Taste Variation in Environmental Features of Bicycle Routes

Thomas Koch
koch@cwi.nl
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands

Elenna R. Dugundji
dugundji@cwi.nl
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands

ABSTRACT

In this paper we look at route choice modeling based on observational GPS traces collected by bicyclists in Amsterdam and surroundings. We consider factors influencing bicycle route choice such as distance and environmental factors such as cycle-way infrastructure, land-use environment, tree cover and the effect of noise emitting roads using data from a noise emission model. We estimate a route choice model, comparing multinomial logit, mixed logit and mixed path size logit specifications. Our results show that cyclists have a highly stochastic behavior that are likely to prefer detours to drive over cycle-way infrastructure, near greener landuse and near water, and on less busy roads. Models such as mixed logit that can estimate the stochasticity of cyclists perform best to capture this behavior.

CCS CONCEPTS

• **Applied computing** → **Transportation**.

KEYWORDS

bicycle route choice, GPS trajectories, environmental variables, geospatial feature engineering, taste variation, mixed logit

ACM Reference Format:
Thomas Koch and Elenna R. Dugundji. 2021. Taste Variation in Environmental Features of Bicycle Routes. In *14th ACM SIGSPATIALInternational Workshop on Computational Transportation Science (IWCTS'21), November 1, 2021, Beijing, China.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3486629.3490697

1 INTRODUCTION

Since the 1970's discrete choice modeling has been a leading method to understand choice behavior of individuals in a wide range of fields such as marketing, economics and transportation. Described by McFadden et al. [20] in 1973, discrete choice modeling has subsequently been extended over the decades in order to overcome specific limitations such as overlapping alternatives and correlations over time.

The study of the specific field of route choice is more complicated than a choice between easily enumerable distinct alternatives, as route choice is typically a sequence of choices at each intersection, each transit stop, each mode, etc. This leads to very large choice set that is theoretically infinite due to loops. Often there can also be a large overlap between different route alternatives leading to difficulties for choice modeling.

An approach established in the 1990s to model route choice using a collection of observed paths and for each observation a set of paths generated by a route choice generator. This approach has been used to estimate models such as multinomial logit (MNL) and mixed logit. This approach comes with limitations: as discussed in Koch et al. [18], these route choice generators do not necessarily create realistic routes; and Frejinger et al. [12] argue that parameter estimates can vary significantly depending on the bias of the route choice generator.

To address the issues with the overlap between different alternative paths and the resulting correlations, multiple extensions have been proposed to attempt to avoid erroneous path probabilities and substitution patterns. The most two popular are path size logit proposed by Ben-Akiva and Bierlaire [1] and C-Logit proposed by Cascetta et al. [7], which decrease the utility of overlapping paths proportional to the overlap with other paths included in the choice set.

A second approach is to achieve a consistent choice set by sampling, as proposed by Frejinger et al. [12]. This approach attempts to set up a sampling protocol in order to obtain unbiased parameter estimates from the route choice sets to neutralize the bias introduced by the route choice generator.

A third approach uses a link-based Markov decision process to model route choice as a series of sequential decisions. First proposed by Fosgerau et al. [11] it uses a linear system of equations to efficiently compute choice probabilities by using a solver to solve Bellman equations. Zimmermann et al. [28] showed that it is possible to estimate bicycle route choice without the restrictiveness of pre-generated route choice sets in this way.

In our previous work [16] we similarly attempted to estimate a recursive logit model, using the same observational data in this study. We found that the bicycling network in Amsterdam however was particularly complex due to the high number of alternatives. This led to numerous numerical issues making it not possible to correctly estimate the model. Another limitation of recursive logit is that it does not allow to have a single link of the network with a negative cost, which can happen in plausible configuration of most models, for example a model that prefers detours in a green park.

In a follow-up study [17] we compared our previous recursive model with a simple multinomial logit model containing a full set of environmental features, with promising results. In the present research, we focus on extending this model to consider taste variation of cyclists for the environmental features of the routes.

2 BACKGROUND

In 2010, Menghini et al. [21] published a seminal route choice model for bicyclists estimated from a large sample of GPS observations in a

IWCTS'21, November 1, 2021, Beijing, China

ACM ISBN 978-1-4503-9117-7/21/11.
https://doi.org/10.1145/3486629.3490697

revealed preference study with 2435 persons logging 73,493 trips in Zurich, Switzerland. Using this data they estimated a multinomial logit model and used breadth-first search link elimination (BFS-LE) to generate choice alternatives for each observed trip. They included six different variables in the choice model: length of the route, average absolute gradient change, maximum gradient change, percentage of marked bicycle paths along the route, number of traffic lights and the path size measure. Accounting for the similarity between alternatives with the path size vector, their model showed that the elasticity with respect to trip length was nearly four times larger than that with respect to the percentage of bicycle paths along the route. The only other explanatory variable that had an impact albeit small, was the product of length and the maximum gradient along the route.

Sener et al. [25] in 2009 estimated a choice model using a stated preference survey on Texas bicyclists and analyzed a comprehensive set of attributes that influence bicycle route choice, such as the bicyclists characteristics, on-street parking, bicycle facility types and amenities, roadway physical characteristics, roadway functional characteristics, and roadway operational characteristics. To estimate the model they used a panel mixed multinomial logit model. Their results indicate that travel time (for commuters) and motorized traffic volume are the most important attributes in bicycle route choice. Other attributes with a high impact include number of stop signs, red light, and cross-streets, speed limits, on-street parking characteristics, and whether there exists a continuous bicycle facility on the route.

Hood et al. [15] estimated a route choice model on GPS data collected by cyclists with a smartphone in San Francisco. Their choice set was generated by the double stochastic method of Bovy and Fiorenzo-Catalano [5]. They avoid the issues with the overlap between different alternative paths by opting for a Path Size Multinomial Logit model and there results showed that bicycle lanes were preferred to other facility types, that steep slopes were disfavored. Other negative attributes were length and turns. Traffic volume, traffic speed, number of lanes, crime rates, and nightfall had no effect.

Broach et al. [6] looked at are revealed preference dataset collected by 164 cyclists using GPS units. Broach et al. [6] used 1449 utilitarian trips to estimate a bicycle route choice model. To generate the choice-set they used an algorithm based on multiple permutations of path attributes and formulated to account for overlapping route alternatives. Their results from the Path Size Logit model suggest that bicyclists are sensitive to the effects of distance, turn frequency, slope, number of traffic signals, and traffic volumes. Additionally, bicyclists appear to highly value off-street bike paths, enhanced neighborhood cycleways with traffic calming features, and bridge facilities. Bike lanes more or less exactly offset the negative effects of adjacent traffic.

In 2017, Ton et al. [26] reported on a route choice model for bicyclists using the same dataset as in the study in this paper. Ton et al. consider the construction of choice sets via an empirical approach, using only the observed trips in the data set to compose the choice alternatives. On the basis of their specific focus (inner city travel in Amsterdam) Ton et al. selected 6 variables: distance, percentage of separate cycle paths, number of intersections, rain, sunset and sunrise times and trip purpose. Their findings suggest that bicyclists in Amsterdam are insensitive to dedicated cycle paths, attributed to an inner city characterized by a dense road network where cycling is the most prominent mode of transport. Additionally they found that cyclists in Amsterdam were found to minimize travel distance and the number of intersections per kilometer. Furthermore they found that for early morning trips there was a stronger impact of distance on route choice than outside these hours. In a subsequent paper Ton et al. [27] looked at a data-driven path identification approach, combining all unique routes observed for one origin-destination pair into a choice set and comparing this approach with two commonly used choice set generation methods (breadth-first search on link elimination and labelling).

Ghanayim and Bekhor [13] analyzed bicycle route choice for commuter trips using a dataset from a GPS-assisted household travel survey in the Tel Aviv metropolitan area. Their results indicate an expected tendency to ride in longer routes, but on separated bike lanes. In the absence of such lanes, riders prefer to use local streets and avoid busy arterial streets and highways. Their route choice generation calculated 20 alternatives using 3 methods: link elimination, link penalties and a simulation method that calculated a shortest path using link impedance after each draw from a log normal distribution. The paper estimated a choice model using 3 model forms: multinomial logit, C-logit and Path Size-logit.

Bernardi et al. [2] analyzed the GPS traces recorded by 280 cyclists in a mobility panel throughout the Netherlands, that made approximately 3500 bike trips over a four week period in 2014. The choice sets were composed by the shortest-path for the origin destination and 4 alternative paths that were observed for that origin-destination pair. Route choice models were estimated using a binomial logit model and a mix multinomial logit model with path size logit formulation. Their results show that trip lengths and trip distribution over time reveal a population sample much used to cycling, frequently and over long distances.

Zimmermann et al. [28] showed it is possible to estimate route choice without the restrictiveness of pre-generated route choice sets and model route choice as a sequence of choices. Zimmermann et al. [28] look at bicycle route choice problem in the city of Eugene, Oregon. By using 648 observations of bike trips collected from 103 users. They test a long list of 14 potential parameters: length; link constant to penalize paths with many constants; length interacted separately with upslope, medium traffic, heavy traffic, regional multi-use path, bicycle boulevard, bike lane; bridge; bridge interacted with bike facilities; no turn; no turn interacted with crossroad; left turn interacted with crossroad separately for medium traffic and for heavy traffic.

Chen et al. [8] examines the effects of built environment features, including factors of land use and road network, on bicyclists route preferences using GPS datasets collected in the city of Seattle. The choice model was estimated using a path-size-based mixed logit model. Chen et al. [8] identifies five core factors that influence route choice behavior: trip length, speed limit, slope, bicycle lanes, and street lights.

The study by Prato et al. [23] focused on observing bicyclist behavior in the cycling oriented country of Denmark, exploiting rich data about the cycling environment, estimating the model in value of distance rather than preference space. Prato et al. [23] not only focused on preferences for traditional variables such as

distance, hilliness and road characteristics. but also on aspects such as bicycle facilities and land-use designations. They estimated a model on 3384 cycling trips using mixed path size logit.

Dane et al. [9] looked at the route choice behavior of cyclists on e-bikes based on 17626 trips from 742 users extracted from GPS data. In this paper a mixed logit model with addition of the path-size attribute is applied on the route choice of respondents. Choice sets were generated using the K-shortest path algorithm.

In our prior research we found in Koch et al. [18] that bicyclists in Amsterdam often deviate from the shortest path, more than car drivers, indicating that there are different and possibly also more factors at play in the route choice of bicyclists in Amsterdam. In Koch et al. [19] we focused on the concept of route complexity: counting the number of locations where people deviate from the shortest path, in the interest of improving route choice generation techniques and potentially get more insight into the motivations for the route choice for bicyclists.

In another previous study [16] using the same observational data, we explore other effects on route choice using different methodologies, without looking at route complexity or where people deviate from the shortest path. To this end, we first attempted to estimate a recursive logit model. However due to the complexity of cycleway road infrastructure in Amsterdam, the number of possible options is higher than we would see in car route choice or in a city without two cycle-paths on both sides of major roads or two roads in both directions (for cyclists) along the canals. The high number of alternatives led to numerous numerical issues making it not possible to correctly estimate the model. Another limitation of recursive logit is that it does not allow to have a single link of the network with a negative cost, which can happen in plausible configuration of most models, for example a model that prefers detours in a green park.

In Koch and Dugundji [17] we extended this work and compared the recursive logit model with a simple multinomial logit model using a synthetically generated choice set. The synthetic approach allows us to consider additional plausible route alternatives outside the set of observed routes. This means that we can include all observations even between origin destinations pairs with only a single observation, unlike the study by Ton et al. [26] that is limited to trips traversing the inner city of Amsterdam, due to a insufficient density of trips in the suburbs for this empirical approach to work there. In the present research, we focus on extending our previous model to consider taste variation of cyclists for environmental features of the routes. For this purpose we will estimate a series of mixed logit models.

3 METHODS

In this section we briefly explain the three discrete choice modeling techniques used in this paper.

3.1 Multinomial Logit

In the multinomial logit the utility U for each observation n for each alternative i is

$$U_{ni} = \beta x_{ni} + \varepsilon_{ni} \tag{1}$$

$$\varepsilon_{ni} \sim \text{iid extreme value} \tag{2}$$

The probability that for observation n alternative i is chosen is:

$$P_{ni} = \frac{e^{\beta_n X_{ni}}}{\sum_j e^{\beta_n X_{nj}}} \tag{3}$$

The model is then estimated by maximizing the log-likelihood. First we estimated multinomial logit (MNL) models with a single variable per model and based on those results we removed some variables with high correlation between them. Firstly we decided to include only the highest and lowest level of traffic noise exposure. While the four variables were significant on their own, there was not much difference for the estimated coefficient values between 60 and 70 decibels. Secondly we only included the absolute number of traffic signals as the frequency of traffic signals per kilometer had a lower t-test score. The choice model was estimated both using PandasBiogeme[3].

3.2 Taste Variation: Mixed Logit

A limitation of the multinomial logit model is that it does not take random taste variation among the thousands of cyclists in our dataset into account. This is why in this study we opted for mixed logit, which is an extension of multinomial logit. In the mixed logit the utility is generalized by allowing B_n to be random, this makes the utility of observation n for each alternative i:

$$U_{ni} = \beta_n x_{ni} + \varepsilon_{ni} \tag{4}$$

$$\beta_n \sim f(\beta|\theta) \tag{5}$$

The probability conditional on β_n that for observation n alternative i is chosen is:

$$L_{ni}(\beta_n) = \frac{e^{\beta_n X_{ni}}}{\sum_j e^{\beta_n X_{nj}}} \tag{6}$$

As β_n is random, the choice probability is the integral of the logit formula in equation 6 over the probability density function f.

$$P_{ni} = \int L_{ni}(\beta) f(\beta|\theta) d\beta \tag{7}$$

3.3 Path size logit

With logit the utility of overlapping paths is overestimated. When δ is large, there is some sort of double counting. The idea of path size logit is to correct for that:

$$V_p = -\beta x_p + \beta \ln PS_p \tag{8}$$

where

$$PS_p = \sum_{(i,j)\in p} \frac{c(i,j)}{c_p} \frac{1}{\sum_{q\in C} \delta^q_{i,j}} \tag{9}$$

and

$$\delta^q_{i,j} = \begin{cases} 1, & \text{if link (i,j) belongs to path q} \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

Where c is the cost function, in study we will simply use length as the cost function.

4 CASE STUDY

In this section we describe the GPS data collected by a panel of volunteers and how we generated environmental variables and how we generated the set of alternatives for the choice modeling.

4.1 GPS Route Trajectory Data

For this study we used the 2016 FietsTelweek ("Bicycle Counting Week") data set (Bikeprint [4]) that is available at their website. It contains 282,796 unique trips (although the corresponding infographic http://fietstelweek.nl/data/resultaten-fiets-telweek-bekend/ mentions 416,376 trips having a total distance of 1,786,147 kilometer). During the week of the 19th of September 2016 approximately 29,600 bicyclists volunteered to track their bicycle movements using a smartphone app. For this case study we limited the study to bicycle trips to and/or from the city of Amsterdam, Diemen, Amstelveen and Ouder-Amstel, leaving around 29,684 trips.

This app ran in the background collecting all bicycle movements by the participants using the phone's GPS and acceleration sensors. The participants in the study travel by bicycle for daily activities in a way as often seen in the Netherlands, as transportation from and to work, supermarket, school, etc. For privacy reasons the resulting data was anonymized by the data provider before making it publicly available: (i) by the removal of user information to make it impossible to trace multiple trips to a single person; (ii) by rounding of the trip departure time into one-hour bins to the nearest hour; and (iii) removal of the random number between 0 and 400 meters from the start and the end of the trip to obfuscate the true origin and destination of each trip.

4.2 Choice Set Generation

To find out what kind of alternatives exist for each observed path we applied synthetic route choice generation using the Double Stochastic Generation Function (DSGF) method described by Nielsen [22]. The DSGF approach produces heterogeneous routes because both the cost and parameters used in the cost function for the links are drawn from a probability function. This way it can generate random paths, just by calculating the shortest path since the cost of each route is based on random factors. Halldórsdóttir et al. [14] showed that DSGF has a high coverage level of replicating routes taken by bicyclists and that it performs well up to 10 kilometer. Furthermore Bovy and Fiorenzo-Catalano [5] state that the method guarantees, with high probability, that attractive routes are included in the choice set, while unattractive routes are left out.

We used an existing implementation of DSGF, specifically POSDAP by ETH-Zurich [10] working on a street network provided by the data collection team of the Fietstelweek, that they imported from OpenStreetMap. We slightly modified POSDAP to execute at most a given number of $M = 128$ iterations (instead of running for a given duration) so that it behaves identically on different machines. For some origin destination pairs POSDAP was not able to find as many as N_0 routes in M iterations, in which case we will use all found routes. The choice sets are written to CSV files for further processing.

Additionally we also run an implementation of Breadth First Search Link Elimination (BFS-LE) by Rieser-Schüssler et al. [24], but we opted to leave out these alternatives since there was not much variance in the route choice set. In Koch et al. [18] we published on the coverage and consistency of both synthetic choice sets.

Table 1: Descriptive statistics on the variables of the observed bicycle trips.

	Min	Median	Avg	Std-dev	Max
Length	500.03	2640.41	4032.32	4935.01	149427.61
55+dB noise	0	0.73	0.68	0.27	1
70+dB noise	0	0.39	0.42	0.29	1
Near green	0	0.18	0.22	0.21	1
Residential	0	0.53	0.52	0.26	1
Near retail	0	0.18	0.22	0.20	1
Near tram	0	0.10	0.24	0.29	1
Tree cover	0	0.40	0.39	0.17	1
Near water	0	0.30	0.34	0.22	1
Traffic signals	0	1	1.99	2.64	25
Cycleway	0	0.53	0.52	0.27	1
$\ln PS$	-2.73	-1.30	-1.28	0.65	0

Table 2: Descriptive statistics on the variables of the generated alternative bicycle trips.

	Min	Median	Avg	Std-dev	Max
Length	175.59	2663.10	3956.69	4487.86	150758.47
55+dB noise	0	0.74	0.69	0.26	1
70+dB noise	0	0.43	0.43	0.26	1
Near green	0	0.15	0.20	0.19	1
Residential	0	0.54	0.52	0.24	1
Near retail	0	0.18	0.23	0.19	1
Near tram	0	0.15	0.24	0.26	1
Tree cover	0	0.39	0.39	0.16	1
Near water	0	0.27	0.31	0.22	1
Traffic signals	0	1	2.19	2.72	28
Cycleway	0	0.48	0.47	0.24	1
$\ln PS$	-2.78	-1.45	-1.44	0.49	0

4.3 Geospatial Feature Engineering

To collect a set of variables that would reasonably impact route choice of bicyclists we collected and processed open data sources to compute various explanatory variables describing each route. The procedure for the generation of the variables is described below. Descriptive statistics of the variables of the observed trajectories are given in Table 1 and for the complete set of generated alternatives in Table 2.

First of all for each link in the network we include the length of that link as ***distance*** and if that link is a dedicated cycleway, we include the length as ***oncycleway***. Additionally we have a variable ***traveltime*** based on the length and an estimated speed based on the GPS observations.

To include data about the environment of each link we extracted information of data made openly available by the city of Amsterdam. Firstly we pulled potentially relevant variables from a geographical data-set with land-use zones. To combine the street-network with other relevant geographical data-sets, we cut each street link into small segments of 5 meters and determined the distance of each segment to a geographical feature in the land use data-set.

The variable ***nearwater*** is the sum of the length of each segment that is situated close to water bodies such as the canals of Amsterdam, (small) lakes, rivers and other water bodies wider than 6 meters. To determine a preference for routes through parks and forests we did the same with the variable ***neargreen***, measures the length of streets that are situated within a 25 meter radius of 'green' land used for parks, forests and meadows.

For a more fine-grained indication of the level of green and trees along a route, we used a data-set of the location of each individual tree in Amsterdam to determine what portion of each street segment is covered by trees. Our reasoning is that the number of trees has an influence on route choice as they can provide shade on hot days and function as a cover against the wind in storm conditions. To determine the variable ***neartree*** we measured the distance of the street within 30 meters left or right from one or more tree(s). This way a street along a row of trees would have the full distance. We determined the distance of 30 meters between road and tree based on various situations where trees are situated along bicycle roads in Amsterdam.

To measure the effect of residential areas the variable ***nearresidential*** measures the distance of streets in residential areas. . The variable ***nearretail*** describes distance within land used for *'Shops, malls and hotels-restaurants-pubs'*, *'Public offices and services'* and *'Cultural, social, medical, educational'*.

To see if the vicinity of busy roads, a major source of noise and pollution, has any impact on route choice we used a data set with the noise contours map of road traffic in Amsterdam as shown in Figure 1. This data-set is produced by a model that estimates the level of exposure to traffic noise. Tn this map there are four noise levels with respectively at least 55, 60, 65 or 70 decibels of noise. The variables ***nearXdb*** represent the distance of the street passing through these exposure zones.

Figure 1: Example noise contour map in Amsterdam, used for computing the variable that indicates the distance of a route trajectory along roads with noisy traffic

Based on the idea that tramlines in Amsterdam form a radial artery towards the heart of the city, we construct the variable ***neartram*** indicating the portion of the route that is situated 100 meter from tram rails either to the left or right of the path, measured using segments of 10 meters.

Finally we wanted to see if the number and frequency of traffic signals has a measurable effect on route choice. To create this variable we used a dataset with traffic signals in Amsterdam and counted the number of signals that were 10 meters to the left or the right of the path. We included this in two ways: first the exact number of traffic signals with ***ntrafficsignal*** and secondly the frequency of traffic signals ***trafficsignalfreq*** where the number of signals is divided by the length of the route.

Since Amsterdam has no elevation changes beyond the occasional bridge, we did not include any elevation changes as a variable.

5 RESULTS

5.0.1 Multinomial Logit. The estimated parameters for our baseline multinomial logit model are presented in the first column of Table 3. The results are mostly as expected except for length. The number of traffic signals on the route and being near roads with (heavy) noise emission all have a negative utility. Being close to water and/or green land-use and dedicated cycle-way infrastructure all have a strong positive utility. Being near retail land-use has a positive utility while residential land-use has a negative utility. The positive utility of being near tram lines may be a correlation effect of tramlines in Amsterdam forming a guide way into the city center. The unexpected positive effect of additional distance warrants further investigation by adding mixing for taste variation.

5.0.2 Mixed Logit. The results for the series of mixed logit models are listed in the second through twelfth columns of Table 3. These results are based on running the model with 5000 draws. In each column we list the estimated parameters for a mixed logit model with the standard deviation (sigma σ) on the variable indicated in the header of the column. The estimated standard deviation (sigma) for that variable is listed at the bottom of the table, this is not applicable for the baseline multinomial logit model. The log likelihood for each model is listed in Table 4.

5.0.3 Path size logit. In Table 5 we listed results for models we ran with the path size logit as a variable. The results for a multinomial logit and mixed logit models were estimated using 1000 draws. The log likelihoods are listed in Table 6. We also see that the natural log of path size variable is statistically significant in every model that we estimated. Again just like as in the mixed logit model we see here that the model with taste variation on route length has the best log likelihood.

6 DISCUSSION

Based on the log likelihood ratio-tests conducted and included in Table 7 we see that each extension brings an improvement of the log-likelihood at a very high level of statistical significance. The series of mixed logit models with path size correction are the best performing model.

We also see that adding the taste variation on the length attribute results in the largest improvement, this intuitively makes sense as a large number of anonymous cyclists observed all have their own taste for how much distance they are prepared to cycle. This leads to different tolerances to base behavior on attributes such as land-use and cycle-way infrastructure.

A limitation in this study is that the cyclists data used in this study was anonymous, so we were not able to group multiple observations with the same taste preferences of an individual using a panel mixed logit model. A panel mixed logit model will likely yield a further improvement of the model.

While we see statistical significant improvement of path-size, we also see that it brings a smaller improvement than taste variation. This is likely due to a low overlap in the alternatives generated by the route choice generator that we used, namely POSDAP[10] with doubly stochastic choice set generation. This is also shown in the distributions of path sizes in observed and generated paths as shown in Figure 2 and 3. The low overlap is what we have already found in our earlier study on the quality of these choice sets in our previous work in Koch et al. [18].

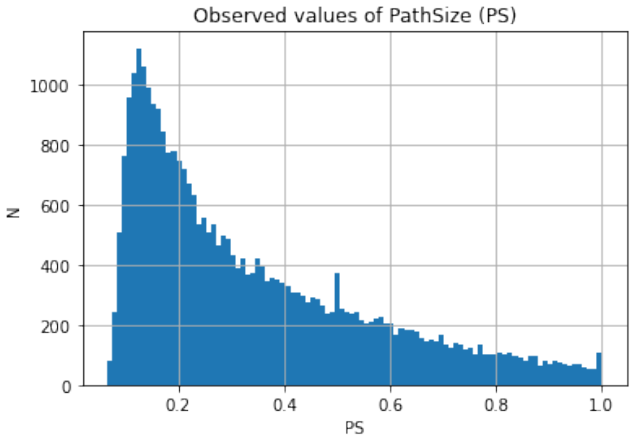

Figure 2: Observed values of PathSize (PS) in the observed trajectories

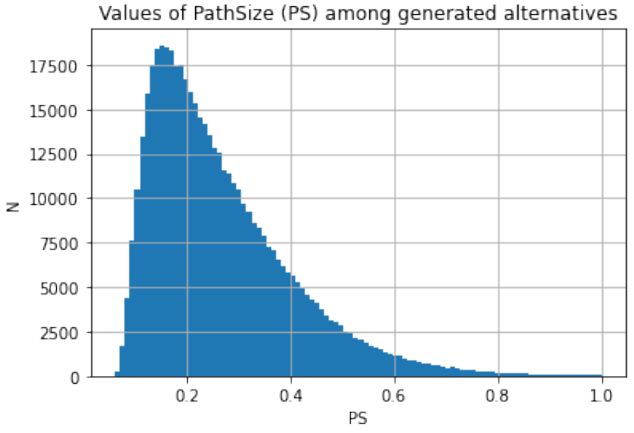

Figure 3: Histogram of PathSize (PS) in the collection of generated alternatives

Our cycling models show that cyclists route choice comes with a high level of stochasticity: how much distance bicyclists are prepared to take detours can widely vary person by person. This can make it hard to simulate route choice by bicyclists, harder than the route choice for car drivers, as shown in our previous study Koch et al. [18] on route complexity for bicyclists and car drivers.

7 CONCLUSIONS

In this paper we have estimated a route choice model on 29,684 bicyclist trips conducted to/from the City of Amsterdam, Diemen, Amstelveen and Ouder-Amstel. We have generated alternative routes using the doubly stochastic generation function in POSDAP [10]. Using this choiceset we have estimated 3 kinds of discrete choice models, a simple multinomial logit model, mixed logit and both extended with the path size logit variable. Our results show that the mixed logit with path size correction and taste variation on length results in the best performing model.

This model shows that bicyclists are prepared to make detours to use dedicated cycle-way infrastructure, avoid roads near heavy noise (70 decibels or more) emitting roads and avoid traffic signals. Cyclists prefer roads along tram lines, probably a correlation with how cyclists find their way into the city center. Cyclists prefer to ride along water, under the cover of trees and along green land-use zones such as parks. An interesting finding is that cyclists prefer to cycling in retail land-use zones while avoiding residential land use. In a future study we would like to repeat our work on a dataset where we can identify individuals in order to estimate a panel mixed path-size logit model.

ACKNOWLEDGMENTS

This research has been conducted in the framework of the Impact Study North/Southline, funded in part by the Municipality of Amsterdam and the regional transportation authority of Amsterdam. Hereby we also thank CWI/VU guest researcher Luk Knapen for useful discussions related to the research topic and assistance with the data preparation.

Table 3: Baseline multinomial logit and series of mixed logit models with taste variation per explanatory variable.

	MNL	**Route Length**	**55+dB Noise**	**70+dB Noise**	**Near green**	**Resid-ential**	**Near retail**	**Near tram**	**Tree cover**	**Near water**	**Traffic signals**	**Cycle way**
βroute length	1.86	-2,21	1.67	1,66	1,76	1,74	1,77	1,61	1,76	1,77	1,8	1.9
t-test	62.1	-14,4	51.6	50,7	55,2	54.5	57,2	50,8	54,3	53,9	57,2	56.2
β55+dB noise	-0.324	-0,496	0.484	-0,711	-0,284	-0,0492	-0,32	-0,403	-0,046	-0,519	-0,348	-0.356
t-test	-3.32	-3,91	3.36	-6,06	-2,7	-0,45	-3,05	-3,66	-0,416	-4,57	-3,47	-3.15
β70+dB noise	-2.09	-2,54	-2.68	-2,25	-2,28	-2,37	-2,13	-2,42	-2,39	-2,08	-2,2	-2.52
t-test	-26.3	-24,9	-29.2	-18,1	-26,4	-26,8	-25,1	-26,3	-26,7	-22,6	-26,8	-26.9
βnear green	2.47	2,84	2.73	3,1	3,73	2,79	2,59	2,69	2,68	2,59	2,43	2.85
t-test	28.9	25,3	27.5	30,1	29	28,7	28,7	27,9	27,4	26,2	27,8	29.1
βresidential	-0.373	-0,415	-0.094	-0,263	-0,261	-0,259	-0,43	-0,0838	-0,224	-0,511	-0,376	-0.324
t-test	-5.51	-4,83	-1.2	-3,3	-3,54	-2,62	-5,86	-1,1	-2,96	-6,59	-5,39	-4.18
βnear retail	0.639	0,921	0.967	0,994	0,592	1,14	0,612	0,56	0,818	0,776	0,589	0.933
t-test	6.9	7,95	9.24	9,26	6,02	10,8	4,79	5,39	8	7,35	6,15	8.96
βnear tram	1.3	1,89	1.34	1,37	1,59	1,5	1,41	0,655	1,59	1,45	1,38	1.67
t-test	19.2	21,7	17	16,6	21,3	19,9	19,4	5,43	20,9	18,5	19,5	21.2
βtree cover	1.46	2,44	2.17	1,56	1,65	1,71	1,66	1,77	3	2,27	1,71	1.81
t-test	11.5	14,7	14.3	10	11,7	11,9	12,3	12,1	15,5	15,4	13	12.5
βnear water	2.11	2,22	2.26	1,93	2,16	2,31	2,23	2,15	2,29	2,11	2,09	2.09
t-test	36	29,5	33.6	27,7	33,7	35,1	35,5	33	34,8	21,2	34,6	30.8
βtraffic signals	-0.496	-0,501	-0.591	-0,733	-0,465	-0,499	-0,513	-0,649	-0,501	-0,479	-1,21	-0.476
t-test	-9.84	-7,61	-10.6	-12,5	-8,86	-9,28	-9,75	-11	-9,33	-8,71	-15,4	-8.64
βcycleway	4.3	4,63	4.91	4,87	4,61	4,67	4,55	4,89	4,71	4,86	4,35	6.43
t-test	78	70	78.5	76,4	77,8	77	77,7	79	77,5	78,3	76,9	64.8
σ	N/A	16,9	10.6	10,3	8,93	8,22	8,94	10,1	15,8	9,3	4,89	6.53
t-test	N/A	65,9	50.2	57,4	38,8	44,6	34,7	50,4	46,6	52,7	35	64.7

Table 4: Log likelihood of baseline multinomial logit and series of mixed logit models with taste variation per explanatory variable.

	MNL	**Route Length**	**55+dB Noise**	**70+dB Noise**	**Near green**	**Resid-ential**	**Near retail**	**Near tram**	**Tree cover**	**Near water**	**Traffic signals**	**Cycle way**
LL	-79817	-75922	-78500	-77851	-79215	-79028	-79407	-78092	-78963	-78478	-79365	-77705

Table 5: Baseline multinomial logit and series of mixed logit models with path size and taste variation per explanatory variable.

	MNL	**Route Length**	**55+dB Noise**	**70+dB Noise**	**Near green**	**Resid-ential**	**Near retail**	**Near tram**	**Tree cover**	**Near water**	**Traffic signals**	**Cycle way**
βroute length	1.27	-5.03	1.33	1.37	1.29	1.29	1.29	1.29	1.3	1.37	1.28	1.44
t-test	38.4	17.8	36.5	36.7	36.7	36.6	37.7	36.3	36.4	37	36.8	37.7
β55+dB noise	-0.32	-0.353	0.41	-0.681	-0.291	-0.081	-0.312	-0.394	-0.075	-0.504	-0.337	0.334
t-test	-3.39	-2.85	2.98	-5.93	-2.87	-0.766	-3.1	-3.66	-0.702	-4.56	-3.48	-3.03
β70+dB noise	-1.93	-2.3	-2.54	-2.17	-2.13	-2.23	-2	-2.3	-2.25	-1.97	-2.04	-2.37
t-test	-25	-23.1	-28.4	-18.1	-25.5	-25.9	-24.4	-25.5	-25.9	-21.9	-25.6	-25.9
βnear green	2.35	2.76	2.66	3.03	3.47	2.69	2.47	2.61	2.58	2.51	2.33	2.76
t-test	28.5	25.2	27.5	29.9	28.7	28.7	28.3	27.7	27.4	26.1	27.5	28.9
βresidential	-0.347	-0.395	-0.103	-0.264	-0.263	-0.262	-0.408	-0.086	-0.225	-0.493	-0.351	-0.317
t-test	-5.3	-4.7	-1.35	-3.38	-3.7	-2.81	-5.78	-1.15	-3.06	-6.54	-5.18	-4.18
βnear retail	0.665	0.953	-0.964	0.996	0.632	1.11	0.631	0.579	0.826	0.775	0.606	0.931
t-test	7.41	8.42	9.42	9.46	6.64	10.8	5.34	5.69	8.32	7.53	6.53	9.15
βnear tram	1.31	1.97	1.36	1.38	1.55	1.49	1.4	0.716	1.56	1.45	1.37	1.67
t-test	20.1	23.2	17.8	17.1	21.6	20.5	20.1	6.22	21.2	19	20.1	21.6
βtree cover	1.42	2.57	2.1	1.53	1.59	1.64	1.59	1.73	2.78	2.19	1.63	1.77
t-test	11.6	23.2	17.8	10	11.7	11.8	12.1	12.1	15.1	15.2	12.8	12.4
βnear water	2.03	2.13	2.2	1.9	2.09	2.23	2.16	2.11	2.21	2.05	2.02	2.04
t-test	35.6	28.8	33.6	27.8	33.7	35	35.4	33.1	34.6	21.7	34.5	30.8
βtraffic signals	-0.477	-0.439	-0.575	-0.712	-0.449	-0.485	-0.494	-0.635	-0.486	-0.468	-1.13	-0.466
t-test	-9.68	-6.72	-10.5	-12.3	-8.76	-9.22	-9.62	-10.9	-9.25	-8.67	-15.1	-8.6
βcycleway	4.26	4.6	4.86	4.85	4.55	4.63	4.48	4.86	4.66	4.82	4.31	6.32
t-test	78.6	70.2	78.8	76.8	78.2	77.6	78.1	79.3	78	78.6	77.6	65.2
βln PS	0.56	1.06	0.365	0.313	0.473	0.458	0.484	0.339	0.462	0.396	0.496	0.439
t-test	30.5	37.9	17.2	14.3	23.9	22.6	24.6	16.1	22.8	18.8	26	20.9
σ	N/A	17.8	9.71	9.7	7.77	7.32	7.45	9.35	14.3	8.58	4.31	9.08
t-test	N/A	67.3	45.7	58.3	34	39.7	28.5	47.3	42.1	48.7	30.8	62.1

Table 6: Log Likelihood multinomial logit and series of mixed logit models with path size and taste variation per explanatory variable.

	MNL	**Route Length**	**55+dB Noise**	**70+dB Noise**	**Near green**	**Resid-ential**	**Near retail**	**Near tram**	**Tree cover**	**Near water**	**Traffic signals**	**Cycle way**
LL	-79372	-75221	-78356	-77758	-78951	-78793	-79123	-78030	-78805	-78307	-79044	-77501

Table 7: Overview of log likelihood ratio tests

Reference Group(s)	d.f.	LR	LU	-2[LR- LU]	χ^2 d.f.(0.05)	p-value	Comments
LR Test: Mixed logit (unrestricted) vs. MNL (restricted)							
Length	1	-79817	-75922	7790	3,84	0	Reject restrictions
55+ dB noise	1	-79817	-78500	2634	3,84	0	Reject restrictions
70+ dB noise	1	-79817	-77851	3932	3,84	0	Reject restrictions
Near green	1	-79817	-79215	1204	3,84	0	Reject restrictions
Residential	1	-79817	-79028	1578	3,84	0	Reject restrictions
Near retail	1	-79817	-79407	820	3,84	0	Reject restrictions
Near tram	1	-79817	-78092	3450	3,84	0	Reject restrictions
Tree cover	1	-79817	-78963	1708	3,84	0	Reject restrictions
Near water	1	-79817	-78478	2678	3,84	0	Reject restrictions
Traffic signal	1	-79817	-79365	904	3,84	0	Reject restrictions
Cycleway	1	-79817	-77705	4224	3,84	0	Reject restrictions
LR Test: Mixed logit with Pathsize (unrestricted) vs. MNL with Pathsize (restricted)							
Length	1	-79372	-75221	8302	3,84	0	Reject restrictions
55+ dB noise	1	-79372	-78356	2032	3,84	0	Reject restrictions
70+ dB noise	1	-79372	-77758	3228	3,84	0	Reject restrictions
Near green	1	-79372	-78951	842	3,84	0	Reject restrictions
Residential	1	-79372	-78793	1158	3,84	0	Reject restrictions
Near retail	1	-79372	-79123	498	3,84	0	Reject restrictions
Near tram	1	-79372	-78030	2684	3,84	0	Reject restrictions
Tree cover	1	-79372	-78805	1134	3,84	0	Reject restrictions
Near water	1	-79372	-78307	2130	3,84	0	Reject restrictions
Traffic signal	1	-79372	-79044	656	3,84	0	Reject restrictions
Cycleway	1	-79372	-77501	3742	3,84	0	Reject restrictions
LR Test: MNL with Pathsize (unrestricted) vs. MNL (restricted)							
Baseline	1	-79817	-79372	890	3,84	0	Reject restrictions
LR Test: Mixed logit with Pathsize (unrestricted) vs. Mixed logit (restricted)							
Length	1	-75922	-75221	1402	3,84	0	Reject restrictions
55+ dB noise	1	-78500	-78356	288	3,84	0	Reject restrictions
70+ dB noise	1	-77851	-77758	186	3,84	0	Reject restrictions
Near green	1	-79215	-78951	528	3,84	0	Reject restrictions
Residential	1	-79028	-78793	470	3,84	0	Reject restrictions
Near retail	1	-79407	-79123	568	3,84	0	Reject restrictions
Near tram	1	-78092	-78030	124	3,84	0	Reject restrictions
Tree cover	1	-78963	-78805	316	3,84	0	Reject restrictions
Near water	1	-78478	-78307	342	3,84	0	Reject restrictions
Traffic signal	1	-79365	-79044	642	3,84	0	Reject restrictions
Cycleway	1	-77705	-77501	408	3,84	0	Reject restrictions

REFERENCES

[1] Moshe Ben-Akiva and Michel Bierlaire. 1999. Discrete choice methods and their applications to short term travel decisions. In *Handbook of transportation science*. Springer, 5–33.

[2] Silvia Bernardi, Lissy La Paix Puello, and Karst Geurs. 2018. Modelling route choice of Dutch cyclists using smartphone data. *Journal of transport and land use* 11, 1 (2018), 883–900.

[3] Michel Bierlaire. 2018. *PandasBiogeme: a short introduction*. Technical Report. Technical Report TRANSP-OR 181219, Transport and Mobility Laboratory, EPFL.

[4] Bikeprint. 2017. Download bestanden Nationale Fietstelweek 2015 en 2016. http://www.bikeprint.nl/fietstelweek/

[5] Piet HL Bovy and Stella Fiorenzo-Catalano. 2007. Stochastic route choice set generation: behavioral and probabilistic foundations. *Transportmetrica* 3, 3 (2007), 173–189.

[6] Joseph Broach, Jennifer Dill, and John Gliebe. 2012. Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice* 46, 10 (2012), 1730–1740.

[7] Ennio Cascetta, Agostino Nuzzolo, Francesco Russo, and Antonino Vitetta. 1996. A modified logit route choice model overcoming path overlapping problems. Specification and some calibration results for interurban networks. In *Transportation and Traffic Theory. Proceedings of The 13th International Symposium On Transportation And Traffic Theory, Lyon, France, 24-26 July 1996*.

[8] Peng Chen, Qing Shen, and Suzanne Childress. 2018. A GPS data-based analysis of built environment influences on bicyclist route preferences. *International journal of sustainable transportation* 12, 3 (2018), 218–231.

[9] Gamze Dane, Tao Feng, Floor Luub, and Theo Arentze. 2019. Route choice decisions of E-bike users: Analysis of GPS tracking data in the Netherlands. In *International Conference on Geographic Information Science*. Springer, 109–124.

[10] ETH-Zurich. 2012. Position Data Processing. https://sourceforge.net/projects/posdap/.

[11] Mogens Fosgerau, Emma Frejinger, and Anders Karlstrom. 2013. A link based network route choice model with unrestricted choice set. *Transportation Research Part B* 56 (2013), 70–80. https://doi.org/10.1016/j.trb.2013.07.012

[12] Emma Frejinger, Michel Bierlaire, and Moshe Ben-Akiva. 2009. Sampling of alternatives for route choice modeling. *Transportation Research Part B: Methodological* 43, 10 (2009), 984–994.

[13] Muhammad Ghanayim and Shlomo Bekhor. 2018. Modelling bicycle route choice using data from a GPS-assisted household survey. *European Journal of Transport and Infrastructure Research* 18, 2 (2018).

[14] Katrín Halldórsdóttir, Nadine Rieser-Schüssler, Kay W Axhausen, Otto A Nielsen, and Carlo G Prato. 2014. Efficiency of choice set generation techniques for bicycle routes. *European journal of transport and infrastructure research* 14, 4 (2014), 332–348.

[15] Jeffrey Hood, Elizabeth Sall, and Billy Charlton. 2011. A GPS-based bicycle route choice model for San Francisco, California. *Transportation letters* 3, 1 (2011), 63–75.

[16] Thomas Koch and Elenna Dugundji. 2020. A review of methods to model route choice behavior of bicyclists: inverse reinforcement learning in spatial context and recursive logit. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on GeoSpatial Simulation*. 30–37.

[17] Thomas Koch and Elenna Dugundji. 2021. Limitations of Recursive Logit for Inverse Reinforcement Learning of Bicycle Route Choice Behavior in Amsterdam. *Procedia Computer Science* 184 (2021), 492–499.

[18] Thomas Koch, Luk Knapen, and Elenna Dugundji. 2019. Path complexity for observed and predicted bicyclist routes. *Procedia Computer Science* 151 (2019), 393–400.

[19] Thomas Koch, Luk Knapen, and Elenna Dugundji. 2021. Path complexity and bicyclist route choice set quality assessment. *Personal and Ubiquitous Computing* 25, 1 (2021), 63–75.

[20] Daniel McFadden et al. 1973. *Conditional logit analysis of qualitative choice behavior*. Institute of Urban and Regional Development, University of California .... 105–142 pages.

[21] Gianluca Menghini, Nelson Carrasco, Nadine Schüssler, and Kay W Axhausen. 2010. Route choice of cyclists in Zurich. *Transportation research part A: policy and practice* 44, 9 (2010), 754–765.

[22] Otto Anker Nielsen. 2000. A stochastic transit assignment model considering differences in passengers utility functions. *Transportation Research Part B: Methodological* 34, 5 (2000), 377–402.

[23] Carlo Giacomo Prato, Katrín Halldórsdóttir, and Otto Anker Nielsen. 2018. Evaluation of land-use and transport network effects on cyclists' route choices in the Copenhagen Region in value-of-distance space. *International journal of sustainable transportation* 12, 10 (2018), 770–781.

[24] Nadine Rieser-Schüssler, Michael Balmer, and Kay W Axhausen. 2013. Route choice sets for very high-resolution data. *Transportmetrica A: Transport Science* 9, 9 (2013), 825–845.

[25] Ipek N Sener, Naveen Eluru, and Chandra R Bhat. 2009. An analysis of bicycle route choice preferences in Texas, US. *Transportation* 36, 5 (2009), 511–539.

[26] Danique Ton, Oded Cats, Dorine Duives, and Serge Hoogendoorn. 2017. How Do People Cycle in Amsterdam, Netherlands?: Estimating Cyclists' Route Choice Determinants with GPS Data from an Urban Area. *Transportation research record* 2662, 1 (2017), 75–82.

[27] Danique Ton, Dorine Duives, Oded Cats, and Serge Hoogendoorn. 2018. Evaluating a data-driven approach for choice set identification using GPS bicycle route choice data from Amsterdam. *Travel behaviour and society* 13 (2018), 105–117.

[28] Maëlle Zimmermann, Tien Mai, and Emma Frejinger. 2017. Bike route choice modeling using GPS data without choice sets of paths. *Transportation research part C: emerging technologies* 75 (2017), 183–196.