Delft University of Technology

How do people cycle in Amsterdam, Netherlands?
Estimating cyclists' route choice determinants with GPS data from an urban area

Ton, Danique; Cats, Oded; Duives, Dorine; Hoogendoorn, Serge

DOI
10.3141/2662-09

Publication date
2017

Document Version
Accepted author manuscript

Published in
Transportation Research Record

Citation (APA)
Ton, D., Cats, O., Duives, D., & Hoogendoorn, S. (2017). How do people cycle in Amsterdam, Netherlands? Estimating cyclists' route choice determinants with GPS data from an urban area. *Transportation Research Record*, *2662*(spec. issue: Bicycles), 75-82. Article 09. https://doi.org/10.3141/2662-09

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

How do people cycle in Amsterdam? Estimating cyclists' route choice determinants using GPS data from an urban area

Danique Ton (corresponding author)
Department of Transport & Planning; Faculty of Civil Engineering and Geosciences
Delft University of Technology
Stevinweg 1, PO Box 5048, 2600 GA Delft, The Netherlands
Phone +31 15 278 39 58; Fax +31 15 278 31 79
E-mail d.ton@tudelft.nl

Oded Cats
Department of Transport & Planning; Faculty of Civil Engineering and Geosciences
Delft University of Technology
Stevinweg 1, PO Box 5048, 2600 GA Delft – The Netherlands
Phone: +31 15 278 13 84; fax: +31 15 278 31 79
e-mail o.cats@tudelft.nl

Dorine Duives
Department of Transport & Planning; Faculty of Civil Engineering and Geosciences
Delft University of Technology
Stevinweg 1, PO Box 5048, 2600 GA Delft – The Netherlands
Phone: +31 15 278 63 04; fax: +31 15 278 31 79
e-mail d.c.duives@tudelft.nl

Serge Hoogendoorn
Department of Transport & Planning; Faculty of Civil Engineering and Geosciences
Delft University of Technology
Stevinweg 1, PO Box 5048, 2600 GA Delft – The Netherlands
Phone: +31 15 278 54 75; fax: +31 15 278 31 79
e-mail s.p.hoogendoorn@tudelft.nl

Word count	
Abstract	182
Main text	5525
References	512
Figures (2 x 250)	500
Tables (3 x 250)	750
Total	**7469**

TRR Paper number: 17-01931

Submission Date: March 3, 2017

ABSTRACT

Nowadays, the bicycle is seen as a sustainable and healthy substitute for the car in urban environments. The Netherlands is the leading country in terms of bicycle use, especially in urban environments. Yet route choice models featuring inner-city travel that include cyclists are lacking. This paper estimates a cyclists' route choice model for the inner-city of Amsterdam, based on 3,045 trips collected with GPS data. The main contribution of this paper is the construction of the choice set using an empirical approach which uses only the observed trips in the dataset to compose the choice alternatives. The findings suggest that cyclists are insensitive to separate cycle paths in Amsterdam, which is a city characterized by a dense cycle path network in which cycling is the most prominent mode of travel. In addition, cyclists are found to minimize travel distance and the number of intersections per kilometer. The impact of distance on route choice increases in the morning peak where schedule constraints are more prevalent. Furthermore, overlapping routes are more likely to be chosen by cyclists given everything else being the same.

1. INTRODUCTION

Governments worldwide nowadays acknowledge the advantages of cycling as mode of transport. First, there are health benefits for individual cyclists. Second, the bicycle can help reduce emissions when substituting the car (*1*). Cycling is most attractive in urban areas without large changes in altitude (e.g. the Netherlands or Denmark), where distances covered are relatively small and car usage is often discouraged and associated with greater travel impedance. Furthermore, most European governments have set goals of increasing the modal share of cycling over the next years (*2*).

The Netherlands is the leading country in terms of bicycle use, with 27% of all trips performed by bicycle (*3*). When focusing on the urban environment, the modal share for bicycles increases further, for example in Amsterdam this was 37% in 2011 (*4*). Other cities such as Groningen, Delft and Leiden have a comparable share of bicycle trips (*3, 5*). Despite the fact that so many people cycle in the Netherlands, models aiming at understanding and predicting cyclists' choice behavior are lacking (*6*).

This shift towards cycling, combined with a lack of models incorporating cycling, calls for the development of models to assess related policy implications. Many cities use forecasting models to estimate if, when and where changes to infrastructure or policy are needed. However, these models are still mainly focused on motorized traffic (*7*). The cycling component is either missing, walking and cycling are combined or the model assumes that cars and cyclists behave similarly. Ideally, in forecasting models mode specific activity and route choices are incorporated. Since both choice processes are currently underdetermined, this study starts out by estimating the route choice determinants for cyclists. Before choosing a route, the traveler has already decided to cycle and which activity to perform, therefore the implications of researching route choice first are expected to be minimal.

Recently, a number of studies have estimated bicycle route choice models for locations where bicyce modal shares range between 1% and 6% (*3*). Arguably, the determinants of route choice behavior and their impact might be different from a city such as Amsterdam, where cycling is prominent. These studies used revealed preference (RP) data, more specifically GPS data for estimating the route choice model *(7-10*). Before, most of the data used for model estimation came from stated preference (SP) surveys where the respondents were asked what they would do in a hypothetical situation or route recall surveys where researchers relied on the respondents' ability to recollect chosen routes (e.g. *11, 12*).

This study aims at estimating cyclists' route choice determinants in a context where cycling is the primary mode of transport. Furthermore, the inner-city of Amsterdam is characterized by a densely built area with well-developed cycling infrastructure. This paper presents the findings from a cyclists' route choice model estimated for the inner-city of Amsterdam, using GPS data to identify the determinants influencing route choice in a network dominantly used by cyclists.

This study contributes to the previous cyclists' RP route choice models by introducing a new approach for choice set identification. Previous RP studies have used choice set generation algorithms to identify the feasible choice set from which the cyclist chooses a route. This approach does not guarantee that the chosen route is generated and may include a large number of alternatives that are not selected by any cyclist. Conversely, an empirical approach is proposed which uses only the observed routes to identify the considered choice set. This implies that the chosen route is by definition included in the choice set. Because all routes in the choice set are chosen at least once, it is likely that the alternative routes are considered by the cyclists in

the sample. Furthermore, a behavioral comparison can be made with environments where cyclists make a small minority, because the data is collected in an environment dominantly used by cyclists.

In this paper, Section 2 details the data processing phase, going from GPS data to route alternatives and characteristics. In Section 3, the processed data is analyzed and the results of the estimated route choice models are reported and discussed. Finally, Section 4 provides the conclusions of the paper.

2. DETERMINING ROUTE ALTERNATIVES AND CHARACTERISTICS

This section describes the collection (2.1) and map matching (2.2) of GPS trajectory data. Furthermore, an empirical approach for identifying the route choice set is proposed (2.3), which requires clustering (2.4) and filtering (2.5) of the data. Finally, the potential determinants for cyclists' route choice are discussed (2.6).

2.1. Collection of GPS data

GPS data was collected during a nationwide initiative called the 'Bicycle Counting Week' (BCW), which took place on 14-20 September 2015. The event was organized as a joint initiative of national agencies and companies with the goal of gaining a better insight into the cycling behavior of Dutch cyclists. Nationwide, a total of 38,000 cyclists participated in this initiative. Participants' cycling patterns were tracked using an App. In addition, they once filled in a socio-demographic and travel habit survey to complement the GPS data. Several bicycles were put up for raffle under the participants (*13*).

During the initiative, data of 377,321 cycling trips was collected nationwide. The respondents' sample includes equal shares of male and female participants. The majority of the participants are in the age group 31-65 (80%), while young people (18-) and old people (65+) are underrepresented. This probably stems from the need for using a smartphone to work with the App and the requirement to have consent from ones' parents if younger than 18 years. Most trips registered are work related (69%), explaining why the group of participants aged 31-65 is overrepresented. Participants could mention multiple reasons for cycling. The most dominant reasons mentioned are health (80%), speed (47%) and comfort (46%) (*13*).

As mentioned before, this research focusses on the cycling trips within the city of Amsterdam, where a total of 12,413 trips performed by approximately 5,000 participants were recorded. The Amsterdam sample is similar in terms of gender and age composition to the national sample. However, the share of commuting is higher in Amsterdam (77%). The majority of the cyclists' cycles between 25 and 100km a week (72%), while only 3% cycles less than 10km a week, suggesting that most participants cycle at least to and from their work on a daily basis (*13*). All the cycling trips included in this research are superimposed on the map depicted in FIGURE 1a (Figure 1).

2.2. Map matching the GPS trajectory data

The map matching is executed by the organizers of the BCW, for a more detailed description on this procedure the reader is referred to Van de Coevering et al. (*14*). In the GPS trajectory data, most consecutive GPS data points are measured with an accuracy of 3-4 meters with respect to the infrastructure network. However, outliers up to 50 meters are observed, mainly in dense urban areas. To reduce the impact of these outliers on the analysis, the speed between each two consecutive GPS points is calculated and compared to the actual GPS speed determined by

means of Doppler techniques. If the discrepancy between the actual speed and the computed speed is too large, the GPS records are removed from the dataset (*14*).

The remaining GPS trajectories are matched to the OpenStreetMap network. The map matching algorithm deployed by the BCW organizers generates all possible routes from origin to destination and selects the best match for the GPS records. If no match is found, it could be that links are missing (for example in case of desire lines). In that case the route is partitioned and the same procedure is repeated for the sub-routes (*14*).

2.3. Identifying the considered route choice set

In literature, several approaches for choice set identification have been used, with most studies focused on cycling applying a choice set generation algorithm (e.g. *7, 8*). The aim of these algorithms is to obtain feasible choice sets (*15*), consisting of attractive alternatives. These algorithms however do not guarantee that the chosen route is generated and may include a large number of alternatives that are not chosen by any individual.

An alternative approach for constructing the route choice set is to compile it based on the trips and routes observed in the data. This empirical approach assures that the chosen route is per definition part of the choice set. While the choice set from which each individual cyclist eventually chooses his route (considered choice set (*15*)) cannot be observed directly, it is assumed that the observed alternative routes in the data for a given OD pair are included in this set. Unlike the algorithm approach, the empirical approach implies that not all feasible routes are included in the choice set, but rather only routes that are all actually used by the cyclists in the collected dataset. Consequently, the choice set depends on the observed choices and might thus vary for different samples.

A discrete choice model estimated using the realized routes (empirical approach) is expected to have lower explanatory power than a model estimated based on possible routes (algorithm approach). The first approach identifies alternatives that are chosen by at least one cyclist in the data, whereas the second approach also identifies alternatives that are not chosen. As a result, the offset between the chosen route and the alternatives is smaller when estimating a model using only the realized routes.

Two prerequisites exist for applying the empirical approach: each OD pair considered in the analysis should contain multiple trips and at least two distinct routes. For this study a maximum of 19 realized routes for one OD pair is identified.

2.4. Clustering of the origin and destination GPS data

Since trip origin and destinations are not likely to be recorded at the exact same geographical location when using high-resolution GPS data (approximately 50% in the BCW database), the GPS origin and destination data points are clustered into larger OD pairs, resulting in more trips and possibly more routes per OD pair.

The k-means clustering method is applied, based on the distance between GPS locations of the origins and destinations (*16*). The algorithm minimizes the intra-cluster distances and maximizes the inter-cluster distances. Two downsides of this method are that the solution can get stuck in a local minimum (*16*), which results in a suboptimal distribution of GPS locations over the clusters. Furthermore, in case the number of clusters is set too low, the routes in one OD pair cannot be compared, because the origin or destination points are too far apart. The first downside can be (partially) mitigated by setting multiple starting points for the algorithm. This way it is less likely to converge into a local minimum.

This method was applied for different k-values; 150, 200, 250 and 300 clusters. If the number of clusters is set too high, the number of trips per cluster becomes too low and the

advantages of clustering the trips diminish. As mentioned before, if the number of clusters is set too low, routes in one OD pair cannot be compared. We find that defining 200 clusters provides the best balance between intra-cluster distance and number of trips per OD pair for the BCW dataset. The number of random starting points is set to 20. FIGURE 1b shows the geographical distribution of the cluster-centers over the inner-city of Amsterdam.

The 200 clusters result in a maximum intra-cluster distance (i.e. diameter) of 444 meters, while the average is 168 meters. The cluster with the largest diameter is located around a park, however the routes chosen are still comparable. Therefore, this is an acceptable diameter for a rather dense network. After clustering, only 30% of the OD pairs consists of one trip, instead of 50% before clustering.

2.5. Data filtering process

Not all trips in the dataset can be used, mostly because of how the choice set is composed. Therefore, several filtering steps are necessary (see FIGURE 2). In the BCW dataset many cycling trips are made in the inner-city, whereas the density of cycling trips in the suburbs is very low. Therefore, only the trips (partially) traversing the inner-city are used, which limits the available trips to 7,984. Not all trips are included completely, because the boundaries of the inner-city are specified on GPS coordinate level and not on trip level. It is, for example, possible that one trip crosses the inner-city more than once. In this case the trip is split into multiple trips. This demarcation means that some cyclists are observed during the entire trip, whereas others are only observed during part of the trip. We assume that the route choice for a section of the route is not fundamentally different from choosing the complete route.

Due to splitting trips some very short routes are created, for which it is unlikely that route choice is possible. Therefore, a filter is applied on the possibility for route choice, which is defined here as crossing at least two intersections during the trip, resulting in 8,847 trips. When applying the empirical approach to identify the choice set, it is necessary to filter out all OD pairs with only one trip, resulting in 6,208 trips. Also, more than one route needs to be chosen per OD pair. The result is a final dataset of 3,045 trips (see FIGURE 1a). Since other GPS based route choice models have been estimated using less trips (*7-10*), the filtered data set seems large enough to estimate a route choice model for cyclists in inner-city areas. Furthermore, the initial dataset and the final dataset show similar patterns with respect to time of departure and day of travel. The distances covered are slightly larger in the initial sample, due to the geographical demarcation of the inner-city. However, no structural behavioral issues are expected due to the filtering process (Figure 2).

2.6. Potential determinants of cyclists' route choice

Previous research has identified a wide range of attributes that might influence the route choice behavior of cyclists, where the attributes selected for research mainly depend on the type of data used (RP or SP) and the availability of the data (in case of RP). Both Hunt & Abraham (*12*) and Sener et al. (*17*) have reviewed many (mostly SP) studies to find attributes that potentially influence bicycle route choice. Based on these reviews (*12, 17*) and previous RP studies (*7-10*) three categories of explanatory variables are identified: individual, network and contextual attributes. TABLE 1 shows an overview of all attributes, including how they influence route choice for cyclists (Table 1).

Individual attributes are commonly incorporated in SP studies, mainly as interaction terms, to identify differences in attitude between individuals with respect to network attributes. Looking at RP studies, this means that next to observing actual behavior, a questionnaire for

socio-demographics is necessary. Although, the privacy of the respondent needs to be preserved. In the RP studies, Hood et al. (*7*) have included gender and cycling experience in their model, but for example Menghini et al. (*8*) did not have these personal attributes at the individual level.

The network attributes that were found to be most influential on route choice behavior are distance, gradient and cycle path percentage (e.g. *8, 9*). Regarding gradient different approaches are applied in literature. For example, Broach et al. (*9*) divided sections of the route into different categories of up-slope, whereas Menghini et al. (*8*) adopted the maximum gradient of the route. With respect to cycle paths, Furth (*18*) identifies four categories: shared streets and lanes, cycling lanes, separate cycle paths and standalone paths. Menghini et al. (*8*) only take into account the third category, whereas Hood et al. (*7*) take the first, second and fourth category into account.

Contextual attributes are mostly found in SP studies, however also in RP studies trip purpose is found to be influential (e.g. *7, 9*). Commuting cyclists tend to value distance more negative compared to other purposes.

Based on the literature and the constraints on the availability of information, the following attributes are selected for this study: distance, percentage of separate cycle paths (third category (*18*)), number of intersections, rain, sunset and sunrise times and trip purpose. Due to privacy issues, the BCW dataset does not contain any personal information at the individual level.

3. ESTIMATING A CYCLISTS' ROUTE CHOICE MODEL

This section provides the analysis of the data collected for the inner-city of Amsterdam. First the descriptive statistics for the trips collected in the inner-city are presented (3.1). Then, the specification of the estimated models is described (3.2) and the results of the model estimations are discussed (3.3).

3.1. Analysis of the trips cycled in Amsterdam

For this study the selected network attributes are distance, percentage of separate cycle path and number of intersections per km. TABLE 2 shows the range, mean and standard deviation of these attributes for all alternatives. As can be expected from the restriction of the case study area to the inner-city, the average route distance is relatively small. However, the longest route is relatively long as it exceeds 6km. Separate cycle paths are only encountered on roads with a speed limit of 50 km/h or higher. In the inner-city cyclists share roads with motorized traffic and large volumes of pedestrians (*18*). Therefore, a low percentage of cycle paths is found along the routes of the participants (36%). The number of intersections crossed per km also varies largely and is, as can be expected in a dense urban area, fairly high (Table 2).

The selected contextual attributes are translated into dummy variables. Even though the trip purpose is unknown, two proxy variables can be derived. Firstly, the time of day at which the trip has started is an indicator for commuting to or from work or school (peak hours) versus recreational or social trips (off-peak hours). Secondly, the trip type can be an indicator for cycling only trips or access and egress as part of a multimodal trip. Two train stations are situated within the inner-city boundaries; Amsterdam Centraal and Amsterdam Muiderpoort. Trips starting at one of these stations are considered egress and the trips ending at these stations are considered access, relative to the multimodal trip.

Only 14% of the trips are undertaken in darkness. Most trips are cycled (28%) during the morning peak hours from 7AM to 10AM, followed by trips during daytime from 10AM to 5PM (27%). Almost half of the trips experienced rain showers (46%). Access and egress are equally

represented in the dataset (each 9%), implying that most trips in the dataset are cycling only (not directed to or from a train station).

3.2. Specification of the route choice models

The most commonly used model to estimate cyclists' route choice, like estimated by Casello & Usyukov (*10*), is the MNL model. This model assumes that cyclists interpret each route as a distinct alternative (independence of irrelevant alternatives), while in reality the perception of routes that share common links might be correlated, implying that the MNL model will inadequately assign high probabilities to overlapping routes. The routes included in this study exercise some degree of overlap, therefore violating this assumption.

To account for overlapping routes, multiple solutions have been proposed in literature. The model structure applied in other cyclists' route choice studies is the PSL model (e.g. *7-9*), which introduces a similarity measure in the utility function to account for the overlap. This approach maintains the MNL structure, making it easy to compute. For the calculation of the path size (PS) factor, different approaches have been put forward, however no straightforward answer can be provided to the question which performs best. For example, the PS factors developed in a later stage can have illogical route probabilities (*19*), whereas the earlier versions of the PS factor do not take large differences in route length into account (*20*). In this study the path size factor put forward by Ben-Akiva & Bierlaire (*20*) is adopted, because no large deviations in route lengths are present in our study:

$$PS_{in} = \sum_{a \in \Gamma_i} \left(\frac{l_a}{L_i}\right) \frac{1}{\sum_{j \in C_n} \delta_{aj}} \tag{1}$$

Where Γ_i is the set of links in route i, l_a is the length of link a, L_i is the length of route i and δ_{aj} the link-route incidence variable which equals one if link a is on route j and zero otherwise. The probability of choosing route i given choice set C_n is specified the following way (*20*):

$$P(i|C_n) = \frac{e^{\left(\beta_d * Distance_{in} + \beta_{int} * \frac{Intersections}{km}_{in} + \beta_{PS} * \ln PS_{in}\right)}}{\sum_{j \in C_n} e^{\left(\beta_d * Distance_{jn} + \beta_{int} * \frac{Intersections}{km}_{jn} + \beta_{PS} * \ln PS_{jn}\right)}} \tag{2}$$

Where β_{PS} equals 0 when estimating a MNL model and PS is the path size factor calculated in Equation 1. PS lies between 0 and 1, where 1 means no overlap and 0 full overlap. The natural logarithm of PS is then negative. In this study both the MNL and PSL modeling structure are adopted in order to determine the effect of overlap on the route choice of cyclists. The models are estimated using the BIOGEME package (*21*).

3.3. Estimated cyclists' route choice models

Both a MNL and PSL model are estimated, in order to test for the effect of overlap in the model. To come to these models, all network attributes have been included in the model estimation, and insignificant attributes have been removed to find the most efficient model. For the third model, a stepwise approach is used to add context attributes to the model as interaction terms when a significant and interpretable result is found, resulting in the extended PSL model.

3.3.1. Discussion of the modeling results

The results of three model estimations are summarized in TABLE 3. In the MNL model both distance and the number of intersections per km have a significant influence on route choice. Increasing the average distance with one percent results in a 0.50% decrease of being chosen (ceteris paribus) and increasing the average number of intersections per km with one percent

results in a decrease of 0.53%. Cyclists prefer fewer intersections per km, our hypothesis is that either they want to reduce interaction with other road users and avoid delays or they want to reduce the cognitive effort during the trip. Translating this to the network of Amsterdam, cyclists avoid the historical city center and dense residential areas due to the presence of many intersections per km and they prefer the ring streets because of fewer intersections per km. Cyclists in Amsterdam are willing to cross 7.73 more intersections for a one kilometer shorter route, which is fairly high but reasonable for an urban environment (Table 3).

In the PSL model, the path size term is added. As this term is calculated based on overlap in terms of distance, this factor decreases the impact of distance on the total utility, which is now only significant on a 90% confidence level. Increasing the number of intersections by one percent reduces the probability of being chosen by 0.43%, while one percent increase in distance reduces this probability by 0.36%. The impact of the path size factor depends on the degree of overlap. One percent increase for nearly unique routes increases the probability of being chosen by approximately 0.3%, whereas for routes that are almost identical to other routes this is 2.7%. For a route that is one kilometer shorter, cyclists are willing to cross 6.28 intersections, which is slightly lower than in the MNL model.

The PSL model has been extended to include context attributes as interaction terms. The time of day, in particular the morning peak, was found the only significant explanatory variable. The other contextual attributes (rain, sunset and sunrise times and access/egress) did not yield any interpretable significant influence on the network attributes. Morning peak hours (7AM-10AM) are characterized by commuters heading to work, where schedule constraints are more likely. Model estimates show that morning trips are characterized by a significantly greater repelling effect for distance compared to other times of the day. One percent increase in the average distance results in a decrease in the choice probability of 3.4% for cyclists travelling in morning peak and only 0.1% for other times. In this model, cyclists travelling during morning peak are willing to cross 20.07 more intersections for a one kilometer shorter route, whereas during other times this is only 1.97 intersections. The differences in the tradeoff clearly show the aversion towards distance of cyclists during morning peak hours.

The path size parameter for the PSL models is significant and negative, indicating that paths that have a high degree of overlap are more likely to be chosen than others (ceteris paribus). Previous studies estimating cyclists' route choice models found a significant positive path size parameter (e.g. *7, 9*). Therefore, this finding may seem counterintuitive at first, as it does not penalize the routes that overlap but rather increases their choice probability. However, there is evidence that overlapping routes are sometimes valued higher than non-overlapping routes. This because overlap can reduce the uncertainty of the route followed, as was for example found by Lam and Xie (*22*) in the context of public transport. Cyclists might prefer routes that offer more downstream decision points to improve route choice robustness. In addition, this might be a result of the characteristics of this case study. In Amsterdam the radial routes provide the backbone of many attractive routes, causing overlapping routes to be valued positively. More behavioral research is needed in order to draw more general conclusions.

3.3.2. Comparison of model structures

The PSL model performs significantly better than the MNL model on a 95% confidence level based on the log-likelihood ratio test ($12.64 > \chi^2$). Furthermore, the model fit for the PSL model is higher. This indicates that including the path size factor to incorporate overlap in the model is beneficial for the interpretation of the results and the prediction of route choice for cyclists. The PSL modeling structure is therefore considered more suitable for estimating route choice models

than the MNL structure. Furthermore, the extended PSL model performs significantly better than the PSL model at a 95% confidence level ($4.88 > \chi^2$), meaning that interaction term increases the model fit. The extended PSL model is therefore the best of the three models.

3.3.3. Model fit

The model fit for all models is very low. In previous studies, where choice set generation algorithms were applied, model fit varied between 23 to 28 percent, significantly better than in this study (*7-9*). As mentioned before, our hypothesis is that estimating discrete choice models using the empirical approach for composing the choice set, results in a low model fit. Experiments with adding fictional route alternatives that are inferior to the one most commonly selected confirm that model goodness-of-fit improves substantially by artificially enlarging the choice set. This indicates that the application of a generation algorithm leads to over fitting of the data. Furthermore, variance over the alternatives is low in the dataset, most likely due to the fact that cycling costs effort. For example, the shortest route is chosen in 32.6% of the cases, and in 41.4% of the cases the distance of the chosen route is only 10% more than the shortest route, which means on average only 0.2km difference. This implies that it is more difficult to estimate a distance coefficient in the model estimation when constructing the choice set using the empirical approach compared to the algorithm approach.

4. CONCLUSIONS

This paper presented the findings of a cyclists' route choice model estimated for the inner-city of Amsterdam, aimed at identifying the determinants influencing route choice in a network where cycling is the primary travel mode. Choice models were estimated based on detailed GPS data comprising more than 3,000 trips performed over the course of one week in September 2015.

It is possible to estimate a route choice model for cyclists based on only GPS trajectory data. The results of the estimated route choice models are mostly in line with literature (*7-10*). However, previous cyclists' route choice studies that have used GPS data found that the percentage of separate cycle paths is a very important factor for route choice (*7-10*), whereas this study finds no such significant relation. This is presumably due to guidelines for Dutch infrastructure, where cyclists are specifically taken care of. For example, cyclists and motorized traffic are only mixed on streets where the speed limit is 30 km/h and the traffic volume is under 4000 vehicles/day, mitigating the safety risks (*18*). This finding suggests that when cycling is indeed well-established, separate cycle paths do not necessarily attract cyclists. This might however be due to the location of the cycle paths in the network, which is on the ring streets and not in the center of the city. More research is necessary for studying how an increasingly dense network of bike paths might lead to a reduction in their importance as route choice determinant. Our results for distance are overall in line with previous studies, although the impact of distance is less pronounced than in other studies (*7-10*). Previous studies calculated more specific attributes related to the number of intersections per km, like number of turns per km, number of signalised intersections per km and number of stop signs per km (e.g. *7, 9*), they were all found to influence the route choice behaviour of cyclists negatively, however a proper comparison cannot be made. Distance and the number of intersections per km are evidently important regardless of the level of penetration of the cycling. During the morning peak, when people cycle to work or school, distance looms more negatively than during other times of the day, which is consistent with the findings reported by Broach et al. (*9*).

In this study both the MNL and PSL modeling structure are adopted in order to determine the effect of overlap on the route choice of cyclists. The effect of taking overlap into account in

the model estimation is large as it increases the explanatory value of the model. Routes that are overlapping are valued higher than non-overlapping routes, several explanations can be found for this phenomenon. First of all, an empirical approach for choice set identification is adopted in this study instead of the often used path generation algorithms. This approach allows us to overcome the common shortcomings of not generating the chosen route and having a large number of non-chosen alternatives, by using only the observed routes per OD pair in the dataset in constructing the choice set. Some links are attractive to all cyclists, probably because they form the most direct path to the destination (*23*) or they could have some non-observed advantage. Consequently, it is likely that cyclists choose routes that include these links and because the observed routes form the basis of our choice set, routes with a higher degree of overlap are common and preferred. Another explanation is that the uncertainty of the chosen route is lower when routes overlap and alternatives are present, this can be especially helpful when for example road works are encountered. This explanation also relates to the physical effort needed for cycling. The alternatives available near overlapping routes are usually similar in terms of physical effort, whereas a non-overlapping route might require more physical effort (e.g. longer distance).

The use of the empirical approach for identifying the choice set has its limitations. In particular, the choice set depends on the observed choices and might thus vary for different samples. In addition, the low model fit of the estimated models is attributed to the use of the empirical approach as confirmed by experimenting with the addition of fictive routes. Finally, the positive value found for overlapping routes might be the result of adopting this approach, however this approach should be tested on other datasets in order to draw a more definitive conclusion.

This study was the first to include data from a city where cycling is a well-established and prominent travel mode. Our findings suggest that there are noticeable differences between this case study area which has few comparable cases, and cities where cycling is almost absent.

We recommend also including socio-demographic variables, such as gender, age and cultural background into future data collection and analysis in order to allow identifying their importance. Furthermore, we expect that including more network attributes will help improve interpretation, practical applicability and model fit. Also, for future research we would like to explore more modeling structures, as they might be better suitable for modeling cyclists' route choice. Possible interesting structures are latent class, nested logit and mixed logit. Next to that, we want to explore the sensitivity of the estimation results to the generated choice sets using the empirical approach. Furthermore, we are interested in testing how individual knowledge and familiarity with the network influences route choice when cycling, we expect that this will help understanding the relationship with overlapping routes. Moreover, nowadays more and more people use mobile devices to plan activities and routes, potentially influencing how they travel. Finally, cycling route choice models can be integrated into an activity scheduling and mode choice model, in order to assess their inter-relation with other modes in transport demand forecasting.

ACKNOWLEDGEMENTS

This research was supported by the Allegro (Unravelling slow mode travelling and traffic: with innovative data to create a new transportation and traffic theory for pedestrians and bicycles) project which is financed by the European Research Council and Amsterdam Institute for

Advanced Metropolitan Solutions. The data for this research was provided by the initiators of the 'Fiets Telweek'.

REFERENCES

1. Pan-European Programme. (2014). *Fourth high-level meeting on transport, health and environment*. Paris: World Health Organisation & United Nations. http://www.unece.org/fileadmin/DAM/thepep/documents/ Déclaration_de_Paris_EN.pdf. Accessed May 18, 2016.
2. Kuester, F. (2015). *National cycling policies*. Retrieved from European Cyclists' Federation: https://ecf.com/what-we-do/cycling-all-policies/national-cycling-policies. Accessed May 20, 2016.
3. Pucher, J. and Buehler, R. (2008). Making cycling irresistible: Lessons from The Netherlands, Denmark and Germany. *Transport Reviews*, 495-528.
4. OViN. (2011). *Onderzoek Verplaatsingsgedrag Nederland 1986-2012*. Den Haag: CBS, [in Dutch].
5. Harms, L., Bertolini, L., & Te Brommelstroet, M. (2016). Performance of municipal cycling policies in medium-sized cities in the Netherlands since 2000. *Transport Reviews*, 134-162.
6. Verkeersnet. (2015, December 29). *Onderzoek naar betere fietsmodellen*. Verkeersnet: http://www.verkeersnet.nl/17820/onderzoek-naar-betere-fietsmodellen/. Accessed April 12, 2016. [in Dutch]
7. Hood, J., Sall, E. and Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California. *Transportation Letters*, 63-75.
8. Menghini, G., Carrasco, N., Schussler, N., & Axhausen, K. (2010). Route choice of cyclists in Zurich. *Transportation Research Part A*, 754-765.
9. Broach, J., Dill, J. and Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A*, 1730-1740.
10. Casello, J. M. and Usyukov, V. (2014). Modeling cyclists' route chocie based on GPS data. *Transportation Research Record: Journal of the Transportation Research Board*, 155-161.
11. Howard, C. and Burns, E. K. (2001). Cycling to work in Phoenix: Route choice, travel behavior, and commuter characteristics. *Transportation Research Record: Journal of the Transportation Research Board*, 39-46.
12. Hunt, J. D. and Abraham, J. E. (2007). Influences on bicycle use. *Transportation*, 453-470.
13. Fiets Telweek. (2015). *Regiobrochure Nationale Fiets Telweek 2015*. Breda: Fietstelweek, [in Dutch].
14. van de Coevering, P., de Kruijf, J. and Bussche, D. (2014). *Bikeprint: Policy renewal and innovation by means of tracking technology*. Eindhoven: Colloquium Vervoersplanologisch Speurwerk, [in Dutch].
15. Hoogendoorn-Lanser, S. (2005). *Modelling travel behaviour in multi-modal networks*. PhD Thesis. Delft University of Technology.
16. Harrington, P. (2012). Clustering unlabeled items using k-means clustering. In P. Harrington, *Machine learnign in Action* (pp. 207-223). Shelter Island: Manning.

17. Sener, I. N., Eluru, N. and Bhat, C. R. (2009). An analysis of bicycle route choice preferences in Texas, US. *Transporation*, 511-539.
18. Furth, P. G. (2012). Bicycling infrastructure for mass cycling: A transatlantic comparison. In: J. Pucher, & R. Buehler, *City Cycling* (pp. 105-140). Cambridge, USA: MIT Press.
19. Freijinger, E. and Bierlaire, M. (2007). Capturing correlation with subnetworks in route choice models. *Transportation Research Part B*, 363-378.
20. Ben-Akiva, M. and Bierlaire, M. (1999). Discrete choice methods and their applications to short term travel decisions. In: W. Hall, *Handbook of Transportation Science* (pp. 5-34). Dordrecht, The Netherlands: Kluwer.
21. Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models. *Proceedings of the 3rd Swiss Transportation Research Conference*. Ascona, Switzerland.
22. Lam, S.-H. and Xie, F. (2002). Transit path-choice models that use revealed preference and stated preference data. *Transportation Research Record: Journal of the Transportation Research Board*, 58-65.
23. Bovy, P. H. and Stern, E. (1990). *Route choice: Wayfinding in transport networks.* Dordrecht, The Netherlands: Kluwer.

LIST OF TABLES

TABLE 1 Attributes and their Influence on Cyclists' Route Choice, based on findings in (*7-10, 12, 17*)
TABLE 2 Descriptive Statistics of Cycling Trips in Amsterdam
TABLE 3 Estimated Cyclists' Route Choice Models

LIST OF FIGURES

FIGURE 1 (a) The Network of Amsterdam used for Cycling Trips. In the center lies the Historical City, surrounded by the Ring Canal streets and the Radial Roads heading to and from the City. To the North lies the river IJ, with two Ferries connecting its Shores. (b) All Origin and Destination Cluster Centers resulting from the K-Means Clustering Algorithm.
FIGURE 2 Data Filtering Process

TABLE 4: Attributes and their Influence on Cyclists' Route Choice, based on findings in (*7-10, 12, 17*)

Individual attributes		***Network attributes***		***Contextual attributes***	
Gender	-	Distance	Negative	Sunset & Sunrise times	-
Age	-	% of cycle path	Positive	Weather (rain)	-
Cycling experience	-	Gradient	Negative	Crime rate / Safety	-
Income	-	Travel time	Negative	Aesthetics (Canal / Park)	-
Household size	-	Travel speed	Positive	Sweeping / Snow ploughing	-
		Maximum speed (cars)	Negative	Cycling season	-
		# Stop signs	Negative	Trip purpose	-
		# Intersections	Negative		
		# Bridges	Positive		
		# (Left) turns	Negative		
		% Wrong way	Negative		
		Pavement surface quality	Positive		
		Continuity of cycle paths	Positive		
		Traffic volume (cars)	Negative		
		On-street parking	Negative		
		# Traffic lights	Positive (*8*), Negative (*9*)		

- *Not estimated as a separate attribute*

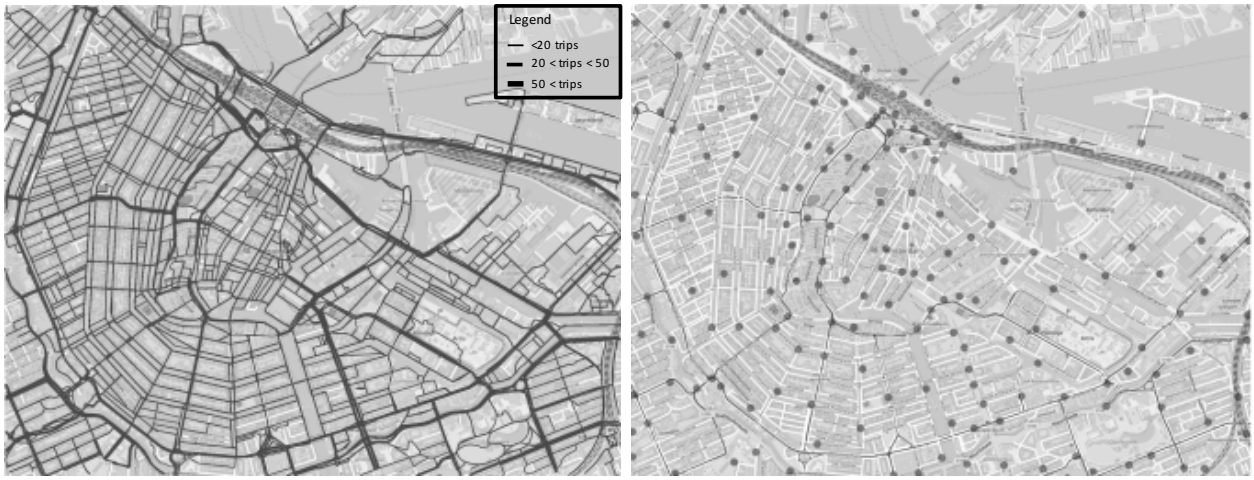
TABLE 5: Descriptive Statistics of Cycling Trips in Amsterdam

Attribute	***Description***	***Range***	***Mean***	***Standard Deviation***
Distance (km)	Route length	0.13 – 6.69	1.96	1.02
Percentage of separate cycle path	Percentage of the route with a cycle path which is separated from motorized traffic	0% – 100%	36.2%	25.5%
Number of intersections per km	Average number of intersections crossed per km (straight and turn)	1.75 – 50.8	16.8	5.8

TABLE 6: Estimated Cyclists' Route Choice Models

	MNL model		PSL model		Extended PSL model	
	Coef.	**t-stat**	**Coef.**	**t-stat**	**Coef.**	**t-stat**
Distance (km)	-0.255	-2.36**	-0.182	-1.65*	-0.057	-0.48
Morning peak	-	-	-	-	-0.525	-1.77*
# Intersections/km	-0.033	-4.84**	-0.029	-4.00**	-0.029	-4.09**
Ln (Path Size)	-	-	-0.252	-2.13**	-0.248	-2.11**
Adjusted rho-square	0.003		0.005		0.005	
Likelihood ratio	31.165		43.809		48.694	
Initial Log-likelihood	-4,167.376		-4,167.376		-4,167.376	
Final Log-likelihood	-4,151.794		-4,145.472		-4,143.030	
# Observations	3045		3045		3045	
Number of parameters	2		3		4	

**Significant on 90% confidence level **Significant on 95% confidence level*



(a) **(b)**

FIGURE 3: (a) The Network of Amsterdam used for Cycling Trips. In the center lies the Historical City, surrounded by the Ring Canal streets and the Radial Roads heading to and from the City. To the North lies the river IJ, with two Ferries connecting its Shores. (b) All Origin and Destination Cluster Centers resulting from the K-Means Clustering Algorithm.

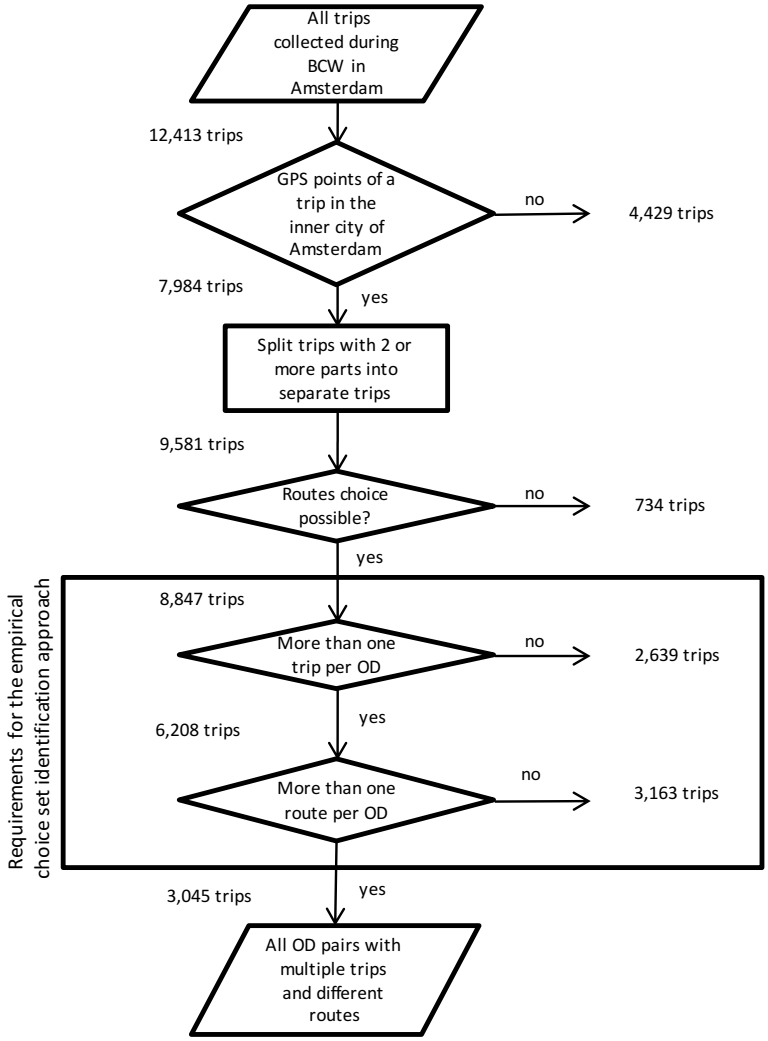


FIGURE 4: Data Filtering Process