoid hills of 2-4%. While the coefficient for the interaction with women and cycle tracks was insignificant, the coefficient is negative, implying women place a higher value on separated facilities. Men have a higher sensitivity to travel distance compared to women and non-respondents, but it is not significant, and is not included Model 2.

The users that reported being comfortable riding in most roadway condition are less sensitive to higher traffic volumes. Those who cycled infrequently (less than a few time per week) did not detour as much for bike lanes compared to more frequent cyclists, but the coefficient was not significant. This might be due to lower knowledge of the bike lane locations. Hood et al. (2011) found that infrequent cyclists have a lower marginal utility for bike lanes, and Menghini et al. (2010) also found that faster cyclists, presumed to be more frequent cyclists, used bike lanes more frequently.

It is interesting to compare the stated preference for bike lanes and a perceptions of safety with the actual route choice attributes. The interaction term of comfort in cycling did reveal that cyclists how are less comfortable in all road way conditions are more sensitive to traffic volumes and women cyclist have a higher MRS for bike lanes. However, 78% of female respondents were comfortable riding in most roadway conditions and 22 % were not, whereas 60% men reported that they were comfortable riding in most roadway conditions and 40% were not. This may show that the perception of comfort and the impact on route choice may vary based on gender. Further research with a representative distribution of gender is required to better understand this issue.

5.4 Model Validation

The model's ability to identify the observed route as the most attractive choice is tested on a retained sample of 300 observed trips, using the coefficients estimated in Model 1. The observed route has the highest probability of being chosen in 20% of cases, and it was the second choice in 19% of cases. The distribution of highest probability routes in the choice set is shown in Figure 5.9.

The alternative A2 was chosen by the model in 20% of cases, and is just as likely to be chosen as the observed route. It is the labeled alternative that most highly discounts (by 60%) the cost for roads with bike lanes. Alternative A14 was chosen in 12% of cases, and its label most highly discounts (by 60%) the use of all bike facilities on the network This was also the route that most often replicates the observed route. The shortest path was only selected in 2% of cases. The route choice model shows that cyclists strongly values routes with cycling infrastructure.

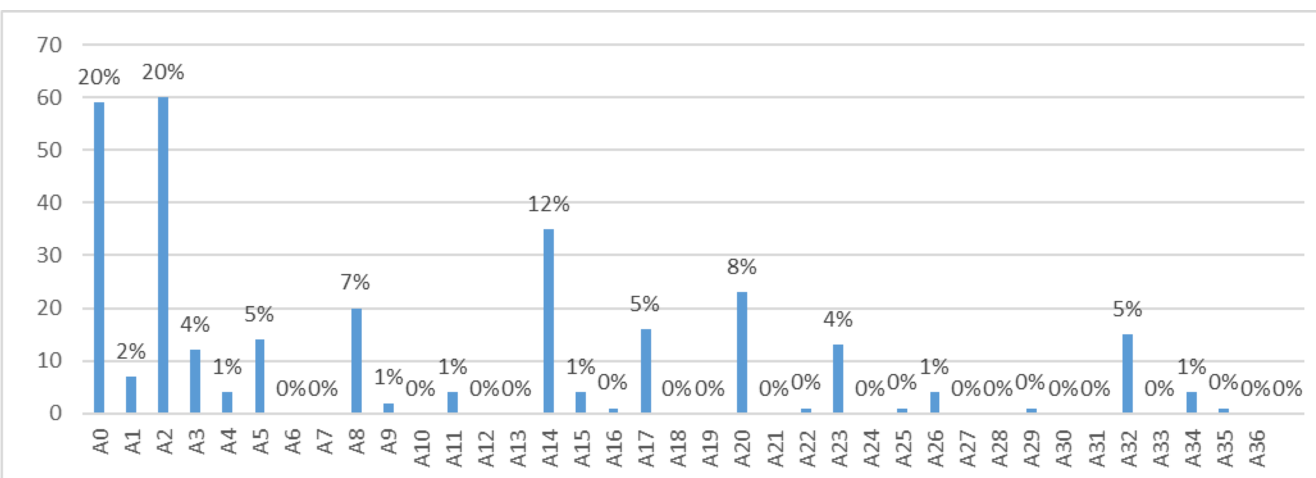


Figure 5.9 Prediction of Observed Routes, Model 1

Other studies have also tried to validate their models by testing the prediction of the observed route. The Waterloo study (Casello and Usyukov, 2014) was able to predict 65% of trips in one of their models, and the San Francisco model (Hood et al., 2011) predicted the observed route 15% of the time.

The model's low prediction of observed routes is most likely due to the fact that there are a number of feasible alternatives and that route choice is a partially random and imperfect process. There is also some level of systematic utility that is not included in the model. For example, quality of pavement is an important factor for the pleasantness and safety of a route (Stinson & Bhat, 2003), but is not often studies in RP studies, most likely because cities do no maintain inventories of pavement quality. It may also be the case that the amount of an individual's taste preference is significant enough to make the chance of predicting the exact observed route quite low. There may also be a higher number of viable routes and so the observed choice of route may be random or based on habit.

Chapter 6

Conclusion

6.1 Summary of Findings

The cycling route choice model developed in this study found that steep hills, high traffic volumes, left turns without signalized intersections and right turns at signalized intersections are all significant physical obstacles for which cyclists are willing to detour from the shortest path.

Cycling infrastructure such as painted and separated bike lanes are significantly preferred by cyclists. They are willing to travel 50% longer to use a bike lane and 75% more for separated cycle track facilities. The preference for cycling infrastructure has been identified in most route choice studies, but compared to cyclists in Portland and Zurich (Broach et al., 2012, Menghini et al., 2010), Toronto cyclists will go further out of their way for bike lanes and less so for off-road bike trails and signed routes in neighbourhoods. Cyclists in Toronto are sensitive to left and right turns, similar to cyclists in Portland, but the Toronto model did not find that the frequency of traffic lights was significant (Broach et al., 2012). Rather the model finds that the presence of a traffic lights reduces the cost of left turns at arterial roads, but the opposite is true for right turns. It may be the case that traffic lights represent areas of higher conflict with right-turning vehicles and pedestrians, and cyclists will choose intersections without traffic lights and with fewer conflicts to make a right turn.

The cycling route choice model in this study did not find a significant difference in the utility coefficients for commuters versus non-commuters or for travel during the AM and PM peak travel times. The interaction term of female respondents was significant for bike lanes and for hills with a grade of 2-4%, which is consistent with other studies that have found that women are willing to travel further to avoid steep hills and to use streets with bike lanes (Hood et al., 2011; Tilahun, 2007).

6.2 Limitations of the Study

The bicycle route choice model developed in this study would benefit from the addition of more attributes of the road network in order to better capture the systematic utility of trips, as well as representative socio-demographic characteristics of the cyclists. There are several attributes of the road network that have been found to have an impact on cycling mode choice and route choice in SP studies, but were not available for this study. These include:

- Tree cover – the presence of mature trees can impact the pleasantness of a route, and in the summer, the shade provided could be sought after by cyclist (Toronto Public Health, 2014). An inventory of tree locations and trunk diameter are available for the City of Toronto, but further research is required to transform this data into an effective measure of tree cover for cycling route choice.
- Pavement quality – A stated preference survey by Stinson and Bhat (2003) found that pavement or surface quality has impact on route choice, but it was not possible to find a complete spatial dataset of pavement quality. It might be possible to approximate the quality based on the most recent date of pavement,.
- Bikes boxes and stop signs are two intersection types that would be important to include in the model to better understand the impact of intersection design on route choice.
- A longer study period can provide a larger sample of trips with complete survey responses that could better estimate the impact of interacting variables such as gender, peak travel time and non-commuting trips. It could also test the variation of parameters by season.

6.3 Future Research

The estimation of cycling route choice in this study can be improved by comprehensively testing choice set generation techniques and new modeling techniques, with special attention given the model's ability to estimate the taste heterogeneity for road attributes based on personal characteristics, which have been found to be significant in SP surveys. A more detailed SP survey of the app users could provide better interaction terms and hypothesises about cycling behaviour. Using a larger sample of trips and selecting a more representative subset of cyclists in the city could help reduce the sampling bias of the dataset.

To make the route choice model of this study useful for policy and planning purposes, a bike trip assignment model could be created and integrated into regional travel demand model. This will require at the least a change in the choice set generation of simplifying the number of labels and combined labels to better reproduce the most attractive route. A predictive cycling assignment model could also be used to improve cycling route planning apps.

Bibliography

Aultman-Hall, L., Hall, F. & Baetz, B. (1997). "Analysis of bicycle commuter routes using geographic information systems: implications for bicycle planning," Transportation Research Record: Journal of the Transportation Research Board, No. 1578, 102–110.

Ben-Akiva, M. & Bierlaire, M. (1999). "Discrete choice methods and their applications to short term travel decisions," in R. Hall, ed., Handbook of Transportation Science, Kluwer Academic Publishers, Norwell, MA, chapter 2, pp. 5–34.

Ben-Akiva and Bierlaire (1999). Discrete choice methods and their applications to short term travel decisions. In R. Hall (Ed.) Handbook of Transportation Science.

Bekhor, S., Ben-Akiva, M. & Ramming, M.S. (2006). Evaluation of choice set generation algorithms for route choice models. Annals of Operations Research, Vol. 144, 235-247.

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models. Proceedings of the 3rd Swiss Transportation Research Conference.

Bovy, P. (2009). On Modelling Route Choice Sets in Transportation Networks: A Synthesis. Transport Reviews, Vol 29(1), 43-68.

Bovy, P. & Fiorenzo-Catalano, S. (2007). "Stochastic route choice set generation: behavioral and probabilistic foundations," Transportmetrica, Vol 3, 173–189.

Broach, J., Gliebe, J. & Dill, J. (2010). Calibrated Labeling Method for Generating Bicyclist Route Choice Sets Incorporating Unbiased Attribute Variation. Transportation Research Record: Journal of the Transportation Research Board, No. 2197, 89-97.

Broach, J., Dill, J. & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. Transportation Research Part A, Vol 46, 1730-1740.

Casello, J. & Usyukov, V. (2014). Modeling Cyclists' Route Choice Based on GPS Data. Transportation Research Record: Journal of the Transportation Research Board, No. 2430, 155-161.

City of Toronto (2011). Briefing Note – Jarvis Street Bike Lane: Traffic Impacts. June 13, 2011. Retrieved August 2016 from https://www1.toronto.ca/City%20Of%20Toronto/Transportation%20Services/Cycling/Files/pdf/J/jarvis-briefing-note-april2011.pdf

City of Toronto, (2014). Official Plan. Section 3.1.1. Retrieved August 2016: http://www1.toronto.ca/static_files/CityPlanning/PDF/chapters1_5_dec2010.pdf

Cycling Think and Do Tank, (2013) Mapping Cycling Behavior in Toronto. Retrieved August 2016: http://www.torontocycling.org/uploads/1/3/1/3/13138411/mapping_cycling_behaviour_in_toronto_final_23_may_printer_tl.pdf

Cycle Toronto (2016). College St moves nearly as many bikes as it does cars during rush hour. Retrieved August 2016: https://www.cycleto.ca/news/college-st-moves-nearly-many-bikes-it-does-cars-during-rush-hour

Dalumpines, R. & Scott, D.M. (2015). GIS-based episode reconstruction toolkit: transferable, modular, and scalable framework for automated extraction of activity episodes from GPS data. 94thAnnual Meeting of the Transportation Research Board, Washington, D.C., January 11-15.

Data Management Group (2013). 2011 Transportation Tomorrow Survey. Retrieved from: http://dmg.utoronto.ca/drs-access

Dill, J. & Carr, T. (2003). Bicycle commuting and facilities in major US cities: if you build them, commuters will use them. Transportation Research Record: Journal of the Transportation Research Board, No. 1828, 116-123.

Dill, J. & Gliebe, J. (2008). Understanding and Measuring Bicycling Behavior: a Focus on Travel Time and Route Choice. Oregon Transportation Research and Education Consortium, Final Report, OTREC-RR-08-03.

Dill, J. & Voros, K. (2007). Factors affecting bicycle demand: initial survey findings from th Portland, Oregon, Region. Transportation Research Record: Journal of the Transportation Research Board, No. 2031, 9-17.

Ehrgott, M., Wang, J.Y.T, Raith, A. & van Houtte, C. (2012). A bi-objective cyclist route choice model. Transportation Research Part A: Policy and Practice, Vol 46(4), 652–663.

Frejinger, E. & Blerlaire, M. (2007). Capturing correlation with subnetworks in route choice models. Transportation Research Part B – Methodological, Vol 41(3), 363–378.

Fomin, R. (2011), TTC Tracks. http://transit.toronto.on.ca/spare/0055.shtml

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

Golledge, R.G. & Gärling, T. (2002). Spatial behaviour in transportation modeling and planning. In: Goulias, K. (Ed.), Transportation Systems Planning: Methods and Applications. CRC Press, New York, pp. 3–1 to 3–21

Hood, J., Sall, E. & Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California. Transportation Letters: The International Journal of Transportation Research, Vol 3, 63-75.

Larsen, J. & El-Geneidy, A. (2011). A travel behavior analysis of urban cycling facilities in Montréal Canada. Transportation Research Part D, Vol 16, 172-177.

Ma, L., Dill, J. & Mohr, C. (2014). The objective versus the perceived environment: what matters for bicycling? Transportation, Vol 41,1135-1152.

Menghini, G., Carrasco, N., Schüssler, N., Axhausen K.W. (2010). Route choice of cyclists in Zurich. Transportation Research Part A, Vol 44, 754-765.

Neilsen, O.A. (2000). A stochastic transit assignment model considering differences in passengers utility functions. Transportation Research Part B, Vol 34(5), 377-402.

Nelson, A.C. & Allen, D. (1997). If you build them, commuters will use them. Transportation Research Record 1578, 79-83.

Papinski, D. & Scott, D. M. (2011). A GIS-based toolkit for route choice analysis. Journal of Transport Geography, Vol 19, 434–442.

Parkin, J. & Rotheram, J. (2010). Design speeds and acceleration characteristics of bicycle traffic for use in planning design and appraisal. Transport Policy, Vol 17, 335-241.

Prato, C.G. (2009). Route choice modeling: past, present and future research directions. Journal of Choice Modelling, Vol 2(1), 65-100.

Prato, C. & Bekhor, S. (2007). "Modeling route choice behavior: How relevant is the composition of choice set?" Transportation Research Record: Journal of the Transportation Research Board, Vol 2003, 64–73.

Pereira Segadilha, A.B. & da Penha Sanches, S. (2014). Analysis of Bicycle commuter routes using GPS and GIS. Procedia - Social and Behavioral Sciences, Vol 162, 198 – 207.

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

Sener, I., Eluru, N., & Bhat, C. (2009). An Analysis of Bicycle Route Choice Preferences in Texas, U.S. Transportation, Vol 36(5), 511-539.

Schuessler, N. & K.W. Axhausen (2009). Map-matching of GPS traces on high-resolution navigation networks using the Multiple Hypothesis Technique (MHT), *Working paper*, 568, institute for transport planning and system (IVT), ETH Zurich, Zurich.

Stinson, M., & Bhat, C. (2003). Commuter bicyclists route choice: analysis using a stated preference survey. Transportation Research Record, No. 1828, 107-115.

Strava Metro (2015). Seattle DOT Case Study. Retrieved August 2016 from: http://metro.strava.com

Teschke, K., Dennis, J., Reynolds, C., Winters, M., & Harris, M.A. (2016). Bicycling crashes on streetcar (tram) or train tracks: mixed methods to identify prevention measures. BMC Public Health, Vol 16, 617.

Tilahun, N.Y., Levinson, D.M., Krizek, K.J. (2007) Trails, lanes or traffic: valuing bicycle facilities with an adaptive stated preference survey. Transportation Research Part A – Policy and Practive, Vol 41(4), 287-301.

Travel Modelling Group (2015). GtaModel v4.0, Design, Background, and Calibration. Technical Report, University of Toronto Transportation Research Institute.

Toronto Public Health (2014). Healthy Streets: Evidence Review. October 2014. City of Toronto, 18-20.

Vernez-Moudon, A.V., Lee, C., Cheadle, A.D., Collier, C.W., Johnson, D., Schmid, T.L (2005). Cycling and the built environment, a US perspective. Transportation research Part D – Transport and Environment, Vol 10(3), 245-261.

Appendix A

User Demographics and Trip Purpose Survey Questions

Cycling frequency
value 1: "Less than once a month"
value 2: "Several times a month"
value 3: "Several times per week"
value 4: "Daily"
How long have you been a cyclist?
value 1: "Since childhood"
value 2: "Several years"
value 3: "One year or less"
value 4: "Just trying it out/ just started"
Your comfort level on a bicycle
value 1: "Not comfortable sharing roadways with vehicles."
value 2: "Only comfortable sharing roadways in clearly designated cycling facilities."
value 3: "Comfortable cycling in most roadway conditions."
Are you a winter cyclist (December to March)?
value 1: "not a winter cyclist"
value 2: "a winter cyclist"
Trip Purpose
value 1: "Commute: This bike trip was primarily to get between home and your main workplace."
value 2: "School: This bike trip was primarily to go to or from school or college."
value 3: "Work related: This bike trip was primarily to go to or from a business related meeting, function, or work-related errand for your job."
value 4: "Exercise: This bike trip was primarily for exercise, or biking for the same sake of biking."
value 5: "Social: This bike trip was primarily for going to or from a social activity, e.g. at a friend’s house, the park, a restaurant, the movies."
value 6: "Shopping: This bike trip was primarily to purchase or bring home good or groceries."
value 7: "Errand: This bike trip was primarily to attend personal business such as biking, doctor, visit, going to the gym, etc."
value 9: "Other: if none of the other reason applied to this trip, you can enter comments bellow to tell us more."