Cycling in Toronto: Route Choice Behaviour and Implications to Infrastructure Planning

Siyuan Li  
M.A.Sc. Student, Department of Civil & Environmental Engineering,  
University of Waterloo,  
Waterloo, ON, N2L 3G1, Canada  
Email: s354li@uwaterloo.ca

Matthew Muresan  
Ph. D. Student, Department of Civil & Environmental Engineering,  
University of Waterloo,  
Waterloo, ON, N2L 3G1, Canada  
Email: mimuresa@uwaterloo.ca

Liping Fu  
Professor, Department of Civil & Environmental Engineering,  
University of Waterloo,  
Waterloo, ON, N2L 3G1, Canada  
School of Transportation and Logistics,  
Southwest Jiaotong University, Chengdu, P. R. China  
Phone: (519) 888-4567 ext 33984, Email: lfu@uwaterloo.ca

Paper prepared for presentation at the 96th Annual Meeting of the Transportation Research Board  
January 2017

Word count = 5998 (text) + 6*250 (figures and tables) ~ 7498
ABSTRACT

This research investigates the route choice behavior of cyclists in the City of Toronto using data collected from a smartphone application deployed to a large number of cyclists in the City. A total 4556 cyclists registered for this study and logged over 20,000 commuting trips over a study period of 9 months. In addition to the time-stamped second-by-second GPS readings of each trip, information on age, gender and residence, work or school place is also collected on a voluntary basis. Multi-nominal logit route choice models are calibrated for the commuting cycling trips. The results have revealed the critical importance of cycling facilities such as bicycle lanes, cycling paths and trails on cyclists’ route choice decisions, providing valuable information for Toronto’s ongoing effort on bicycle network planning.
INTRODUCTION AND BACKGROUND

Increased travel demand from population growth in many regions worldwide creates substantial challenges for planners and engineers. In many regions, these travel demand increases lead to increases in motor vehicle traffic, which in turn increases congestion, collisions, noise and pollution. The Greater Toronto Area (GTA) in Ontario, Canada is one of Ontario’s fastest growing regions, and its population is projected to increase by 40% by 2041 (1). Previous studies have estimated that congestion in Toronto already costs commuters $3.3 billion in lost time and additional vehicle maintenance, and may also cost the wider economy an additional $2.7 billion (2). Additionally, it is estimated that transportation currently represents 24% of total greenhouse gas (GHG) emissions (3). Regional governments therefore have all been taking initiatives to encourage the use of alternative transportation modes. In particular, active modes of transportation, such as walking and cycling, have been promoted as sustainable alternatives to automobile travel. With a population of over 2.6 million Toronto is the fifth largest city in North America. Cycling in Toronto has seen substantial growth over the past five years, with the number of commuting trips made by residents increasing by 33% from 2001 to 2006. Despite this, its popularity as a primary mode choice is still low, and represents only about 1.7% of all commuting trips (4).

The City of Toronto is currently developing a new long term plan to revitalize its cycling network to achieve the goal of increased cycling. A number of cycling network projects are being proposed. These projects are to be evaluated on eight evaluation criteria, including Connectivity, Coverage, Crossing Barriers, Current Demand, Potential Demand, Trip Generators, Safety, and Population and Employment Density. One of the top requirements for estimating the benefits of these projects is quantitative understanding of factors affecting traveler preferences towards different modes as well as different bicycle facilities.

Although past research has investigated the effects of many different factors on cycling, few have considered both trip operational and network factors. Also, of the past studies, most are limited to the evidence of stated preferences with some hypothetical choice options or a small sample of revealed preference data, which can only be used to examine the effects of factors separately with little possibility of revealing hidden interactions and the relative importance of each factor in cyclists’ route choice behavior. The objective of this study is therefore to investigate the cyclists’ choice behavior using reveal preferences data from their real cycling decisions. With a data set of over 20,000 commuting trips, the study has the opportunity to examine the effect of multiple factors, including factors that have not been analyzed in previous studies, such as types of bicycle facilities, energy consumption, and number of bus stops.

The remainder of this paper describes the literature on related topics, the methodology applied, and the results of the analysis for the dataset. The next section presents a summary of the data collected, including an overview of the processing required to prepare the data for input in a logit model. Following this, the next section gives an overview of the model specification, including a discussion on the rationale behind the variables included and overview of how the logit model is applied to solve this problem. Finally, the last two sections review the results and provide the conclusions and recommendations for future research.
Cyclist route choice analysis has received considerable attention in the past few decades, and many studies have been done to identify the factors that capture travelers’ route choice behavior. Research on route choice behavior, however, faces three main challenges. The first is the ability to collect data on the travel patterns and preferences of users, which is also most time consuming (5). Previous studies have frequently employed three different data collection methods: revealed preference surveys, stated preference surveys and global positioning system (GPS). Currently, many studies rely on either revealed or stated preference surveys as the main method to collect user preference data. Revealed preference surveys ask respondents about the cycling routes they took recently for the trips between specific origins and destinations (6, 7, 8). In contrast, stated preference surveys ask users about their preferences by comparing different options; they may ask respondents, for example, to rank the factors they considered when choosing their routes (9) or to choose from some hypothetical choices presented to them by the researchers (e.g. (10, 11, 12, 13).

Although survey-based data collection hampers the ability of researchers to collect accurate data, recent advances in technology, particularly the increased adoption of GPS-enabled smartphones, have opened new possibilities. Collection of this data requires less effort from a participant, as software can be created to automatically log a user’s path, and therefore the scale of the data collection effort can be increased as respondents will find it easier to participate. As a result of this, many studies on route choice behavior using GPS data have been done in the past few years (e.g. (5, 14, 15, 16). Despite these benefits, GPS-based data collection also has issues. For example, the precision the GPS device and the process to match collected data points to a map may cause inaccuracies. Studies have shown that, at best, 6% of the data collected will contain errors (15). Furthermore, if GPS data is collected automatically by a broad category of participating devices, other modes like cars or pedestrians may be mis-detected as bicyclists in the dataset (14).

The second challenge is particularly related to studies using revealed preference data. In these studies, only the routes that were actually chosen or used are known while the true alternative routes that had been considered by the travelers are unknown and can only be speculated. The alternative routes are needed for choice modelling in order to gain quantitative understanding of cyclists’ preference structure related to different choice factors.

The third challenge is to identify the factors that may have an effect on cyclists’ route choices. Previous studies have shown that cyclists choose their routes based on a combination of multiple factors (8), which can be divided into three types: trip operational factors, such as distance and travel time, network factors, such as road surface condition, and demographic and socio-economic factors, such as gender, age and income.

Factors Affecting Cyclist Route Choice

One of the most important factors identified by previous studies is distance (or travel time). According to the 2001 National Household Travel Survey (NHTS), 41% of all trips in 2001 were shorter than 2 miles, and 28% were shorter than 1 mile (17). In 1985, Bovy and Bradley (18) examined cyclists’ route choices in Delft, Netherlands and found that travel time always played
the most important role in the choices cyclists made. Despite the importance of travel time, cyclists do not always choose the shortest possible path. For example, a study done in 1997, Aultman-Hall et al. (6) compared the actual routes cyclists took with the shortest possible paths in Guelph, Ontario and found that only 14.6% of the routes were exactly the same as the shortest paths; however, many of the routes chosen still showed a preference to the shortest path, as 37.5% of routes were within 0.1 km of the minimum distance, with average deviation from the shortest path only being 0.4 km.

In addition to distance or travel time, network conditions also play important part in route choice decisions. Stinson and Bhat (19) did a study that considered 11 link-level and route-level factors collected from stated preference surveys and used them to analyze commuter’s route choice preferences through a discrete choice modeling framework. They found that cyclists preferred streets with lower traffic, with residential streets being the most preferred, followed by minor arterials. The study also found that cyclists avoided streets with parallel parking but preferred streets with bicycle facilities, that totally separate motor and non-motorized traffic, flat ground and routes with smooth pavement. These findings have been corroborated by other studies as well, such as one by Tilahun et al. (13) that grouped facilities into five types based on their characteristics: off-road, in-traffic, bike lanes, and on-street parking. They found that bicycle facilities and on-street parking affect route choice behavior, with cyclists sometimes choosing a longer route to avoid bad cycling facilities. A similar result was also shown from GPS data collected by Broach et al. (2012), 53% of recorded trips were on facilities with bicycle infrastructure. Similarly, a study by Krizek (12) examined the effect of cycling infrastructure on route choice, but used additional factors such as season and demographics to build a generalized mixed logit regression model. They found that facilities contributed positively if travel times were the same, with bike lanes being the most favored followed by streets without parking.

A number of studies have also assessed the effect of road gradient on the route choice behavior of cyclists. For example, Menghini et al. (14) and Hood et al. (5) assessed the effect of average absolute gradient, maximum gradient, and the number of traffic lights on route choice behavior. They found that chosen routes are always shorter, less steep, with fewer lights and more bicycle paths when compared statistically with non-chosen paths. In particular, the study by Hood et al. (5) improved the analysis done in previous studies by further dividing bicycle facilities into three classes, off-street bike paths, separated bike lanes, and on-street shared lane. They concluded that most users preferred bike lanes followed by bike paths. The study also found that cyclists were willing to make 0.49 to 0.92 km detours to choose routes with bicycle facilities.

Finally, some past studies have found that demographic factors can also influence a cyclist’s route choice. Some have found that lower income households tended to have significantly shorter commute times (20) and that personal characteristics, such as age or sex, have an effect on the choice behavior of cyclists (21).

Although considerable research has been done into the factors that affect cyclists’ preferences, contradicting findings are abundant. For example, some studies have found that cyclists prefer flat terrain (5, 14, 19, 21) while others have found that cyclists prefer moderately hilly terrains as they prevent boredom (21).
DATA DESCRIPTION

The data used in this study was collected in the City of Toronto by a tailor-made smartphone application developed by Brisk Synergies - CycleTrack. The app was promoted by the City of Toronto through city-wide campaigning via a variety of media. Participation in the project was voluntary and information was recorded on a per-trip basis. Users can also complete a voluntary survey and provide information about themselves, such as their age, gender, as well as information about each trip they make. Completion of this survey is not required to record trips on the app, and users have full control over what data is submitted to the servers.

A total of 4556 participants participated in the project between May 20th, 2014 and January 25th, 2015. As participants are not required to complete the user survey, some opted not to complete it. Of the collectable fields, gender had the poorest response rate, with 72.8% of participants preferring not to disclose their gender. Response rates to most of the other fields were generally high, with around 80% of respondents answering all other questions, among which 33.8% are between 25 and 34 years old followed by age between 35 and 49 makes up of 30% of all respondents. Around 31.6% of all cyclists come from households with more than $100,000 or greater income per year, although nearly 30% do not have a selection in household income for the privacy concerns. Approximately half of participants reported that they have cycled since they were children and approximately half of participants also reported that they do not cycle in the winter.

In addition to the GPS data, this research has also made use of network data available from Toronto’s Open Data portal containing the street names, locations, and road types as well as the locations of all the city’s bicycle facilities, as shown in FIGURE 1. Road type was considered as a surrogate for factors not available in the dataset, such as roadway speed limits. As only one-direction edges are provided by Toronto Open Data, node data was first extracted from the start and end points of all edges. These edges were then expanded into bi-directional roads or one-way roads with contra-flow bike lanes. There are five types of biking facilities, including Bike Lanes (separate from motor traffic), Cycle Tracks (bicycle only), Multi-Use Trails (all non-motor users - bicycles, pedestrians, rollerblades), and Shared Roadways (shared with motor vehicles but with high priority), as illustrated in FIGURE 1. This network forms the basis by which all subsequent analyses are conducted. The final integrated cycling network used in this analysis contained 14286 km of links and 40808 intersections, of which 46% are residential or local streets.

Although the City of Toronto continues to encourage cycling as a mode choice, only 11.61% links have bicycle facilities and only 1.59% of links have bike lanes. This network was then further augmented with slope information obtained from Digital Elevation Model (DEM) data (23) and annual average daily traffic volume (AADT) obtained from Toronto Traffic Management Center.
FIGURE 1 Cycling Network and Facility Types in Toronto

- **Minor Multi-use Pathway**
- **Sharrows**
- **Signed Routes**
- **No Facility**
- **Bike Lanes**
- **Cycle Tracks**
- **Major Multi-use Pathway**

A Trip Example of GPS Data in August

Network:
- No Facility
- Bike Lanes
- Cycle Tracks
- Major Multi-use Pathway

---

**Bike Lane**

**Shared Lane**

**Cycle Tracks**

**Multi-use Trail**

---

1. Cycling Network of the City of Toronto
2. A Trip Example of GPS Data in August
3. FIGURE 1 Cycling Network and Facility Types in Toronto
Collected GPS data points were then matched to links on the network. This map-matching process was conducted using a multi-step approach. First, erroneous GPS records outside of a 20-meter buffer created around all links were removed. The remaining points not within 20 meters of an intersection were immediately matched to the link they were nearest to. Remaining points were then assigned to a link based on a comparison between the previous 10 and following 10 points. Using this process, 36221512 GPS records representing 63.33% of the total data collected were considered valid. After matching all GPS records, trips can be generated based on the trip’s starting and ending points, starting and ending time as well as the total distance, average speed and the sequence of links cyclists travel on.

Because this data comes from volunteers, there are no quality controls in place for participants and the trips they log, and so it is possible that non-cycling trips could be recorded. To mitigate this, trips with average speeds exceeding 36 km/h (or 10 m/s) were deleted. This value was chosen after reviewing relevant literature and consulting resources from local cycling clubs. For example, Toronto’s Cabbagetown Cycling Club groups cyclists into 5 groups based on riding skill. Cyclists in the group with the most experience (A* level) generally have average speeds of between 30 to 32 km/h (24). In addition, for short trips, the alternative routes are very limited, and it is hard to analyze the effects of other factors, so in this research, we focused on middle-distance trips. Extremely short trips (trip distance less than 2 km) and trips with an unreasonable detour length (distance is double of the shortest distance) were filtered.

The final dataset used in the analysis contains 20730 commuting trips, representing 62.4% of the total trips in the dataset. The distribution of distance is shown in FIGURE 2. Of the trips made, around 11853 trips used more than 10% bike lanes, comparing with 5234 trips using more than 10% shared lanes.

All of the data was processed using a combination of ArcGIS, a GIS platform, and PostgreSQL, a relational database platform with support for PostGIS functions and readymade PL/pgSQL.
scripts to help process data. In this study, creating buffers, spatial joining GPS records to link buffers, separating the records into two datasets, adding surface slope information were performed in ArcGIS Desktop platform. In this study, building bikeway network, map-matching arbitrary GPS records, trips generating are all performed in PL/pgSQL in PostgreSQL database.

5 DEVELOPMENT OF ROUTE CHOICE MODELS

In order to understand the factors that influence the routes cyclists choose, the choices available to them must first be identified. Since the reasoning behind a cyclist’s route choice cannot be known, candidate alternative route choices must instead be developed based on a set of factors that are known to influence the decision. Once potential alternatives have been identified, the problem can be framed and analyzed using a multinomial logit structure. The following sections highlight the process behind these two steps.

Generation of Alternative Route Choice Sets

Although a cyclist ultimately chooses a single route to follow for their trip, there are always a number of other possible choices that could have been considered. Consequently, two cyclists going to and from the same place may end up choosing different routes, depending on the factors they consider when making the decision. For example, distance is likely to be one of the most significant influencers of the decision, and should therefore be considered when developing alternative choices.

Two types of iterative approaches are commonly used in the literature to generate alternative routes: stochastic choice set generation, which predefine the link costs by a probability distribution, and deterministic route choice set generation methods, which include link elimination, link penalty and path labeling (25). Because of the insufficient variation, stochastic choice set generation often produces many duplicated alternative routes. Deterministic route set generation methods are widely used, especially path labeling algorithms, which repeats shortest path generation by using different cost functions called labels. The labels may consider travel time, distance, proportion of different road types or bicycle facilities or even congestion (25). Other approaches such as the link penalty method increases the link costs by a factor, and repeats the shortest path algorithm (SPA) to ensure generation of a diverse set of alternatives, while link elimination uses certain order deletion of existed links to avoid large overlapping among alternatives.

In this study, unique alternatives are generated through a combination of path labeling and link penalties. A link penalty multiplier of 1.5 (50% increase in travel distance) was applied to each link chosen for each iteration. The alternatives were generated based on two scenarios: shortest path routes and minimal energy cost routes. Shortest paths are based on length of the links while energy cost routes are based on a cyclist’s ability to overcome rolling resistance, air resistance and the grade when it is uphill. We adopted the energy function from Olds et al. (26), the energy of cycling is defined as

\[ E_{tot} = E_{Rr} + E_{Ra} + E_{grade} \]
\[ = C_{Rr} \cos(\arctan(S)) (M + M_b)d + E_{Ra} + (M + M_b) \sin(\arctan(S_{uphill}))d \]

where \( E_{Rr} \) = the energy cost in overcoming rolling resistant (J)
The energy cost in overcoming air resistant ($E_{Ra}$), which is assumed to be constant in the context of this research (the average is 0.19 J).

The energy cost in overcoming uphill grade ($E_{grade}$).

$C_{rr}$ = the factor related with tire pressure and tread, wheel radius, and road surface. The default value is 0.0457, which is used in this research.

$S$ = the gradient of the road section

$d$ = distance (m)

$M$ = the mass of the rider and $M_b$ = the mass of the bicycle (kg)

In each scenario, the cyclist tries to minimize the distance or energy cost. SPA was iterated for generating a given number of paths in each scenario. FIGURE 3 shows an example of route set. In this example, the total distance of each of the candidate routes is around 10 km. It was found that the overlapping between the alternative routes is relatively small, which could be due to the link penalty method we adopted.

FIGURE 3 Chosen and Shortest Routes  Chosen and Energy Saving Routes for Trip 18776

Calibration of Logit Route Choice Model

When the chosen routes for all trips in the data are compared with the alternative routes generated, 22.5% overlap with the shortest path by more than 40%. As the average length of the trips recorded in the data is around 8 km, this means that more than 3 km of each trip follows the same path as the shortest path. However, the chosen routes are not exactly the same as the minimum distance paths, which means other factors like bicycle facilities or transit stops may cause cyclists to choose alternative routes. To determine which factors are significant and which are not, a discrete choice model was applied to the candidate route set generated earlier and the factors differentiating each route were considered.

Discrete choice models are the most commonly used method to model route choice. The probability of choosing option $i$ in Multinomial Logit model (MNL) is defined as:
\[
P(A_i) = \frac{e^{V_i}}{\sum_{A_j \in A} e^{V_j}}
\]

The most important characteristic or disadvantage of MNL is the Independence from Irrelevant Alternatives (IIA), which states that the ratio of the probabilities of choosing two alternatives for each individual is independent of the availability or attributes of any other alternatives. However, there are some overlapping between each two links, which is hard to avoid. The Path Size Logit (PSL) model includes a path size factor into the utility function to overcome the overlapping issue (27). The path size factor is specified by Ben-Akiva and Bierlaire (27):

\[
PS_{ln} = \sum_{a \in \Gamma_i} \frac{l_a}{L_i} \frac{1}{\sum_{j \in C_n} \frac{(L_j)}{L_i} \gamma \delta_{aj}}
\]

where \(\Gamma_i\) are the set of links in alternative \(i\), \(l_a\) is the length of link \(a\), \(L_i\) is the length of route \(i\), and \(\delta_{aj}\) is 1 if path \(j\) includes link \(a\) and 0 if not (27). \(\gamma\) is a positive scaling term to penalize long routes. As the alternative routes do not contain unreasonable detour, we adopted \(\gamma = 0\) as suggested by Broach et al. (16). Finally, the probability of choosing option \(i\) in Path Size Logit model (PSL) is defined as:

\[
P(A_i) = \frac{e^{V_i+b+\ln(PS_{ln})}}{\sum_{A_j \in A} e^{V_j+b+\ln(PS_{jn})}}
\]

In addition to the IIA problem, multiple trips are produced by same cyclists can also bias the results, as the preferences of a cyclist making multiple trips may be over-represented in the analysis. However, due to the fact that users were not required to record personal information, it was not possible to conclusively determine if a trip was made by the same person at all times. Each trip was therefore considered independently.

The key parameter of the model requiring specification is the utility function, \(V_i\), which is assumed to be a multi-variable linear function of factors that are believed to influence route choice. For this study, the following functional form was selected:

\[
V_i = b_1 \cdot D_i + b_2 \cdot E_i/1000 + b_3 \cdot T_i + b_4 \cdot AADT_i/1000 + b_5 \cdot MA_i + b_6 \cdot MI_i + b_7 \cdot C_i + b_8 \cdot L_i + b_9 \cdot SL_i + b_{10} \cdot BL_i + b_{11} \cdot MP_i + b_{12} \cdot CT_i + b_{13} \cdot NF_i
\]

where \(D_i\) = the distance of route \(i\) (km)
\(E_i\) = the energy consumption of route \(i\) (kj)
\(T_i\) = the number of transit stops of route \(i\)
\(AADT_i\) = the annual average daily traffic volumes of route \(i\) (veh/day)
\(MA_i\) = the proportion of major arterial of route \(i\)
\(MI_i\) = the proportion of minor arterial of route \(i\)
\(C_i\) = the proportion of collector of route \(i\)
\(L_i\) = the proportion of local streets of route \(i\)
\(SL_i\) = the proportion of shared lane of route \(i\)
\(BL_i\) = the proportion of bike lane of route \(i\)

\(MP_i\) = the proportion of multiuse trail of route \(i\)

\(CT_i\) = the proportion of cycle tracks of route \(i\)

\(NF_i\) = the proportion of streets without bicycle facility of route \(i\)

As discussed in the previous sections, some of the major factors affecting cyclist’s route choice behavior are distance and the presence of cycling facilities, and so their inclusion in the model is logical. Cycling facilities are included through values representing the proportion of the trip between 0 and 1. In addition to assessing the effects of these terms, our study aimed to assess the significance of other factors which have been previously identified in the model, including the energy consumption, effect of road type and volume. Although trip distance is one of the most important parameters in energy consumption function, the correlation between distance and energy consumption (0.85) is relative high. Despite this, as gradient is also related to energy consumption, cyclists may treat energy consumption differently than distance. The effect of road type is expected to arise from differences in the number of lanes, speed of motor vehicles, roadway condition, and level of service of cycling. Despite consideration of traffic volume as one of the factors of classification of road type, the averaged traffic volume (AADT) has low correlations (between 0.1 and 0.55) with the various road types. This suggests that road type is not a good surrogate for volume, and it should be considered separately in the model. As with cycling facilities, these factors are included in the model as a value representing the fraction of the trip occurring on either facility. In addition, the effect of transit routing on cyclist route choice was also considered. In urban regions, transit vehicles often have speeds similar to that of cyclists, but have a different speed regime as they make frequent stops to allow passengers to board and alight. This behavior can make cyclists feel unsafe and may increase the risks of collisions with pedestrians. Although correlation between transit stops and roadway type was not observed, some correlation was observed between the number of transit stops and distance. The correlation was low (0.5) and is caused by the fact that longer trips will invariably pass by more transit stops. Although utility functions can be visualized linearly, the relationship between the factors and a user’s choice are not linear in nature. Iterative approaches can, however, be used to solve this problem, and many statistical software packages have the ability to estimate Multinomial Logit (MNL) models. This study uses Biogeme (28), which is an open source software tool for the maximum likelihood estimation of parametric models in general, with a special emphasis on discrete choice models.

A number of models were calibrated and compared with the following alternative settings:

- Size of alternative route set for each trip: 10 versus 20
- Distance versus its logarithmic form (LN(distance))
- Grouping of trips by trip distance: <10 km, 10-15 km, >15 km and all
- Inclusion versus exclusion of path size factor
- Combination of route attributes: distance, energy consumption, number of transit stops, road types and facility types

The alternative models are evaluated using log likelihood and \(\rho^2\) values. It was found that a model considering a higher number of alternative routes (n=20) performed better than one with ten alternatives, as shown in TABLE 1. Transformation of distance in a logarithmic form
improved the model fitting. TABLE 2 shows the calibration results of models with different market segmentation by trip length, which shows little improvement as compared to a model considering all trips together. When the trip length is more than 10 km, the sensitivity of length is not as obvious as short trips, but the energy consumption becomes sensitive. As most of the trips are between 5 and 10 km, the model for all trips performed similar with the one using trips less than 10 km. However, a comparison between a normal MNL and a Path Size Logit model shows counterintuitive results. The inclusion of the path size factor resulted in a negative coefficient. Furthermore it rendered other factors having signs with counterintuitive interpretation. One of the reasons for this result is that the alternative routes generated by the link penalty method have little overlap. On the other hand, the chosen route has a large overlapping with one of the routes generated, thus leading to positive effect of overlapping. As a result, we decided to drop off the path size factor and the MNL model was chosen in further analysis.
<table>
<thead>
<tr>
<th>Variables</th>
<th>MNL (n = 20)</th>
<th>MNL (n = 10)</th>
<th>MNL (n = 20)</th>
<th>MNL (n = 20)</th>
<th>MNL (n = 20)</th>
<th>PSL (n = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>t-test</td>
<td>Coeff.</td>
<td>t-test</td>
<td>Coeff.</td>
<td>t-test</td>
</tr>
<tr>
<td>Distance (km)</td>
<td>-0.335</td>
<td>-27.87</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(*** )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LN(Distance (km))</td>
<td>-</td>
<td>-</td>
<td>2.12</td>
<td>17.43</td>
<td>-3.19</td>
<td>-31.29</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(*** )</td>
</tr>
<tr>
<td>Energy Consumption (1000s kJ)</td>
<td>-0.504</td>
<td>-21.66</td>
<td>-2.26</td>
<td>-60.00</td>
<td>-0.456</td>
<td>-19.80</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(*** )</td>
<td></td>
<td>(*** )</td>
<td>(*** )</td>
<td>(*** )</td>
</tr>
<tr>
<td>Number of Transit Stops</td>
<td>-0.00314</td>
<td>-3.75</td>
<td>9.39e-005</td>
<td>0.11</td>
<td>-7.09e-005</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(*** )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Annual average daily traffic volumes (1000s veh/day)</td>
<td>-0.0893</td>
<td>-3.69</td>
<td>-0.043</td>
<td>-1.75</td>
<td>-0.0847</td>
<td>-3.48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(*** )</td>
<td></td>
<td></td>
<td>(*** )</td>
<td>(*** )</td>
</tr>
<tr>
<td>Trip proportion (TP) with major arterial²</td>
<td>23.0</td>
<td>104.01</td>
<td>19.0</td>
<td>84.29</td>
<td>22.5</td>
<td>101.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(*** )</td>
<td></td>
<td>(*** )</td>
<td>(*** )</td>
<td>(*** )</td>
</tr>
<tr>
<td>TP with minor arterial³</td>
<td>23.1</td>
<td>99.02</td>
<td>19.4</td>
<td>80.28</td>
<td>22.7</td>
<td>96.72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(*** )</td>
<td></td>
<td>(*** )</td>
<td>(*** )</td>
<td>(*** )</td>
</tr>
<tr>
<td>TP with collector⁴</td>
<td>23.4</td>
<td>90.15</td>
<td>19.7</td>
<td>72.85</td>
<td>23.1</td>
<td>88.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(*** )</td>
<td></td>
<td>(*** )</td>
<td>(*** )</td>
<td>(*** )</td>
</tr>
<tr>
<td>TP with local street⁵</td>
<td>19.5</td>
<td>63.24</td>
<td>17.1</td>
<td>53.22</td>
<td>19.3</td>
<td>62.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(*** )</td>
<td></td>
<td>(*** )</td>
<td>(*** )</td>
<td>(*** )</td>
</tr>
<tr>
<td>TP with shared lane⁶</td>
<td>0.566</td>
<td>3.83</td>
<td>0.186</td>
<td>1.20</td>
<td>0.585</td>
<td>3.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(*** )</td>
<td></td>
<td></td>
<td>(*** )</td>
<td>(*** )</td>
</tr>
<tr>
<td>TP with bike lane⁷</td>
<td>0.937</td>
<td>6.99</td>
<td>0.786</td>
<td>5.57</td>
<td>0.995</td>
<td>7.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(*** )</td>
<td></td>
<td>(*** )</td>
<td>(*** )</td>
<td>(*** )</td>
</tr>
<tr>
<td>TP with multiuse trail⁸</td>
<td>24.8</td>
<td>65.29</td>
<td>21.4</td>
<td>53.71</td>
<td>24.4</td>
<td>64.20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(*** )</td>
<td></td>
<td>(*** )</td>
<td>(*** )</td>
<td>(*** )</td>
</tr>
<tr>
<td>TP with cycle track⁹</td>
<td>10.4</td>
<td>29.39</td>
<td>9.12</td>
<td>25.02</td>
<td>10.4</td>
<td>29.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(*** )</td>
<td></td>
<td>(*** )</td>
<td>(*** )</td>
<td>(*** )</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(*** )</td>
<td></td>
<td>(*** )</td>
<td>(*** )</td>
<td>(*** )</td>
</tr>
<tr>
<td>Path Size Factor</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final log-likelihood</td>
<td>-28767.111</td>
<td>-22805.297</td>
<td>-28678.360</td>
<td>-28678.363</td>
<td>-20323.980</td>
<td></td>
</tr>
<tr>
<td>Rho-Square</td>
<td>0.544</td>
<td>0.541</td>
<td>0.546</td>
<td>0.546</td>
<td>0.546</td>
<td>0.678</td>
</tr>
</tbody>
</table>
### TABLE 2 Model Estimations with Different Market Segmentation by Trip Length

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coeff.</th>
<th>t-test</th>
<th>Coeff.</th>
<th>t-test</th>
<th>Coeff.</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>LN(Distance (km))</td>
<td>-4.00</td>
<td>-34.27</td>
<td>2.12</td>
<td>17.43</td>
<td>3.47</td>
<td>4.40</td>
</tr>
<tr>
<td>Energy Consumption (1000s KJ)</td>
<td>-0.536</td>
<td>-15.54</td>
<td>-2.26</td>
<td>-60.00</td>
<td>-0.507</td>
<td>-6.62</td>
</tr>
<tr>
<td>Number of Transit Stops</td>
<td>0.00136</td>
<td>1.17</td>
<td>9.39e-005</td>
<td>0.11</td>
<td>0.0153</td>
<td>4.76</td>
</tr>
<tr>
<td>Annual average daily traffic volumes</td>
<td>-0.108</td>
<td>-4.29</td>
<td>-0.043</td>
<td>-1.75</td>
<td>-0.337</td>
<td>-1.06</td>
</tr>
<tr>
<td>Trip proportion (TP) with major arterial</td>
<td>20.8</td>
<td>88.06</td>
<td>19.0</td>
<td>84.29</td>
<td>48.4</td>
<td>16.87</td>
</tr>
<tr>
<td>TP with minor arterial</td>
<td>20.8</td>
<td>80.26</td>
<td>19.4</td>
<td>80.28</td>
<td>47.6</td>
<td>21.37</td>
</tr>
<tr>
<td>TP with collector</td>
<td>21.2</td>
<td>73.58</td>
<td>19.7</td>
<td>72.85</td>
<td>45.8</td>
<td>18.99</td>
</tr>
<tr>
<td>TP with local street</td>
<td>18.0</td>
<td>53.08</td>
<td>17.1</td>
<td>53.22</td>
<td>34.6</td>
<td>12.63</td>
</tr>
<tr>
<td>TP with shared lane</td>
<td>0.855</td>
<td>5.49</td>
<td>0.186</td>
<td>1.20</td>
<td>1.06</td>
<td>0.70</td>
</tr>
<tr>
<td>TP with bike lane</td>
<td>1.26</td>
<td>8.78</td>
<td>0.786</td>
<td>5.57</td>
<td>1.53</td>
<td>1.15</td>
</tr>
<tr>
<td>TP with multiuse trail</td>
<td>20.4</td>
<td>48.59</td>
<td>21.4</td>
<td>53.71</td>
<td>56.3</td>
<td>15.87</td>
</tr>
<tr>
<td>TP with cycle track</td>
<td>10.6</td>
<td>28.69</td>
<td>9.12</td>
<td>25.02</td>
<td>4.19</td>
<td>1.34</td>
</tr>
<tr>
<td>TP without bicycle facilities</td>
<td>-2.55</td>
<td>-21.17</td>
<td>-2.91</td>
<td>-24.16</td>
<td>-6.32</td>
<td>-6.40</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>16361</td>
<td>2752</td>
<td>1617</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final log-likelihood</td>
<td>-24622.299</td>
<td>-2142.135</td>
<td>-697.805</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rho-Square</td>
<td>0.506</td>
<td>0.744</td>
<td>0.858</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 Volumes normalized over the length of the trip for each link passed
2 major arterial: primary function is car movement; AADT greater than 2000; specially facilities are desired
3 minor arterial: primary function is car movement; AADT between 8,000 to 20,000
4 collector: provide access to property and traffic movement; 2,500 to 8,000 vehicles per day; special facilities are required
5 local: provide access to property; less than 2,500 vehicles per day
6 shared lane: the cyclists share the lane with vehicles but have the priority
7 bike lane: a dedicated space for cyclists where vehicles are not allowed to stand or drive in
8 multiuse trail: are physically separated from motorized traffic by open space or a barrier, and are sometimes called “off-road paths”. Multi-use paths are often shared with pedestrians, in-line skates and cyclists
9 cycle track: a physical separated bike lane for cycling use only

Route Choice Model (MNL) with 20 alternative routes (plus the chosen one) and logarithmic form of distance shows the best calibration results. According to the Likelihood Ratio Test, the critical likelihood ratio value $\chi^2 = 0.004$ with 1 degree of freedom and level of significance 0.05. LR = 0.006 > 0.004, so number of transit stops should be included into the utility function. TABLE 3 presents the estimation results of distance trade-off as the other variables have marginal rates of substitution with respect to the natural log of distance. The values
represent the perceived equivalent distance value of a unit change in each variable relative to the total distance, and is calculated as follows (16):

\[
\text{Equivalent } \% \Delta \text{distance} = \left( \exp \left( \Delta \text{attribute} \frac{b_{\text{attribute}}}{b_{\ln(\text{distance})}} \right) - 1 \right) \times 100
\]

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Distance value (% dist.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy Consumption (KJ)</td>
<td>0.01537</td>
</tr>
<tr>
<td>Number of transit stops</td>
<td>0.0022</td>
</tr>
<tr>
<td>Annual average daily traffic volumes (1000s veh/day)</td>
<td>2.69</td>
</tr>
<tr>
<td>Trip proportion (TP) with major arterial</td>
<td>-99.91</td>
</tr>
<tr>
<td>TP with minor arterial</td>
<td>-99.92</td>
</tr>
<tr>
<td>TP with collector</td>
<td>-99.93</td>
</tr>
<tr>
<td>TP with local street</td>
<td>-99.76</td>
</tr>
<tr>
<td>TP with shared lane</td>
<td>-16.76</td>
</tr>
<tr>
<td>TP with bike lane</td>
<td>-26.80</td>
</tr>
<tr>
<td>TP with multiuse trail</td>
<td>-99.95</td>
</tr>
<tr>
<td>TP with cycle track</td>
<td>-96.16</td>
</tr>
<tr>
<td>TP without bicycle facilities</td>
<td>159.34</td>
</tr>
</tbody>
</table>

A comparison of the sign and magnitude of the coefficients allows their relative effect to be assessed. The coefficient for distance is negative, suggesting commute cyclists do not want to detour a lot especially in peak hours. A 1% increase in distance would decrease the probability of choosing the route by around 14%. The model also suggests that cyclists are sensitive to energy consumption with 1 kj of additional energy consumed (e.g. from a hill) perceived as equivalent to a distance increase of 0.015%. The number of transit stops has a negative coefficient, which means cyclists try to avoid routes with more transit stops. One additional transit stop perceived as a 0.002% increase in distance, which is more significant in long trips than short trips. The traffic volume also negatively affects the probability of choosing a route. Cyclists are more likely to detour a route 2.69% longer to avoid routes with high traffic volume. No matter which bicycle facilities the link has, more motorized vehicles with high speed may make the route dangerous or at least make cyclists feel unsafe.

The coefficients for the road type suggest that cyclists care about directness more than the type of the road. However, among all road types, the absolute values of minor arterial and collector are higher than major arterial and local streets. If 1% of the route changed from collectors to major arterials, then the probability of choosing that route would decrease by 0.01. Commuting cyclists prefer not to travel on major arterials with higher traffic volume and speed limit. The coefficient for local streets is the lowest of all the road types, suggesting it is the least preferred. This makes sense for commute trips, because cyclists care about directness more than other factors. Choosing local streets often involves long detours, lower speeds, and more bicycles, skaters, and pedestrians, which is in conflict with a cyclists’ idea of the fastest way to arrive the work place.

The model also shows a preference for bicycle facilities, especially separated multiuse trail and cycle tracks. Off-street bicycle facilities are much more preferred than on-street facilities. A 1%
increase in the route’s proportion containing multiuse pathways or cycle tracks was found to be perceived as equivalent to decreasing distance by around 10%. Comparing on-street facilities, the coefficient of bike lanes is greater than shared lanes suggesting that cyclists consider bike lanes safer than shared lanes.

CONCLUSIONS

In this research, we have conducted an extensive modeling analysis of route choice behavior of cyclists in Toronto using a unique data set featuring large sample size, large road network, and detailed tracing of cycle routes. Both trip operational factors such as distance and traffic volume, and network factors such as number of transit stops, types of road, and bicycle facilities were considered. The results show that for commute cyclists, shorter routes with fewer transit stops, less volume, and more off-road facilities are more preferred. Comparing with different types of facilities, off-road facilities like multiuse trails and cycle tracks could increase the probability of choosing the route, while the bike lane are preferred than shared lane for the cyclists. These findings shed lights on importance of establishing a connected cycling network with dedicated cycling facilities such as bike lanes, cycling tracks and multiuse trails for improved cycling experience and ultimately increased share of active travel modes.

ACKNOWLEDGMENT

This research was supported through a project funded by Ontario Centres of Excellence (OCE), Brisk Synergies, and the Natural Sciences and Engineering Research Council of Canada (NSERC). The authors gratefully acknowledge Brisk Synergies for their support in data collection.
REFERENCES


(22) City of Toronto, *Toronto Cycling App*, 2016, [Online; accessed 5-July-2016].


