International Journal of Transportation Science and Technology 8 (2019) 161–175

Contents lists available at ScienceDirect

International Journal of Transportation Science and Technology

journal homepage: www.elsevier.com/locate/ijtst

Metropolis-Hasting based Expanded Path Size Logit model for cyclists' route choice using GPS data

Anae Sobhani [a,*], Hamzeh Alizadeh Aliabadi [b], Bilal Farooq [c]

[a] Laboratory of Innovations in Transportation (LITrans), Department of Civil Engineering, Ryerson University, 341 Church Street, Toronto, Ontario M5B 2M2, Canada
[b] Department of Civil, Geological and Mining Engineering, École Polytechnique de Montréal, C.P. 6079, succ. Centre-Ville, Montréal, Québec H3C 3A7, Canada
[c] Canada Research Chair, Disruptive Transportation Technologies and Services, Laboratory of Innovations in Transportation (LITrans), Department of Civil Engineering, Ryerson University, 341 Church Street, Toronto, Ontario M5B 2M2, Canada

ARTICLE INFO

Article history:
Received 7 February 2018
Received in revised form 1 November 2018
Accepted 13 November 2018
Available online 28 November 2018

Keywords:
Route choice
GPS data
Expanded Path Size Logit
Metropolis-Hastings
Cyclists

ABSTRACT

This study contributes to the field of cycling route choice by adopting the unprecedented combination of the Metropolis-Hastings (MH) path-sampling algorithm and the Expanded Path Size Logit (EPSL) model. The MH sampling approach is used to generate 15 alternative route choice sets for cyclists. The EPSL multivariate route choice framework is utilized to account for the correlation between sampled and non-sampled alternatives (joint MH-EPSL model). The data used in this paper is drawn from GPS data collected by the City of Toronto using a custom-built smartphone application in 2014–2015. The study focuses on non-work-related cycling trips (shopping, leisure, social and others) in downtown Toronto on weekdays.

The estimated results indicate that the presence of bicycle lanes and road medians attractions and number of trees along the path have a positive impact on cyclist route choice. In general, cyclists prefer to take shorter routes on lower speed roads with less public transit stops especially during the evening rush hour, and less willing to take one-way streets, local roads, and steep road segments. These findings are useful to policy makers as well as transportation and urban designers when developing a cycling network aiming to attract more cyclists. Finally, our results indicate that the MH-EPSL model performance is an appropriate framework to study cyclists' route choice decisions.

1. Introduction

Canadian auto ownership has increased from 65% of households owning at least one automobile in 1992, to 84% in 2007 (Natural Resources Canada, 2009; Turcotte, 2005). The same pattern is observed in the U.S., where the household auto ownership increased from 80% in the early 1970s to 92% in 2001 (Polzin and Chu, 2005; Pucher and Renne, 2003). The well-established negative impacts of auto-dependency range from a decrease in economic stability and public health, to impacts on the global climate, noise pollution, urban livability and energy security (Boyle, 2005; Sener et al., 2009). Over the years, these negative impacts have lead local, regional and federal policy makers to propose different transportation strategies

Peer review under responsibility of Tongji University and Tongji University Press.
* Corresponding author at: Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, 2628 BX Delft, The Netherlands.
E-mail address: anae.sobhani@mail.mcgill.ca (A. Sobhani).

https://doi.org/10.1016/j.ijtst.2018.11.002
2046-0430/

aiming to increase and encourage more sustainable travel modes. Cycling has gained a lot of attention due to its personal, societal and environmental benefits such as improving health, alleviating traffic congestion, improving air quality, decreasing fuel consumption, and providing an affordable mode of transport (Boyle, 2005; Natural Resources Canada, 2009; Polzin and Chu, 2005; Sallis et al., 2004; Sener et al., 2009; Turcotte, 2006). Despite efforts and investments to increase bike use, the low cycling mode share has been a subject for investigation in the last decade.

Although cycling is an ideal alternative to the private car for short-distance travel (less than 1.6 kilometer), which adds up to 41% of all trips (Pucher and Renne, 2003; U.S. Department of Transportation, 2001), cycling mode share accounts for only 1.3% of the commuters in Canada and 1% of all trips in the U.S. in 2001 (National Household Traffic Survey of America, 2013; U.S. Department of Transportation, 2001). To increase cycling mode share, a comprehensive understanding of cyclist's needs and perceptions will help researchers, policy makers and transportation agencies to propose better infrastructural designs and policy regulations. To this end, route choice studies have adopted numerous modeling approaches to investigate the effects of different infrastructural, environmental, traffic, and sociodemographic variables to identify attractive features for cyclists (Cheng and Yang, 2015; Lam and Small, 2001; Papinski and Scott, 2011).

Conducting a route choice study requires extensive data collection and processing. Four elements should be considered in these studies: 1) data collection, 2) large dataset processing (such as map matching and variable derivation), 3) alternative choice set generation, and 4) data analysis (e.g. discrete choice estimation) (Papinski and Scott, 2011). In the first element, there are two forms of data that have been used in route choice analysis: stated preference data (see (Amirgholy et al., 2017)), and revealed preference data (e.g. GPS trajectories or video recordings, see (Fan and Gurmu, 2015; Papinski and Scott, 2011). Reviewing literature shows that most studies used stated preference surveys, possibly since the alternative has a higher relative cost and complex network algorithm computations (Hood et al., 2011). More recently revealed preference data is being used, with the widely available combination of smartphone applications with GPS sensors to record trace tracking, travel mode, activity purpose, and sociodemographic information, providing a comprehensive dataset which is ideal for route choice studies.

The second element of a route choice study requires dataset processing. Deriving variables from the available datasets, as well as processing the data. The third element in route choice studies includes alternative route generation. A choice set of random routes from origin to destination does not consider realistic and attractive alternatives and the preferences of individuals (Hess, 2010). This is especially true in the case of cyclists where the road network is a mix of various types of cycling facilities (shared, dedicated, counter-flow, etc.) where road characteristics such as direction, traffic flow volume and speed limit need to be considered. Route choice studies have used different types of alternative choice set generation criteria such as shortest path based on travel time or travel distance, reducing total cost, congestion, intermodal interchanges, minimizing number of turns or obstacles such as intersections, stop signs, and maximizing aesthetics (Golledge and Gärling, 2002; Papinski and Scott, 2011). In past studies, different deterministic and stochastic path generation methods based on repeated shortest path algorithm were used for generating alternative choice sets such as link-labelling, link-elimination, and link-penalty methods (for more details see (Frejinger and Bierlaire, 2010; Prato, 2009)). These studies made use of a sampling approach to compare the chosen path to a set of alternative paths available to the cyclist. Ignoring sampling probability of every alternative in the universal set, or assuming equal sample probability are the main shortcomings of these methods which lead to biased model estimates. To overcome this issue, Frejinger et al. applied a biased random walk to sample a subset of paths and calculated a sampling correction factor to obtain unbiased estimated results (Frejinger et al., 2009). Furthermore, Flötteröd and Bierlaire proposed a more efficient method using the Metropolis-Hastings (MH) algorithm to create sample choice sets and arbitrary probability distributions for each alternative (Flötteröd and Bierlaire, 2013). The MH framework considers the road network and definition of path weight as inputs and the Markov Chain process is employed to sample feasible alternatives and also calculate their probability (Flötteröd and Bierlaire, 2013).

Finally, the fourth element requires detailed analysis of the results. Most of the earlier studies in this field adopted descriptive analysis with small sample sizes and neglected to use multivariate analysis (Hood et al., 2011; Sener et al., 2009). Although descriptive analysis of stated and revealed preference datasets provides valuable information, the application of econometric modeling frameworks will provide more accurate evidence of route choice preferences. Cyclist route choice modeling gained considerable attention in the past decade where studies have been employing several different frameworks and increasingly detailed evaluation methods. Earlier studies used C-Logit and Path-Size Logit (PSL) models in their cycling route choice analysis (see (Ben-Akiva and Bierlaire, 1999; Cascetta et al., 1996)). These modeling approaches focus on the effects of attributes related to the whole trip. In these models, the similarity issue between alternative routes is addressed by adding a correction term to the deterministic part of the utility function. However, the applied correction factors in these models account only for similarities between the considered set of routes. To overcome this limitation, Frejinger et al. proposed the Expanded Path Size Logit (EPSL) model to account for correlations between sampled and non-sampled alternatives (Frejinger et al., 2009). In this approach, a correction factor is applied to the sum of the number of paths using a particular link.

Since cyclists are observed to behave differently in different conditions such as at different times of the day (rush hour and off-peak hours), during weekdays and weekends, for different trip purposes, etc., route choice studies evaluate these datasets separately. Most previous studies on travel behaviour and trip patterns limited their study to work related trips (commute to work and school) (Ben-Akiva and Lerman, 1985; Mannering, 1989; Swait and Ben-Akiva, 1987), and only a few studies focused on non-work related trips (see (Bhat et al., 1999; Boarnet and Crane, 2001; Boarnet and Sarmiento, 1998; Handy, 1992; Reilly, 2002; Sobhani et al., 2013)). Bhat reported that 75% of the daily urban trips in the San Francisco

Bay area are taken for non-work related activities, which is expected to increase as cities expand and people's lifestyle changes (see (Bhat, 1998; Lockwood and Demetsky, 1994)). The flexible nature of non-work trips compared to work related trips, significantly affects the general travel pattern and affects urban traffic and emission, since it brings uncertainty in travel behavior Given the high proportion of these trips, it is important to study the travel behaviour and route choice preferences of individuals traveling for non-work purposes. Furthermore, the U.S. Department of Transportation National Household Travel Survey indicates that more than 72.5% of household trips are taken during weekdays (Lockwood et al., 2005; Parsons Brinkerhoff Quade and Douglas Inc.m, 2000; U.S. Department of Transportation, 2001). Since the temporal and spatial characteristics of weekday and weekend trips are different, their individuals' travel behaviour and route choice is usually studied separately. For instance, Lockwood et al. reported that not only is the weekend travel peak period during the midday, but also the trip lengths are longer compared to peak periods in weekdays (Lockwood et al., 2005). Additionally, since Toronto is one of the cultural and tourist centers in Canada, weekend sporting or cultural events that are mostly located in the downtown area could result in traffic conditions not usual to weekdays such as higher pedestrian volumes, different locations and times for congestion, higher or lower traffic volume in certain areas (Sall and Bhat, 2007; Sobhani et al., 2013).

The factors mentioned above motivate the focus of our research to analyze non-work-related weekday cyclist route choice in downtown Toronto. A large-scale GPS-based travel survey is used as well as Toronto's georeferenced road and cycling network databases. The GPS cyclist trajectory data includes information on route, trip purpose, travel date and time. For modeling purposes, the MH sampling algorithm along with the EPSL model are adopted together, which to best of our knowledge has not been used together for cycling route choice analysis. Our research effort contributes to cyclists' route choice literature by considering a comprehensive set of individual and route-based attributes and evaluating the effects of these variables on cyclist route choice. In the current study, cyclists choose among a maximum of fifteen route alternatives generated by the MH algorithm in addition to the observed route, each with their set of infrastructural and physical variables.

The remainder of the paper is organized as follows. The next section provides a background of earlier studies in cycling route choice modeling and highlights the current study in context. Section 3 describes the methodology detailing the sampling algorithm and modeling methodology used for data analysis. Section 4 outlines the preparation and analysis of the sample data. The empirical results, cyclist's route choice baseline utility profile based on the estimated parameters, and model validation are presented in Section 5. The final section concludes the research findings and presents the study limitations and future work.

2. Earlier studies and current study in context

2.1. Background

There is a large body of literature in micro-economics, behavioural science, psychological, and behavioural travel patterns of individuals, aiming to understand the essentials of individuals' decision-making process. Consequently, several modeling frameworks have been proposed to simulate travelers' route choice behaviour. This section briefly summarizes earlier research on route choice, common variables and models employed for studying cyclist preferences.

Sener et al. categorized earlier cyclist route choice research into two scaled level classes: 1) the aggregated-level studies which analyse the effects of route attributes by aggregating bicycle use factors on cyclists' route choice decisions or by cross-comparing bicycle level of service between different cities or regions (Forester, 1996; Moritz, 1997), and 2) the disaggregate-level studies at the decision makers' level (Sener et al., 2009). Unlike the former group, the interpretations of the latter studies denote essential cyclist's preference. In other words, disaggregate-level studies are able to capture the underlying cyclist route preferences and their mainsprings (Koppelman and Bhat, 2006; Sener et al., 2009). That said, few studies have made use of econometric models to evaluate the elasticity and trade-offs between route variables, while many simply adhered to descriptive analysis methods (see (Antonakos, 1994)).

Among the studies adopting and econometric framework, the multinomial logit model (MNL) is the most popular model used in literature. However, this model in its traditional form is not suitable for route choice analysis especially for cyclists, since it neglects the relationship between routes with shared road segments (Dhakar and Srinivasan, 2014). Over the years, new models have addressed this shortcoming, for example Bekhor and Prato used a modified version of the MNL model which outperformed the traditional MNL especially in the of presence of a large number of alternatives (Bekhor and Prato, 2009). Furthermore, models proposing changes to the deterministic component have been used such as the C-logit, PSL, Path Size Correction Logit (PSCL), and EPSL (see (Dhakar and Srinivasan, 2014) for more detail on these models). The traditional PSL model has been used in recent cyclist route choice studies each employing different methods for choice set generations. For example, Broach developed a modified method of route labeling by maximizing individual criteria while applying multiple distance constrain values (Broach et al., 2012). A study by Hood, Sall and Charlton investigated cyclists' route choice using GPS data collected in San Francisco by combining a stochastic path generation and labeling method developed by Bovy and Fiorenzo-Catalano called a doubly stochastic method (Bovy and Fiorenzo-Catalano, 2007; Hood et al., 2011), and adopted a Path Size Multinomial Logit framework to evaluate the effects of variables on cyclist route choice. To address the shortcomings of these models in route choice evaluation, studies took into account errors across alternatives by conducting generalized extreme value-based models with closed form probability functions (e.g. cross nested logit) or mixed models without closed-form probability functions (e.g. mixed logit) (Dhakar and Srinivasan, 2014). Comparing these two approaches indicates that the computation of C-logit, PSL, PSCL, or EPSL models are more straightforward in general; however, to estimate error components in the latter group, more complex probability functions were used which are time consuming especially with a large number of observations in a dense and large road network which is the case in most transportation studies (Bovy et al., 2008). An application of the sequential link choice method has been applied by Fosgerau et al. in the context of a recursive logit model, which can be consistently estimated without the need of path sampling (Fosgerau et al., 2013). The link-based formulation suggests that drivers have a link-by-link perception of the network and their choices are based on link-level attributes. Since a real network consists of a large number of links, the estimation of these models could be very computationally expensive (Alizadeh et al., 2017; Fosgerau et al., 2013). A different framework has been slowly growing in the past decade for route choice analysis: random regret minimization model (RRM). These models assume that users aim to minimize the experienced regret when considering alternatives (see (Chorus, 2012; Li and Huang, 2017)).

The attributes analysed in route choice vary among studies. Some studies examined travel time delay with regard to free-flow (see (Bekhor and Prato, 2009) for more detail). These studies, mainly focused on the effect of delay on route choice decisions (Dhakar and Srinivasan, 2014). Moreover, some studies examined a route's physical and functional characteristics in their analysis for decision making process (e.g. number of turns, number of speed bumps). Chen et al. examined participants' travel behaviour by taking into account the effects of several route choice criteria assigning criterion weights to these attributes without assuming independency between them (Chen et al., 2001). Most studies in this field were not able to generate a series of essential variables (due to data availability), and also focused solely on bicycle facility characteristics or cycling network continuity as determining factors for cycling route choice decisions. Exceptions to this pattern include Sener et al. who employed six extensive categories of variables including the cyclist's characteristics, on-street parking, bicycle facility type, road physical characteristics, road functional characteristics (traffic volume and speed limit), and road operational characteristics (Sener et al., 2009). Although, distance and travel time are most frequently considered in route choice studies, surprisingly few studies examined the impact of these attributes in the cyclist's context. Studies that did consider the two variables have identified them to be one of the important attributes in cyclists' route choice especially for commuters (see (Hunt and Abraham, 2006; Sener et al., 2009)). Additionally, other than the mentioned study by Sener et al., there has not been any study in the past on cyclists' route choice behaviour that took into account sensitivity variations across cyclists (latent class variables) such as perception of individuals over time (time-conscious versus time-relaxed) (Sener et al., 2009). Also, to be best of our knowledge, only few studies looked into individual and trip level characteristics together as potential attributes that might have an effect on cyclist's route choice decisions such as trip purpose, time of day, day of the week, and cyclist attributes (Dhakar and Srinivasan, 2014).

2.2. The current research effort

The summary of literature presented in the preceding section highlights the recent progress in understanding cyclists' route choice decisions. The current study identifies different attributes that might have significant effects on cyclists' route choice from a Canadian perspective where the road and cycling network as well as cyclist behaviour may differ from other locations. For the first time, to be best of our knowledge, a joint MH-EPSL framework is employed to is employed on cyclist route choice behaviour. The sample in our study is drawn from GPS data collected by the City of Toronto using a custom-built smartphone application in 2014–2015. Downtown Toronto has the third highest concentration of businesses and skyscrapers in North America, and is densely packed with mixed-use residential, commercial, parks, government and other land use areas. The majority of Toronto's cycling facilities are located in the downtown area (City of Toronto, 2017). Also, most non-work activities such as shopping, sporting, cultural, and musical events are located in the downtown area.

The model estimation is performed on a sample of 500 observations and the results are validated based on a holdout sample of 229 observations not considered in the choice sampling. In addition to the several route attributes, road, individual and origin-destination (OD) location factors are also included in the model estimation.

3. Methodology

This section presents the Metropolis-Hastings (MH) algorithm used for choice set generation based on Flötteröd and Bierlaire's (2013) work, and the Expanded Path Size Logit model (EPSL) based on Frejinger et al. (2009) as the econometric framework employed for the analysis. The EPSL model was programmed in Python Biogeme, while Bioroute was used for the MH sampling.

3.1. Choice set generation

The MH method is adopted to generate an un-normalized form on the Markov Chain (MC) with a predefined stationary distribution to address the enumeration of paths (Flötteröd and Bierlaire, 2013). This algorithm requires a road network and a definition of path weight as an input. The underlying Markov Chain process samples alternatives and calculates their sampling probability without the need to normalize over the full choice set. The MH sampling algorithm starts with an arbitrary path, such as the shortest path between an OD pair, and makes random modifications to the path, which are accepted or rejected based on the known probability of the modification.

Let us consider h as the iteration counter, a^h as an arbitrary initial state, d as a candidate state, $e(a^h, d)$ as an irreducible proposal distribution which defines the probability of proposing a transition from state a^h to state d, $\varsigma(a^h, d)$ as the acceptance probability which is specified such that the desired stationary distribution is attained. The predefined stationary distribution can be defined in un-normalized form through positive weights $\{g(a)\}$ $a \in O$ where O is the MC's finite state space and g(a) is proportional to the stationary probability of state $a \in O$.

The steps of the algorithm are (Flötteröd and Bierlaire, 2013):

1. Set iteration counter h = 0
2. Select arbitrary initial state ah
3. Repeat beyond stationarity:

   Draw candidate state d from $\{e(a^h, d)\}_d$

   Compute acceptance probability $\varsigma(a^h, d) = min\left(\frac{g(d)e(d,a^h)}{g(a^h)e(a^h,d)}, 1\right)$

   With probability $\varsigma(a^h, d)$, let $a^{h+1} = d$; otherwise let $a^{h+1} = a^h$

   Increase h by one

Based on the steps in this algorithm, we are able to compute the *e(a,d)/e(d,a)* for every proposed transition *a* to *d* in an efficient manner automatically eliminating the probability of self-loops (Flötteröd and Bierlaire, 2013). The observed chosen route of the cyclist is also added to the choice set along with the paths generated by the MH algorithm (Dhakar and Srinivasan, 2014; Elgar et al., 2015; McFadden, 1978; Nur Arifin, 2012).

The MH algorithm is adopted to generate a maximum of fifteen feasible alternatives per observation and the observed route is added to the choice set (maximum of sixteen in total) (see (Bekhor et al., 2008; Bovy and Fiorenzo-Catalano, 2007; Elgar et al., 2015) for extensive discussion on choosing an optimum sample size). To apply this algorithm on downtown Toronto's dense road network, 298 separate input files were prepared to provide the possibility of parallel calculation of paths. Files are imported into a cluster of 26 computers (2 processors Intel(R) Xeon(R) X5675 @ 3.07 GHz) taking up to 880 min to generate the output files. The generated alternatives choice sets are then added to the data sample for estimation.

3.2. Econometric model structure

The EPSL model has a correction factor Expanded Path Size (EPS) in the utility function that applies to each alternative with all the possible paths in the true choice set. That said, this factor, unlike in the PSL model, takes into consideration the correlation between non-sampled paths (for more details see (Frejinger et al., 2009)). In the traditional path size modeling framework, the choice sets contain all possible paths connecting each OD. The similarity issue between alternatives is addressed by adding a correction term to the deterministic part of the utility function, which alters the utility of paths based on their similarities. However, the applied correction factor in this model accounts only for similarities between the considered set of paths (see (Frejinger and Bierlaire, 2007)). To address this, the EPS correction factor corrects the alternatives' path utilities based on the sampling prototype for estimating asymptotically unbiased parameters affecting route choice (Frejinger et al., 2009).

Let us consider n as an individual who chooses path i, μ as the scale factor, Γ_i as the set of links in path i, Z_{kn} as the empirical frequency or the actual number of times path k is drawn, L_b as the length of path b and L_i the length of path i. The equation $\sum_{k\in\varphi_n}\rho_{bk}\omega_{kn}$ refers to the number of paths in the considered choice set (φ_n) using link b. And so, $\rho_{bk} = 1$ if path k contains link b, and $\rho_{bk} = 0$ otherwise. ω_{kn} is the expansion factor defined in Eq. (2). Finally, R_n indicates the total number of paths drawn with replacement from the universal choice set (C_n) and $q(k)$ is the sampling probability of path k. The correction factor EPS can be defined as Eq. (1) (Frejinger et al., 2009):

$$EPS_{in} = \sum_{b\in\Gamma_i}\frac{L_b}{L_i\sum_{k\in\varphi_n}\rho_{bk}\omega_{kn}} \tag{1}$$

where ω_{jn} is defined by:

$$\omega_{kn} = \begin{cases} 1, if \ \ \rho_{bk} = 1 \quad or \quad (q(k)R_n) \geq 1 \\ \frac{1}{q(k)R_n}, Othewise \end{cases} \tag{2}$$

Once the EPS factor is defined, the EPSL's conditional probability for individual n choosing path i from the universal choice set (C_n) can be defined as Eq. (3) (Frejinger et al., 2009):

$$\mathrm{P_{EPSL}(i|C_n)} = \frac{\mathrm{e}^{\mu(\mathrm{V_{in}+lnEPS_{in}})+\ln\left(\frac{\mathrm{Z_{in}}}{\mathrm{q(i)}}\right)}}{\sum_{\mathrm{k\epsilon C_n}}\mathrm{e}^{\mu\left(\mathrm{V_{kn}+ln(EPS_{kn})}\right)+\ln\left(\frac{\mathrm{Z_{kn}}}{\mathrm{q(j)}}\right)}} \tag{3}$$

Finally, V_i represents the value term of the utility which can be measured by s number of exogenous variables (Xsi) and their relative estimated coefficients (βs) (Eq. (4)) (Frejinger et al., 2009):

$$V_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_s X_{si} \quad (4)$$

4. Empirical analysis

4.1. Data sources and sample formation

The data for this study is drawn from the *Toronto Cycling* smartphone application with GPS trip traces and additional information from cyclists who used the application in Toronto during 2014–2015. The application was manually activated by users and at the end of each trip they were asked to enter their trip purpose. The dataset includes trip level information of the cyclists such as reason of travel, day of the week, season, average speed, start and end time and trip duration. Other sociodemographic and individual-level attributes were also collected which were not available to us for this study. It should be noted that since participants are anonymized, the traveller cannot be identified. For this reason, all trips are assumed to be taken independently and the agent-effect cannot be considered in the model. It should be noted that this data does not represent all bicycle trips in Toronto as the participants were limited to the smartphone application users.

For analysis purposes, the sample generation exercise requires a series of transformations to the original GPS and road datasets:

- The duration of each trip is calculated and trips shorter than two minutes are removed from the data.
- The activity purposes (except work/school or commute related activities) are compiled and classified into four categories: leisure, shopping, social, and others.
- Trip characteristics such as travel time, time of day, day of the week (weekends versus weekdays), and season are appended to the database.
- In the process of data assembly, the origin and destination of trips are mapped to Toronto's road network (obtained from Open Street Map), and trips with the same start and end location are deleted from the data due to chain effects which is not the focus of this research.
- The map-matching process for assembling GPS traces and the road network adopts a direction-based nearest link point-to-curve algorithm. Since GPS points are matched to their closest link, the results might not be accurate in dense networks. This issue is more noticeable at intersections as the mentioned method does not consider trip direction (see (Zhou and Golledge, 2006) for more details). To address this shortcoming, we associated the GPS records to their nearest links with respect to their azimuths to ensure GPS points are matched to closer links with respect to their related trip direction. At the end, the distance-based shortest-path algorithm is applied between consecutive GPS records to deduce the entire path for each trip while considering one-way streets as a restriction.
- By merging the information from various secondary data sources (Toronto open data and DMTI Spatial Inc.) link and route level characteristics are added to the GPS database (see Section 5.1).
- The database is split into two components based on whether the observed trip occurred during a weekday or weekend.
- There are 5123 observed trips in downtown, 4391 (86%) of which are during weekdays. Furthermore, 1929 of these weekday downtown Toronto trips are non-commute trips (38 % of all trips), 500 of which are chosen randomly for model estimation and analysis, and 229 trips are selected as a holdout sample for model evaluation.

4.2. Cyclists' trip data and Downton Toronto road network descriptive analysis

Toronto is the most populous metropolitan area in Canada and was ranked the fifth largest city in North America in 2011 (Statistics Canada, 2015). Our study area is in downtown Toronto between Sherbourn street and Batherst street from east and west, and Bloor street and Queens Quay from north and south (Fig. 1). This area covers 15.7 square kilometers, consisting of 6447 nodes and 2312 links; 53.1% of the area's land use is residential, 5.3% is commercial, 29.7% is governmental and industrial, 6.6% is parks and green spaces and the rest is other types of land use. In this area, 7.7% of the 232.6 road kilometers are highways, 31.5% are major roads (arterials), and 60.8% are local roads. Moreover, there are only 34.0 kilometers of bicycle facilities in this area covering 17.3% of the road network. The bicycle network in this area consists of 47.9% painted bike lanes[1], 29.9% are physically separated cycle tracks[2], and 22.2% are assigned to shared roadways (sharrows[3]). One-way streets cover 40.4% of the network which is a big constraint for road users and might lead to longer routes. Moreover, 44.3% of the roads' speed limit in downtown Toronto are more than 60 kilometers per hour which can affect the route choice of cyclists who have safety concerns.

[1] Bike lanes are panted lanes on the road separating bikes and motor vehicles with the exception when vehicles have to cross the painted lane to park or enter a driveway.

[2] A cycle track is physically separated from motor traffic along the road.

[3] Sharrows are designated roadways where cyclists and vehicles share the same space and are not separated by any means. Itis usually indicated by a painted bicycle on the road.

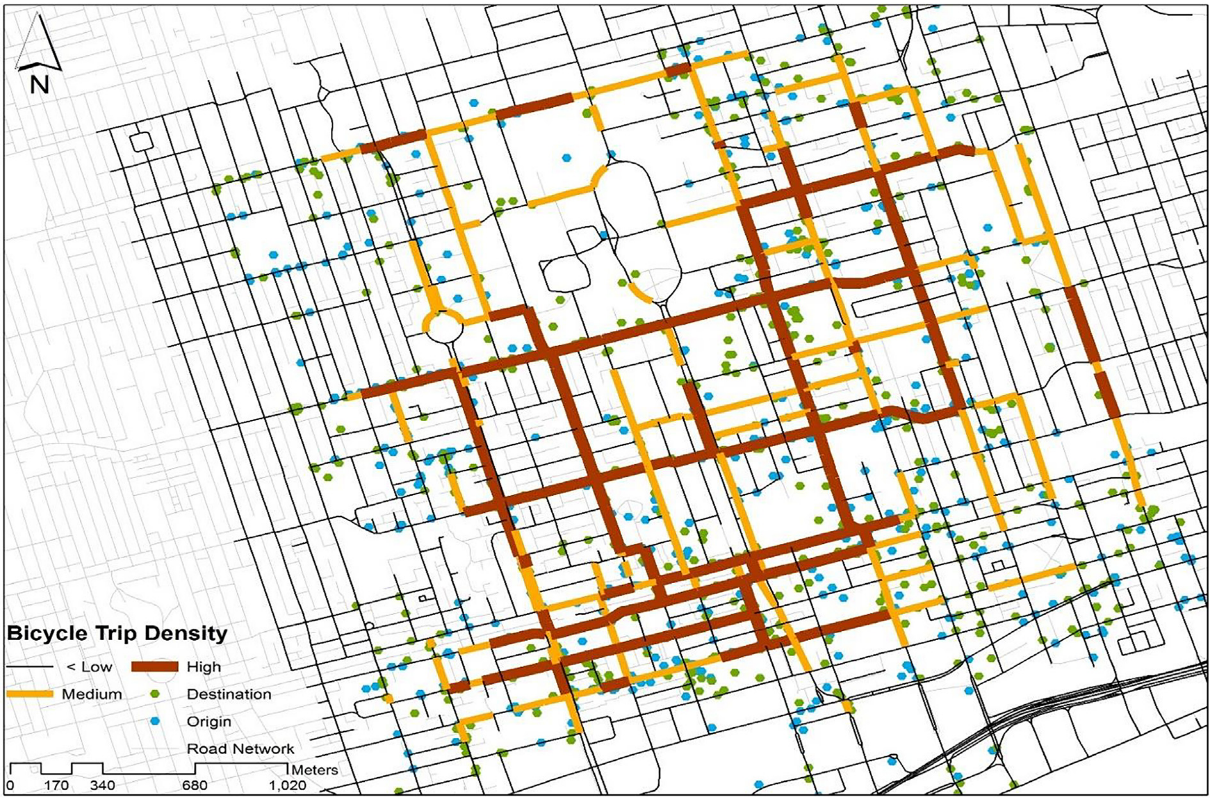

Fig. 1. Study area indicating observed high demand roads for cyclists.

The observation of Torontonian cyclists' profiles (Tables 1 and 3) indicates that cyclists traveled on average 743.6 meters during weekdays to participate in non-work activities. Among the activity types, the most common reason of travel is the "other" non-work activity type (49.2%) followed by shopping (22.6%) and social (22.6%) activities. Cyclists in downtown Toronto traveled longer for leisure activities (805.3 meters) while less for shopping (670.6 meters), indicating that the downtown area has a well-mixed land use with shopping, restaurants, sport centers, theaters and other recreational locations that are accessible in less than one kilometer. Looking at the time distribution of activities' start time indicates that 53.2% of the observed trips started after 8 PM. This is reasonable since during weekdays, participants prefer to be involved in non-work activities after business hours. Furthermore, 53.6% of the reported weekday non-work-related trips occurred during the summer season, and 4.2% in the winter which is expected. Investigating the number of turns shows that only around 15% of the cyclists, used straight routes to get to their destinations (no turns) which is expected since downtown Toronto has a dense network where turns are inevitable.

Further, analyzing the participants' observed routes shows that 81.2% of the trips used bicycle facilities along their routes, where 77.8% of cyclists rode on major roads such as arterials and collectors (See Table 3).

4.3. Variable specification

Several types of variables are considered in this study which are selected based on past cyclist route choice studies (see Section 2.1 for more details). The independent variables are classified into four categories (see Tables 2 and 3):

1) road-level physical attributes: road type, speed limit, road segment elevation, bicycle facility type, bicycle racks, intersections, one-way streets, presence of median, trees, public transit stop type, attraction places[4] and land use type along road segments;
2) route level attributes: route length and number of turns;
3) individual level attributes: reason of travel, activity start time, and season; and
4) origin and destination location: distance from central business district (CBD).

[4] Places of attraction include museums, the beach, historic buildings, parks, waterfront port, urban open space (e.g. Dundas square), and theaters.

Table1
Downtown Toronto cyclists' sample GPS observation rate and length.

Variables		Trip rates (%)	Average length (meter)
Activity type	Leisure	5.60	805.25
	Shop	22.60	670.62
	Social	22.60	794.83
	Other activities	49.20	746.53
	Total length	100.00	743.58
Activity start time	0 to 8 AM	10.60	751.43
	9 AM to 2 PM	12.60	773.33
	3 PM to 7 PM	23.60	729.42
	8 PM to 11 PM	53.20	745.04
Season	Winter	4.20	676.31
	Spring	13.60	832.51
	Summer	53.60	717.78
	Autumn	28.60	759.50
Bicycle facility type	Bike lane	49.80	37.49
	Bike track	36.80	35.36
	Sharrow	45.60	17.74
	Total bike segments	81.20	48.98
Road type	Major roads	94.80	56.26
	Local roads	77.80	43.74

For variables in category one and two besides estimating the impact on different bicycle route alternatives (sixteen alternatives), deviations of reasonable variables from all four categories are evaluated through interaction variables, for example leisure * local, and 9 AM to 2 PM * length, etc.

5. Model results

5.1. Model parameters

Table 4 presents the parameter estimates corresponding to exogenous variables affecting the baseline utility specification for downtown Toronto cyclists who participated in no-work and non-commute activities during the week. Various combinations of the mentioned three attribute categories are tested (e.g. combination of leisure activity and local road type). It should be noted that in the estimation process, observed route is considered as the base alternative.

5.1.1. Scale and correction parameters

The EPSL scale (μ) and EPS correction parameters estimated in our model are significant (Table 4). Hence, the sampling correction is valid, and the parameter estimates are asymptotically unbiased. Based on these parameters and the results presented in our study, the importance of correcting the utilities are well established for sampling and employing MH algorithm (which also provides sampling correction) (see supporting results in (Frejinger et al., 2009)).

5.1.2. Estimated parameters and discussion of results

Two attributes are used to define route level characteristics: route length and number of turns. The findings presented in Table 4 indicate a coefficient of -1.36 for the route length, reflecting a preference for shorter distance trips to participate in all non-work activities (see (Sener et al., 2009) for similar results). Reviewing the literature shows that this attribute has the same affect for both non-work and work related trips. (see (Broach et al., 2012; Tilahun et al., 2007) for similar results in work related trips). Interaction effects of the mentioned variable with activity type, distance of trip origin from CBD, season, presence or number of trees, attractions along route, time of day, presence of median, presence and type of bicycle facilities and presence of public transits stops are also considered. Among the mentioned combined variables, it is observed that during the summer, cyclists travel longer routes which is reasonable since weather conditions and activities justify longer journeys (see same travel pattern for commute and work related trips in (Tilahun et al., 2007)).

Analysing the distance of trip origin from CBD along with route length shows that cyclists whose trips originate further from CBD take the shortest route, which is reasonable since trips originating closer to the CBD have more restrictions in the dense CBD area such as avoiding high traffic volume and congestion, more stops, higher number of turns etc. The results corresponding to the time of the day in association with route length shows a preference for longer paths when departing in the morning and afternoon hours (9 AM to 2 PM) compared to late evening (8 PM to 11 PM). This is intuitive, since most people are tired during late evening period and would prefer to take other modes for longer trip lengths; while in the morning they are fresh and enthusiastic toward traveling longer trips for their non-work activity. Safety concerns during nighttime (reduced visibility due to lack of light and presence of drunk drivers) can be an explanation for this behaviour as well.

Table 2
Attributes chosen for model estimation.

Attributes		Attribute Levels	Description
Road-level physical attributes	Road type	*Local road* *Arterial*	Length/total path length
	Bicycle facility type	*Bike lane* *Cycle track* *Sharrow* *No cycling facility*	Length/total path length
	Speed limit	*Average* *Maximum*	Average and Maximum in each link in the total path
	Road segment elevation (considering segment length)	*Average* *Maximum* *Minimum*	Average, Maximum, and Minimum of the grade link length/the total path
	Bicycle rack occupancy		Number of bike racks/total path length
	Number of intersections		o Number of intersections o Number of intersections/total path length
	Number of one-way streets		o Number of one-way streets/total path length o length of one-way streets/total path length
	Presence of median		Dummy variables based on if there is any medium in the path
	Presence and number of trees		o Dummy variables based on if there is any trees in the path o Number of trees/ total path length
	Number of public transit stops		Number of stops/total path length
	Number of attraction places		Number of attractions/total path length
	Land use type	*Commercial* *Residential* *Parks and green spaces* *Governmental and industrial*	Area of land use type/ area of the path in the 5 m buffer
Route level attributes	Length Number of turns		Sum of the length of the links in the path Number of turns/total path length
	Reason of travel	*Leisure* *Shop* *Social* *Other activities*	Dummy variables based on the generated activity type
	Activity start time	*12 AM to 8 AM* *9 AM to 2 PM* *3 PM to 7 PM* *8 PM to 11 PM*	Dummy variables based on the generated time slot
Individual-level attributes	Season	*Winter* *Spring* *Summer* *Autumn*	Dummy variables based on the generated season

Furthermore, the implication of activity type and trip length interaction shows that trip length does not differentially impact cyclist's route choice for different non-work activity types (for similar results see (Dhakar and Srinivasan, 2014)).

The second variable in this attribute category is the number of turns per meter. The positive coefficient corresponding to this variable surprisingly underlines the preference among cyclists for trips with more turns (see (Bailenson et al., 1998; Golledge and Gärling, 2002; Papinski and Scott, 2011) for different results). This could be because of the dense downtown grid road network and also presence of many one-way streets along the routes which leads to less straight-like paths in the study area. That said, the exact reasons for this impact requires further investigation. On the other hand, results from literature on commute and work related trips shows that cyclists prefer paths with less turns in general (e.g. see (Broach et al., 2012)).

Table 3
Descriptive information of the bike trips' sample in Downtown Toronto.

Attributes		Attribute levels	Mean	Std. Dev.	Minimum	Maximum
Road-level physical attributes	Road type	*Local road*	0.44	0.29	0.00	0.97
		Arterial	0.56	0.28	0.03	1.00
	Bicycle facility type	*Bike lane*	0.37	0.26	0.00	1.00
		Cycle track	0.35	0.27	0.00	1.00
		Sharrow	0.18	0.21	0.00	1.00
	Speed limit	*Average*	53.96	6.23	40.00	90.00
		Maximum	57.3	7.12	40.00	90.00
	Road segment elevation (considering segment length)	*Average*	-11.41	81.57	-533.06	400.09
		Maximum	257.21	304.60	0.00	2068.98
		Minimum	-376.75	568.02	-3687.51	0.00
	Bicycle racks occupancy		0.09	0.04	0.00	0.40
	Number of Intersections		0.009	0.003	0.002	0.020
	One-way streets		0.41	0.29	0.01	1.00
	Presence of median		0.17	0.08	0.09	0.43
	Number of trees		0.24	0.08	0.03	0.64
	Number of public transit stops (considering path length)		0.002	0.001	0.000	0.008
	Number of attraction places		0.003	0.002	0.000	0.013
	Land use type	*Commercial*	0.09	0.17	0.00	0.59
		Residential	0.26	0.46	0.00	0.72
		Parks and green spaces	0.06	0.03	0.00	0.13
		Governmental and industrial	0.59	0.16	0.00	0.91
Route level attributes	Length		743.58	390.84	151.57	2312.76
	Number of turns over the total length		0.004	0.002	0.001	0.011
Individual-level attributes	Reason of travel * Length	*Leisure*	805.25	448.31	212.66	1779.20
		Shop	670.62	318.10	157.78	1545.35
		Social	794.83	413.60	158.52	1862.57
		Other activities	746.53	400.50	151.57	2312.76
	Activity start time * Length	*12 AM to 8 AM*	751.43	451.16	158.52	2286.23
		9 AM to 2 PM	773.33	377.18	151.57	1545.35
		3 PM to 7 PM	729.42	350.40	169.23	1780.43
		8 PM to 11 PM	745.04	401.47	157.78	2312.76
	Season * Length	*Winter*	676.31	280.43	164.73	1080.38
		Spring	832.51	422.04	164.30	1780.43
		Summer	717.78	370.14	157.78	2312.76
		Autumn	759.50	422.02	151.57	2286.23

Analyzing the effect of speed limit shows that cyclists are less willing to ride along high-speed roads. This is justified since 70% of the cycling network is not physically separated from motor vehicles meaning cyclists are sharing the road with vehicles and their safety perception is affected by vehicle speeds (for similar results see (Akar and Clifton, 2009; Caulfield et al., 2012; Habib et al., 2014)).

Estimating the effects of presence of attractions and number of trees along the route indicates that Torontonian cyclists prefer to take routes with more attractions and green area.

Another attribute that has a significant effect on cyclist route choice is the number of public transits stops along the route. Results show that cyclists are less willing to ride on paths with more public transit stops especially during the evening rush hour compared to late evening (8 PM to 11 PM). This is reasonable since cyclists do not want to be restricted or stopped by streetcars especially during the evening rush hour when there is a higher frequency of streetcars and pedestrians getting on or off.

The presence of median on roads along the cyclists' route has a positive effect on cyclist route choice. This can be due to the fact that medians restrict left turns providing a safer environment for cyclists. Moreover, studying one-way streets shows that cyclists are less willing to take one-way streets possibly since it sets a restriction in their direction of movement.

The next set of variables in Table 4 corresponds to road type, which show a preference for arterial roads compared to local roads (see (Stinson and Bhat, 2003a) for different results for commute and work related trips). While this result may seem counterintuitive, it might be reflecting the fact that these cyclists perceive a health benefit from being able to ride at

Table 4
Effects of exogenous variables on Downtown Toronto cyclists' baseline utility in the MH-EPSL route choice model.

Attributes			Coefficient	t-statistics
EPSL parameters	*μ*		5.81	2.23
	EPS		0.06	2.23
Length (meter)	*General*		−1.36	−4.42
	Summer		0.50	1.39
	9 AM to 2 PM		0.54	2.49
	Origin distance from CBD		−0.12	−1.37
Number of turns (per meter)	*General*		0.20	1.84
Average of maximum speed limit (km/h)	*General*		−1.56	−1.28
Presence of attraction	*General*		0.21	1.32
Number of trees (per meter)	*General*		0.32	1.48
Number of public transit stops (per meter)	*General*	*<=2*	−0.44	−1.38
		=>3	−1.80	−1.55
	3 PM to 7 PM		−0.30	−1.28
Presence of a median	*General*		0.14	1.57
One-way street	*General*		−0.23	−1.43
Local road (base: Arterial road)	*General*		−0.88	−2.11
	Social		−0.44	−1.54
Bike lane (base: Cycle track)	*General*		0.33	1.43
Average road segment elevation (considering segment length) in the total path	*General*		−0.36	−7.21
	Bike lane		−0.13	−1.37
	Sharrow		−0.28	−1.55
	Social		−0.11	−1.21

relatively higher speeds on arterials (similar results found here (Sener et al., 2009)). Furthermore, cyclists are travelling to participate in non-work activities that are mostly located on major roads. In addition, most of the local and minor roads in downtown Toronto are one-way, therefore the preference arterials can be expected. The interaction of this variable with activity type shows that cyclists traveling for social activities are more likely to take arterial roads which may provide a faster more direct route to their destination.

In the group of road characteristics, the effect of bicycle facility type is introduced as well as their interactions with other variables. This attribute is categorized into bike lane cycle track, sharrow and links with no cycling facility as illustrated in Table 4, cyclists in our study are more willing to ride on painted bicycle lanes compared to physically separated cycle tracks (for similar results for commute and work related trips see (Stinson and Bhat, 2003a; Tilahun et al., 2007)). As supported by other studies, this can be due to the cyclist's preference to have more maneuvering room by not being limited to a cycle track and having the psychological freedom to move around as needed (see (Sener et al., 2009; Wilkinson et al., 1994) for similar results and (Forester, 1993, 1996; Pucher et al., 1999) for reasons behind this behaviour). Another explanation of the fact could be that around 48% of the cycling network is composed of bike lanes, and cycle tracks take up to 30% of the downtown Toronto cycling network.

The coefficient of average road elevation along the segment indicates that cyclists participating in non-work activities during weekdays avoid routes with steep hills (see (Bhat and Lockwood, 2004; Sener et al., 2009; Stinson and Bhat, 2003b,a) for similar results), especially when participating in social activities which is expected since cyclists prefer not to sweat or run out of breath before attending their social activity. The interaction of this variable with bike facility type shows that cyclists are willing to ride a hill if they are on a cycle track while they are less willing if they are riding on a sharrow. This could be due to the fact that when cyclists choose to ride on a safer and more comfortable space provided by the cycling facility, they are willing to overlook the road grade.

5.2. Cyclist's route choice baseline utility profile

The estimates presented in Table 4 provide an indication of how different variables influence cyclists' route choice. To illustrate how cyclist route choice preference for different road, route and individual attributes has changed, Fig. 2 presents the route choice utility profile of a synthetic cyclist by plotting the changes in the baseline utility. The cyclist is assumed to travel to participate in a leisure activity during the summer after 8 PM, traveling a straight route on major roads with no trees or attraction places along the route. Further, the syntactic cyclist prefers roads with medians but does not choose one-way streets or cycling facility.

In the analysis process, the defined variables (season, time of the day, number of turns, presence of attraction places, number of trees, number of public transport stops, presence of median, one-way street) are changed each time, and the rote choice baseline utility values for the new synthetic person is plotted. The results of these exercises are presented in Fig. 2.

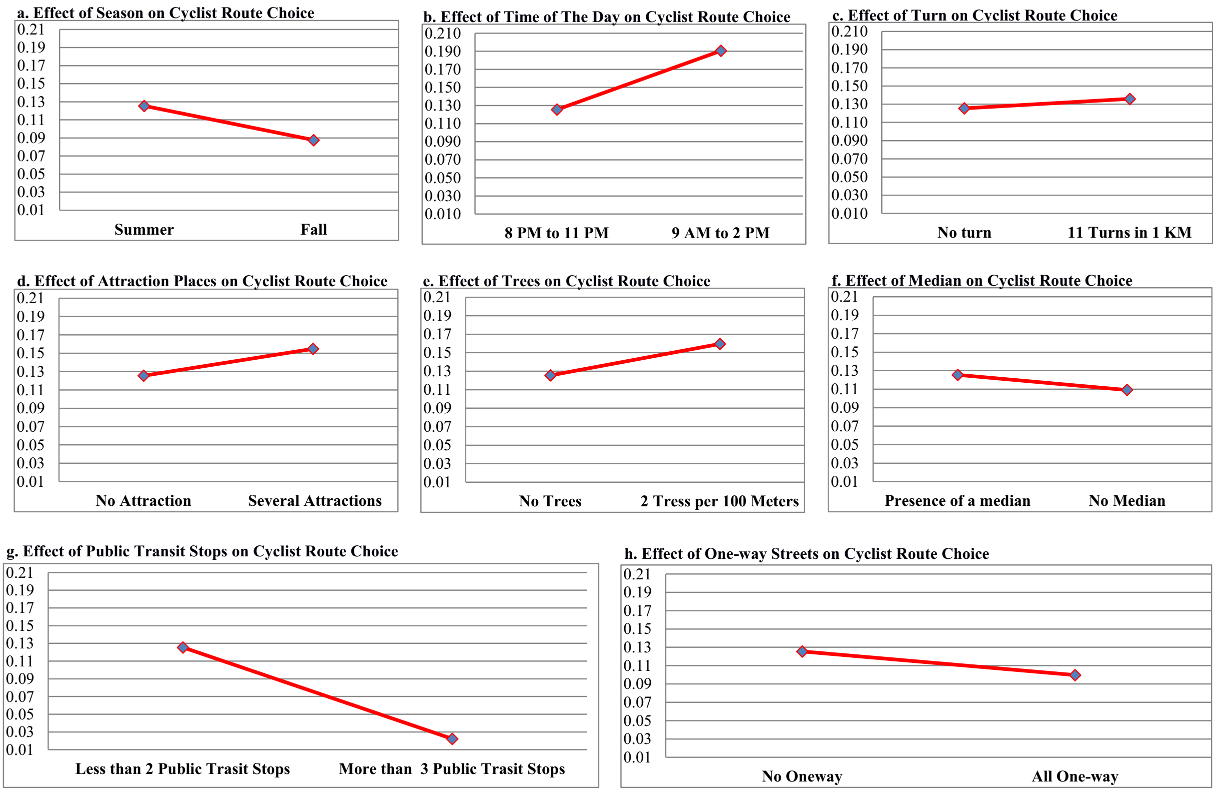

Fig. 2. Exogenous attributes' effect on cyclist route choice baseline.

Fig. 2a illustrates the baseline utility profile changes due to season. As it is observed, during the fall, the cyclist's route choice utility decreased compared to summer time. This is reasonable since during the relatively colder fall season in Canada, cycling is less desirable. Fig. 2b represents the utility profile of the synthetic cyclist based on the time of the day (night time vs. morning/early afternoon). The route choice profiles across morning/early afternoon had a positive impact on the cyclist's route choice and increased the utility (52%). Fig. 2c denotes the effects of turns on cyclist route choice. As illustrated in this figure, when the synthetic person took a route with multiple turns compared to a straight route, their route choice utility increased slightly (8%) which indicates a preference for routes with more turns.

Fig. 2d–f represent the effect of presence of attraction places, trees, and median on the road on cyclist route choice respectively. It is observed that absence of these attributes reduces the choice utility of a cyclist route. In other words, the synthetic cyclist route choice utility improved when there were more attraction places (23% percent increase) (Fig. 2d), green areas (Fig. 2e), and medians (13% percent increase) (Fig. 2f) along the route.

Finally, Fig. 2g and h illustrate the influence of number of public transport stops and one-way streets on cyclist route choice. The figures show that when the number of public transport stops and one-way street along the route increases, the likelihood of choosing that route for cycling decreases by 82% (Fig. 2g), and 21% (Fig. 2h) respectively. This denotes that among the effective variables on cyclist route choice, number of public transport stops along the route has a stronger negative effects on cyclist route choice compared to the rest of the attributes. Hence, it can be concluded that public transport stops have a stronger impact on cyclist route choice compared to other variables.

5.3. Model validation

To confirm the estimated results and ensure estimation did not over fit the dataset, the route choice prediction exercise was undertaken on a holdout sample of 229 observations. For validation purposes, the predicted route is computed based on the EPSL model estimated parameters and applied to the holdout sample. It should be noted that similar to the estimation sample, fifteen alternatives were generated using the MH approach plus the observed route for the holdout sample and the approached used by Alizadeh et al. is adopted for route prediction (Alizadeh et al., 2017). Subsequently, the calculated predicted trip results are compared to the actual observed trip in the holdout dataset (see Fig. 3).

The comparison between observed and predicted route probabilities highlights the high route choice accuracy offered by the MH-EPSL model. It should be noted that in this study area a very high proportion of the roads are one-way streets (40%), imposing a high restriction on cyclists to choose between possible routes, which was well captured in the MH route sample

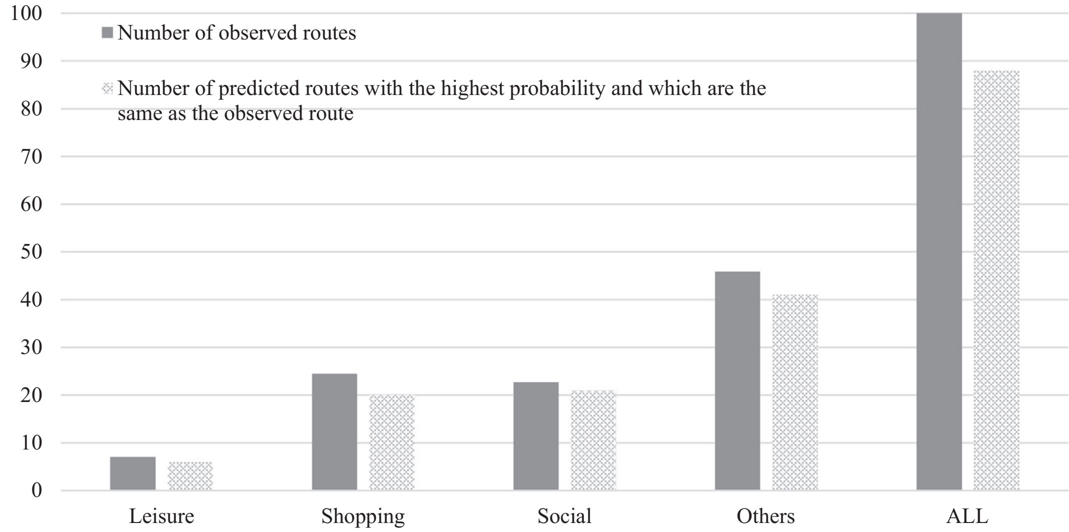

Fig. 3. MH-EPSL cyclist route choice prediction for holdout sample and reported for four activity types to compare with the observed route.

generation. Analyzing the MH generated alternative routes shows a high disparity between feasible and unfeasible alternatives given the one-way restriction i.e. the presence of these one-way streets resulted in alternatives that were too long or had too many turns. Therefore, some of the MH generated alternatives are considered unreasonable routes where by applying the EPSL model, the more reasonable routes that were chosen were in fact the actual observed route. This shows that despite this high disparity, the MH-EPSL model is very well calibrated and it is able to capture the nature of the network and chose the most feasible route among the alternatives based on the estimated results.

It can be concluded that the MH-EPSL model performance capability emphasizes the importance of bringing the concept of sampling correction in bicycle route choice modeling which leads to more consistent results with regards to cyclist's actual behaviour.

6. Conclusion

The data sample used in this paper is drawn from downtown Toronto's cyclists using the GPS-based *Toronto Cycling* smartphone application in 2014–2015 travelling to participate in non-work activities (shopping, leisure, social and others) during weekdays. The increase of non-work related cycling trips (Bhat, 1998; Lockwood and Demetsky, 1994), highlights the importance of studying their behaviour compared to the previously popular focus on commute travel behaviour. Similarly, since weekday and weekend travel behaviours are different, our study focuses on non-work-related cycling trips during weekdays since behavioural attributes may have an effect on cyclists and other road users who are traveling for work purposes. Given the large Toronto area and the limited cycling network concentrated in the downtown area, we focus our attention to cycling trips with origins and destinations in downtown Toronto (a dense area which consists of 6447 nodes and 2312 links).

A joint MH-EPSL framework is adopted to generate a maximum of sixteen route alternatives and evaluate the effect of 18 variables along with their combined interactions (around 250 feasible combinations) on cyclist route choice behaviour. The major advantage of MH sampling algorithm over conventional methods (e.g. link labelling, link elimination) is that it provides researchers with path sampling probabilities, so that model estimates based on these sets are asymptotically unbiased. The variables that may influence cyclist's route choice decisions include road physical attributes (such as road type, speed limit, segment elevation, bicycle facilities, intersections, one-way streets, etc.), route attributes (such as route length and number of turns), individual attributes (such as reason of travel, activity start time, day of the week, and season), and origin-destination location attribute (distance from CBD). The joint MH-EPSL framework which predicted the route choice with high accuracy was employed in the bicycle route choice field for the first time. The results from our study confirm the suitability of using the MH-EPSL model for cyclist route choice behaviour. Our results provide policy makers and planners with a better understanding of cyclists' behaviour which would result in more informed decisions made for implementing and improving cycling infrastructure.

The estimated results show that in general, cyclists prefer to take shorter routes on lower speed roads with less public transit stops especially during the evening rush hours. Moreover, they are less willing to take one-way streets, local roads and steep road segments. The presence of bicycle lanes and road medians as well as attractions and number of trees have a positive impact on cyclist route choice. Hence, policy makers, transportation and urban designers can incorporate these features when planning the cycling network in order to attract more cyclists.

The MH sampling algorithm proved to perform well despite downtown Toronto's dense road network, which provides the possibility of many road link combinations (routes) between each origin-destination pair, and also presence of many one-way streets (40% of the network) which resulted in alternative routes that were too long and had too many turns (i.e. high disparity between feasible and non-feasible routes). The MH-EPSL model was very well calibrated to capture the nature of the network and chose the most feasible route among the alternatives based on the estimated results. In other words, the estimated results clearly highlight the fact that sample generation algorithm combined with utility correction and scale factors (EPS and μ respectively) improve model performance. That said, the MH-EPSL model performance capability emphasizes the importance of bringing the previously overlooked concept of sampling correction in bicycle route choice modeling which leads to more consistent results with regards to a cyclist's actual behaviour. Hence, it is safe to say that sampling correction is valid, which means the parameter estimates are asymptotically unbiased.

Our study is not without limitations. The lack of more individual level information (age, gender, etc.) along with their perception variables (latent variables), level of experience, road traffic flow and parking information reduces the versatility of the impacts that can be examined in our analysis. Hence, it is possible that in the presence of these data their impacts on cyclist behaviour might offer significant inputs for policy makers. Moreover, employing the spatial and temporal transferability of the proposed framework on different datasets along with adopting multilevel nested models can be implemented in future work to explore the effects of multiple cyclists' route choice decision levels. Another appealing aspect to investigate further would be extending the Recursive Logit (RL) model proposed by Fosgerau et al. (see (Fosgerau et al., 2013)) by considering activity type, destination and start time as the three dimensions of cyclist route choice decisions. Comparison of the findings of this paper with recursive logit and Regret theory-based route choice approach by Chorus (see (Chorus, 2012)) can be interesting not only from a behavioural point of view but also from the modeling goodness of fit and performance point of view.

Acknowledgements

The authors would like to acknowledge Gunnar Flötteröd for providing the initial MH algorithm code. We would also like to thank the City of Toronto for providing us with the GPS based cycling dataset and Natural Sciences and Engineering Research Council (NSERC) for financially supporting this research project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ijtst.2018.11.002.

References

Akar, Gulsah, Clifton, Kelly J., 2009. Influence of individual perceptions and bicycle infrastructure on decision to bike. Transp. Res. Rec. 2140, 165–172.
Alizadeh, Hamzeh, Farooq, Bilal, Morency, Catherine, Saunier, Nicolas, 2017. On the role of bridges as anchor points in route choice modeling. Transportation, 1–26.
Amirgholy, Mahyar et al, 2017. An advanced traveler navigation system adapted to route choice preferences of the individual users. Int. J. Transp. Sci. Technol. 6 (4), 240–254 http://linkinghub.elsevier.com/retrieve/pii/S2046043017300345.
Antonakos, C.L., 1994. Environmental and travel preferences of cyclists. Transp. Res. Rec. 1435, 25–33.
Bailenson, Jeremy N., Shum, Michael S., Uttal, David H., 1998. Road climbing: principles governing asymmetric route choices on maps. J. Environ. Psychol. 18 (3), 251–264 http://www.sciencedirect.com/science/article/pii/S0272494498900950.
Bekhor, S., Prato, C.G., 2009. Methodological transferability in route choice modeling. Transp. Res. Part B 43 (4), 422–437.
Bekhor, Shlomo, Toledo, Tomer, Prashker, Joseph N., 2008. Effects of choice set size and route choice models on path-based traffic assignment. Transportmetrica 4 (2), 117–133 http://www.tandfonline.com/doi/abs/10.1080/18128600808685682.
Ben-Akiva, M., Lerman, S.R., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. The MIT Press.
Ben-Akiva, Moshe, Bierlaire, Michel, 1999. Discrete choice methods and their applications to short term travel decisions. Handbook Transp. Sci. 23, 5–33.
Bhat, C.R., Carini, J.P., Misra, R., 1999. Modeling the generation and organization of household activity stops. Transp. Res. Rec. 1676, 153–161.
Bhat, Chandra R., 1998. Analysis of travel mode and departure time choice for urban shopping trips. Transp. Res. Part B 32 (6), 361–371.
Bhat, C.R., Lockwood, A., 2004. On distinguishing between physically active and physically passive episodes and between travel and activity episodes: an analysis of weekend recreational participation in the San Francisco bay area. Transp. Res. Part A 38, 573–592.
Boarnet, M.G., Crane, R., 2001. The influence of land use on travel behavior: specification and estimation strategies. Transp. Res. Part A 35, 823–845.
Boarnet, M.G., Sarmiento, S., 1998. Can land use policy really affect travel behavior? A study of the link between nonwork travel and land use characteristics. Urban Studies 35 (7), 1155–1169.
Bovy, P.H.L., Bekhor, S., Prato, C.G., 2008. The factor of revisited path size: alternative derivation. Transp. Res. Rec. 2076, 132–140.
Bovy, Piet H.L., Fiorenzo-Catalano, Stella, 2007. Stochastic route choice set generation: behavioral and probabilistic foundations. Transportmetrica 3 (3), 173–189 http://www.tandfonline.com/doi/abs/10.1080/18128600708685672.
Boyle, G. 2005. ''An overview of alternative transport fuels in developing countries: drivers, status, and factors influencing market deployment, hydrogen fuel cells and alternatives in the transport sector: issues for developing countries." In United Nations University International Conference, Maastricht, Netherlands.
Broach, Joseph, Dill, Jennifer, Gliebe, John, 2012. Where do cyclists ride ? A route choice model developed with revealed preference GPS data. Transp. Res. Part A 46 (10), 1730–1740. https://doi.org/10.1016/j.tra.2012.07.005.
Cascetta, Ennio, Nuzzolo, Agostino, Russo, Francesco, Vitetta, Antonino, 1996. A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks. Transp. Traffic Theory, 697–711 http://www.alkox.informatik.hu-berlin.de/lehre/lvws0809/verkehr/logit.pdf.
Caulfield, Brian, Brick, Elaine, McCarthy, Orla Thérèse, 2012. Determining bicycle infrastructure preferences – a case study of Dublin. Transp. Res. Part D 17 (5), 413–417.
Chen, Ting Y., Chang, Hsin L., Tzeng, Gwo Hshiung, 2001. Using a weight-assessing model to identify route choice criteria and information effects. Transp. Res. Part A 35 (3), 197–224.

Cheng, Huan, Yang, Xiaokuan, 2015. Random parameter nested logit model for combined departure time and route choice. Int. J. Transp. Sci. Technol. 4 (1), 93–106. https://doi.org/10.1260/2046-0430.4.1.93.
Chorus, Caspar G., 2012. Regret theory-based route choices and traffic equilibria. Transportmetrica 8 (4), 291–305.
City of Toronto. 2017. "Toronto Cycling Network."
Dhakar, Nagendra, Srinivasan, Sivaramakrishnan, 2014. Route choice modeling using gps-based travel surveys. Transp. Res. Rec. 2413, 65–73 http://trrjournalonline.trb.org/doi/10.3141/2413-07.
Elgar, Ilan, Farooq, Bilal, Miller, Eric J., 2015. Simulations of firm location decisions: replicating office location choices in the greater Toronto area. J. Choice Modelling 17, 39–51. https://doi.org/10.1016/j.jocm.2015.12.003.
Fan, Wei, Gurmu, Zegeye, 2015. Dynamic travel time prediction models for buses using only GPS data. Int. J. Transp. Sci. Technol. 4 (4), 353–366 http://linkinghub.elsevier.com/retrieve/pii/S204604301630168X.
Flötteröd, G., Bierlaire, M., 2013. Metropolis – Hastings sampling of paths. Transp. Res. Part B: Methodol. 48, 53–66.
Forester, J., 1993. Effective Cycling. The MIT Press, Cambridge, MA. http://www.sciencedirect.com/science/article/pii/0191260786900452.
Forester, J., 1996. How to make biking a real alternative. Transp. Environ., 59–61
Fosgerau, Mogens, Frejinger, Emma, Karlstrom, Anders, 2013. A link based network route choice model with unrestricted choice set. Transp. Res. Part B: Methodol. 56, 70–80.
Frejinger, E., Bierlaire, M., 2007. Sampling of Alternatives for Route Choice Modeling, Technical Report. Transport and Mobility Laboratory, ENAC, EPFL.
Frejinger, E., Bierlaire, M., Ben-Akiva, M., 2009. Sampling of alternatives for route choice modeling. Transp. Res. Part B: Methodol. 10, 984–994.
Frejinger, Emma, Michel Bierlaire. 2010. "On path generation algorithms for route choice models." In Choice Modelling: The State-of-the-Art and the State-of-Practice: Proceedings from the Inaugural International Choice, 307–315.
Golledge, R.G., Gärling T. 2002. "Spatial behaviour in transportation modeling and planning." Chapter 3 for the Transportation and Engineering Handbook.
Habib, Khandker Nurul, Mann, Jenessa, Mahmoud, Mohamed, Weiss, Adam, 2014. Synopsis of bicycle demand in the city of toronto: investigating the effects of perception, consciousness and comfortability on the purpose of biking and bike ownership. Transp. Res. Part A: Policy Pract. 70, 67–80.
Handy, S.L., 1992. Regional versus local accessibility: neo-traditional development and its implications for nonwork travel. Built Environ. 18, 253–267.
Hess, Stephane, Daly, Andrew, 2010. Choice Modelling: The State-of-the-Art and the State-of-Practice. Emerald Group Publishing. http://books.google.com/books?id=drlO3MdyrmIC.
Hood, Jeffrey, Sall, Elizabeth, Charlton, Billy, 2011. A GPS-based bicycle route choice model for San Francisco, California. Transp. Lett. 3 (1), 63–75.
Hunt, J.D., Abraham, J.E., 2006. Influences on bicycle use. Transportation 34 (4), 453–470.
Koppelman, Frank S, Bhat, Chandra, 2006. A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models. US Department of Transportation, Federal Transit Administration.
Lam, Terence C., Small, Kenneth A., 2001. The value of time and reliability: measurement from a value pricing experiment. Transp. Res. Part E 37 (2–3), 231–251.
Li, Meng, Huang, Hai-Jun, 2017. A regret theory-based route choice model. Transportmetrica A 13 (3), 250–272 https://www.tandfonline.com/doi/full/10.1080/23249935.2016.1252445.
Lockwood, Allison, Srinivasan, Sivaramakrishnan, Bhat, Chandra, 2005. Exploratory analysis of weekend activity patterns in the san francisco bay area, California. Transp. Res. Rec. 1926, 70–78.
Lockwood, P.B., and M.J. Demetsky. 1994. "Nonwork travel - a study of changing behavior." Annual Meeting of the Transportation Research Board.
Mannering, F.L., 1989. Poisson analysis of commuter flexibility in changing routes and departure times. Transp. Res. Part B 23 (1), 53–60.
McFadden, Daniel, 1978. Modelling the choice of residential location. Spatial Interact. Theory Plan. Models 673 (477), 75–96 http://cowles.econ.yale.edu/P/cd/d04b/d0477.pdf.
Moritz, W.E. 1997. "A Survey of North American Bicycle Commuters - Design and Aggregate Results." Transportation Research Board Conference.
National Household Traffic Survey of America, 2013. Traffic Safety Facts 2011 Data. U.S. Department of Transportation, National Highway Traffic Safety Administration.
Natural Resources Canada. 2009. Canadian Vehicle Survey Summary Report. http://oee.nrcan.gc.ca/publications/statistics/cvs09/pdf/cvs09.pdf.
Nur Arifin, Z. 2012. "Route Choice Modelling Based on {GPS} Tracking Data: The Case of {Jakarta}."
Papinski, Dominik, Scott, Darren M., 2011. A GIS-based toolkit for route choice analysis. J. Transp. Geogr. 19 (3), 434–442. https://doi.org/10.1016/j.jtrangeo.2010.09.009.
Parsons Brinkerhoff Quade and Douglas Inc. 2000. "Comparative Analysis of Weekday and Weekend Travel with NPTS Integration for the RT-HIS: Regional Travel-Household Interview Survey."
Steven E. Polzin Xuehao Chu. 2005. Public Transit in America : Results from the 2001 National Household Travel Survey National center for transit research
Prato, Carlo Giacomo, 2009. Route choice modeling: past, present and future research directions. J. Choice Modell. 2 (1), 65–100.
Pucher, John, Komanoff, Charles, Schimek, Paul, 1999. Bicycling renaissance in North America? Recent trends and alternative policies to promote bicycling. Transp. Res. Part A 212, 1999.
Pucher, John, Renne, John L, 2003. Socioeconomics of urban travel: evidence from the 2001 NHTS. Transp. Q. 57 (3), 49–78.
Reilly, M.K., 2002. The influence of urban form and land use on mode choice: evidence from the 1996 bay area travel survey. Annu. Meet. Transp. Res. Board.
Sall, E.A., Bhat, C.R., 2007. An analysis of weekend work activity patterns in the San Francisco bay area. Transportation 34 (2), 161–175.
Sallis, J.F., Frank, L.D., Saelens, B.E., Kraft, M.K., 2004. Active transportation and physical activity: opportunities for collaboration on transportation and public health research. Transp. Res. Part A: Policy Pract. 28 (4), 249–268.
Sener, Ipek N., Eluru, Naveen, Bhat, Chandra R, 2009. An analysis of bicycle route choice preferences in Texas, US. Transpotation 36 (5), 511–539.
Sobhani, A., Eluru, N., Pinjari, A.R., 2013. Evolution of Adults' Weekday Time Use Patterns from 1992 to 2010: A Canadian Perspective. Transportation Research Board.
Statistics Canada, 2015. Population of Census Metropolitan Areas. Statistics Canada. http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/demo05a-eng.htm (July 1, 2016).
Stinson, M.A., Bhat C.R. 2003. "An analysis of commuter bicyclist route choice using a stated preference survey." In Transportation Research Board Conference.
Stinson, Monique, Bhat, Chandra R., 2003b. Commuter bicyclist route choice: analysis using a stated preference survey. Transp. Res. Rec. 1828, 107–115.
Swait, Joffre, Ben-Akiva, Moshe, 1987. Empirical test of a constrained choice discrete model: mode choice in São Paulo, Brazil. Transp. Res. Part B 21 (2), 103–115.
Tilahun, N., Levinson, D., Krizek, K., 2007. Trails, lanes, or traffic: the value of different bicycle facilities using an adaptive stated preference survey. Transp. Res. Part A 41 (4), 287–301.
Turcotte, M., 2006. Time Spent with Family during a Typical Workday, 1986 to 2005. Statistics Canada.
Turcotte, Martin, 2005. Dependence on cars in urban neighbourhoods. Can. Social Trends Stat. Canada 11 (8). http://www.statcan.gc.ca/pub/11-008-x/2008001/article/10503-eng.pdf.
U.S. Department of Transportation, 2001. Summary of Travel Trends 2001 National Household Travel Survey (December 2004). Federal Highway Administration, pp. 1–135.
Wilkinson W., Clarke A., Epperson B., Knoblauch R. 1994. National Technical Information Service, Great Falls, VA The Effects of Bicycle Accommodations on Bicycle/Motor Vehicle Safety and Traffic Operations. Washington DC: Springfield.
Zhou, J., Golledge, R., 2006. A three-step general map matching method in the GIS environment: travel/transportation study perspective. Int. J. Geog. Inf. System 8 (3), 243–260.