



Delft University of Technology

Evaluating a data-driven approach for choice set identification using GPS bicycle route choice data from Amsterdam

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

DOI

[10.1016/j.tbs.2018.07.001](https://doi.org/10.1016/j.tbs.2018.07.001)

Publication date

2018

Document Version

Final published version

Published in

Travel Behaviour and Society

Citation (APA)

Ton, D., Duives, D., Cats, O., & Hoogendoorn, S. (2018). Evaluating a data-driven approach for choice set identification using GPS bicycle route choice data from Amsterdam. *Travel Behaviour and Society*, 13, 105-117. <https://doi.org/10.1016/j.tbs.2018.07.001>

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.



Evaluating a data-driven approach for choice set identification using GPS bicycle route choice data from Amsterdam



Danique Ton^{a,*}, Dorine Duives^a, Oded Cats^a, Serge Hoogendoorn^a

^a Department of Transport & Planning, Delft University of Technology, Stevinweg 1, PO Box 5048, 2600 GA Delft, The Netherlands

ARTICLE INFO

Keywords:

Data-driven choice set generation
Cyclists' route choice
Travel behaviour analysis comparison
Breadth-first search on link elimination
Labelling

ABSTRACT

Specifying the choice set for travel behaviour analysis is a non-trivial task. Its size and composition are known to influence the results of model estimation and prediction. Most studies specify the choice set using choice set generation algorithms. These methods can introduce two types of errors to the specified choice set: false negative (not generating observed routes) and false positive (including irrelevant routes). Due to increased availability of revealed preference data, like GPS, it is now possible to identify the choice set using a data-driven approach. The data-driven path identification approach (DDPI) combines all unique routes that are observed for one origin-destination pair into a choice set. This paper evaluates this DDPI approach by comparing it to two commonly used choice set generation methods (breadth-first search on link elimination and labelling). The evaluation considers the three main purposes of choice sets: analysis of alternatives in the choice set, model estimation and prediction. The conclusion is that the DDPI approach is a useful addition to the current choice set identification methods. The findings indicate that in analysing alternatives in the choice set, the DDPI approach is most suitable, as it reflects the observed behaviour. For model estimation the DDPI approach provides a useful addition to the current choice set generation methods, as it provides insights into the preferences of individuals without requiring network-data for additional information or generating routes. In terms of prediction, the DDPI approach is not suitable, as it is not able to perform well with out-of-sample data.

1. Introduction

In the context of travel behaviour, many choices must be made by an individual before a trip is made, e.g. destination, mode and route choice. These choices are all discrete in nature, meaning that only one option can be chosen at a time. The choice set from which an individual chooses one, forms an important aspect in the analysis of travel behaviour. Three different purposes of choice sets can be identified. First, it is essential in analysing different travel options in the network (e.g. number of alternatives, characteristics or composition of the alternatives), second it is used for demand model estimation (estimating behavioural parameters), and third it is instrumental in predicting choice probabilities and thereof flow distribution over alternatives/the network (Bovy, 2009). The size and composition of the choice set influence the results of the model estimation and prediction, and consequently the interpretation of the estimated behavioural parameters (Bovy, 2009). This issue is for example relevant in route choice analysis, as many possible alternatives can be identified by the researcher, but only few will be known to the individual, leading to possible mismatches in the choice set identification.

Route choice sets are often specified using *choice set generation algorithms* (e.g. k-shortest paths or labelling), which compute a set of routes based on characteristics of the network (-links) (e.g. distance or travel time). The use of these algorithms can introduce two types of errors in the choice set: false negative and false positive errors. False negative errors arise when the algorithm is not able to reproduce the chosen alternatives. The generated alternatives might not match the behaviour and preferences of the individual, and as a result the chosen route is not reproduced. The impact of this error decreases when the ability of the choice set generation algorithm to capture the individuals' behaviour and preferences increases. False positive errors occur when a choice set generation algorithm also generates routes that are not considered by the individual, resulting in a too large choice set. In conclusion, the use of choice set generation algorithms potentially comes with several flaws.

In recent years, large improvements have been made in revealed preference data collection methods. New data sources, such as GPS data that contain detailed spatial and temporal information on the movement pattern of individuals, help creating insights into the individuals' choice behaviour. By combining the GPS records belonging to one

* Corresponding author.

E-mail addresses: d.ton@tudelft.nl (D. Ton), d.c.duives@tudelft.nl (D. Duives), o.cats@tudelft.nl (O. Cats), s.p.hoogendoorn@tudelft.nl (S. Hoogendoorn).

individual into separate trips, the observed trips can be used for route choice research (e.g. Menghini et al., 2010; Hood et al., 2011). Next to generating the choice set based on a set of assumptions on network properties, it is then also possible to use the observed trips from GPS data to identify the choice set directly. Every trip between an origin and destination follows a certain route, the unique routes that are observed can then be combined into one choice set. Consequently, the potential false negative error associated with choice set generation algorithms cannot occur and the potential false positive error is negligible because all the routes included in the choice set have been chosen by the individual.

Governments worldwide have shown increasing interest in promoting and understanding cycling usage, due to the potential health, congestion and emissions benefits. Consequently, goals have been set to increase the cycling modal share (Pan-European Programme, 2014). Several studies investigated bicycle route choice using GPS data, primarily in areas where cycling is relatively scarce, with the goal of identifying determinants that influence route choice, so that substantiated infrastructure investments can be made (Menghini et al., 2010; Hood et al., 2011; Broach et al., 2012; Casello and Usyukov, 2014; Montini et al., 2017; Zimmerman et al., 2017; Chen et al., 2017; Li et al., 2017; Ghanayim and Bekhor, 2018). Other studies have taken in place in urban environments with a larger share of cyclists, like Copenhagen (Halldórsdóttir et al., 2014; Prato et al., 2018; Skov-Petersen et al., 2018). These studies have applied different types of choice set generation algorithms, such as labelling, stochastic methods, link elimination, and link penalty. However, none of the studies has applied a data-driven method for choice-set identification as proposed and examined in this study. This approach is applied to a bicycle route choice study for the city of Amsterdam, the Netherlands (Ton et al., 2017). Amsterdam is known for its well developed bicycle infrastructure and high share of bicycling trips (37%) (OVIN, 2011). To evaluate the potential of this data-driven method for choice set identification, we compare the method using the dataset from the city of Amsterdam, to other choice set generation algorithms previously applied in the cycling route choice literature.

This paper evaluates the use of a data-driven approach for choice set identification in travel behaviour analysis. The goal is to investigate whether a data-driven approach can be a valuable addition to the current choice set identification methods. Bicycle GPS data from Amsterdam, the Netherlands, is used to identify the choice set and this choice set is used in the estimation and validation of a route choice model. The evaluation of the data-driven approach is done by means of a comparison study, where it is compared to two commonly used choice set generation methods, to assess and compare their performance and results. Based on computation time, sensitivity to false negative errors and, number of applications, two approaches have been selected: the breadth-first search on link elimination (BFS-LE) introduced by Rieser-Schussler et al. (2013) and the labelling approach introduced by Ben-Akiva et al. (1984). The evaluation is performed on the three above-mentioned purposes of choice sets; 1) analysing the composition of the choice set, 2) understanding behaviour (model estimation) and 3) application of the model on out-of-sample data (model validation).

The rest of the paper is outlined as follows. Section 2 reviews contemporary choice set generation procedures. In Section 3, the data-driven approach is elaborated upon in terms of requirements of data, opportunities, limitations of the method, and sensitivity with respect to data collection duration. Section 4 describes the methodology for evaluating the specified choice sets as well as the route choice model estimation and validation. Section 5 provides background on the data that was collected and prepared for this study. Section 6, then details the evaluation of the generated choice sets in comparison to the observed routes and Section 7 covers the evaluation of the route choice model estimation and validation. Finally, Section 8 concludes the paper and provides directions for future research.

2. Choice set generation methods

This section discusses different choice set generation methods that have been proposed in the past and selects two methods as reference for the evaluation of the data-driven approach.

Many different methods have been proposed for identifying route choice sets (for detailed reviews see Fiorenzo-Catalano (2007) and Ramming (2002)). Bovy (2009) and Prato (2009) identify four categories of choice set generation methods: deterministic methods, stochastic methods, probabilistic methods and constrained enumeration methods. Most choice set generation methods belong to the *deterministic category* and consist of repeated shortest path searches in the network. These shortest path methods have different input variables such as search criteria, route constraints and link impedance (Prato, 2009). They are computationally attractive due to the efficiency of shortest path algorithms. *Stochastic methods* are also based on repeated shortest path searches, but additionally the computation of optimal paths is randomised based on link impedances or individual preferences drawn from probability distributions, mostly done using simulation. These methods have been applied in the bicycle route choice context by Hood et al. (2011), Halldórsdóttir et al. (2014), Ghanayim and Bekhor (2018), and Prato et al. (2018). *Constrained enumeration methods* are not only based on shortest routes, but also make additional behavioural assumptions (Prato, 2009). These assumptions reflect different behavioural thresholds that can be specified, e.g. excluding loops and only including links that bring the individual closer to the destination. These methods have been applied in the bicycle route choice context by Halldórsdóttir et al. (2014), but did not prove to outperform the deterministic or stochastic methods. *Probabilistic methods* assign a probability for each alternative to be included in the choice set. A fully probabilistic approach, as proposed by Mansky (1977), which includes the choice set generation and selection in the utility function, is often deemed infeasible due to its computational complexity. As a consequence, these methods have not yet been applied in the bicycle route choice context.

Recently, two alternative approaches have been proposed that address the choice set identification implicitly (i.e. no need for explicit enumeration of alternatives). The first is the *sampling approach* (Frejinger et al., 2009; Flötteröd and Bierlaire, 2013), that assumes a universal choice set and by means of importance sampling selects a subset of these routes. The second approach is the *link-based approach* (Fosgerau et al., 2013), which assumes that individuals make successive choices at each node. The link-based approach was applied in the bicycle route choice context by Zimmerman et al. (2017).

Due to its prevalence in the general and bicycle route choice literature, computational efficiency and deterministic nature (which relates more to the cognitive aspects of the decision-maker rather than conceived a computational instrument), *deterministic methods* are selected as reference methods for comparison in this study. Four categories of deterministic methods are identified: shortest paths, link elimination, labelling and link penalty. Previous findings suggest that the *shortest path methods* have the lowest performance in terms of reproducing the observed routes (Bovy, 2009). Furthermore, the *link penalty methods* are known for their large computation times (Bekhor et al. 2006). Therefore, the focus lies with the link elimination and labelling methods.

The *link elimination method* iteratively removes links that are on the shortest path and finds new shortest paths (Bellman and Kalaba, 1968). Prato and Bekhor (2007), Bekhor et al. (2006), and Ghanayim and Bekhor (2018) evaluated this approach and found that in about 40% of the cases false negatives are produced. Azevedo et al. (1993) proposed an alternative approach, where the entire shortest path is eliminated, after which a new shortest path is calculated. This approach is more drastic, as it eliminates overlap but can result in an unrealistic choice set (e.g. large detours). Rieser-Schussler et al. (2013) adapted the link elimination method by applying a breadth-first search technique on link

elimination (BFS-LE), meaning that one starts eliminating links closest to the origin, repeats the shortest path search and moves stepwise towards the destination, before going one level deeper and eliminating two links at once (the one removed in the first level and again the first link of the new shortest route). They found lower error percentages compared to previous implementations of the link elimination method. Furthermore, this method appears to be computationally efficient and is suitable for high density networks (Rieser-Schussler et al., 2013). It has been applied in different contexts, e.g. cars (Rieser-Schussler et al., 2013; Prato et al., 2012; Dhakar and Srinivasan, 2014; Montini et al., 2017), bicycles (Menghini et al., 2010; Halldórsdóttir et al., 2014; Montini et al., 2017), heavy goods vehicles (Hess et al., 2015), and public transport (Montini et al., 2017).

Ben-Akiva et al. (1984) introduced the *labelling approach* which searches for the most optimal alternative given a certain label (e.g. distance, time, number of turns etc.). Prato and Bekhor (2007) applied this method to an urban network for cars in which they minimise for distance, free-flow time, travel time and travel delay. They report a false negative rate of 60%. Bekhor et al. (2006) specified and examined 16 different labels in their study. They found that each individual label generates only between 8% and 34% of the observed alternatives, while combined they can reproduce 72% of the observed routes. This method has been applied in the bicycle route choice context by Chen et al. (2017), Li et al. (2017), and Skov-Petersen et al. (2018). Unfortunately, none of them evaluate the performance of this method. Dial (2000) proposed a generalised approach of the labelling method for generating efficient paths. This method minimises a linear combination of labels. Broach et al. (2010) extended the labelling approach by generating multiple optima for one label by varying the label cost function parameter. They applied the method to bicycle traffic and identified eleven different labels, among others the distance of upslope travel and the number of turns. Their method generated more observed alternatives than the labelling method, however, the computation time also increased manifold. They also applied this method in a later study (Broach et al., 2012).

Table 1 provides an overview of the performance of the discussed methods in terms of producing false negatives in comparison to the number of alternatives generated. Note that the studies mentioned before are only included in the table if these numbers were provided. In general, when generating more alternatives, the false negative error percentage should decrease (where the false positive error potentially increases). Next to that, computation time of the methods is compared.

Because the studies use different datasets, it is hard to objectively compare the results. Most studies have resulted with a relatively high number of alternatives in the choice set, indicating that both relevant and irrelevant alternatives are included in the choice set. The different studies have also addressed different modes; the false negative error

percentage is higher for the non-motorised modes compared to the motorised modes for each algorithm. This is most likely due to the higher complexity of the network for bicycles compared to cars and trucks.

From the link elimination methods, the BFS-LE approach introduced by Rieser-Schussler et al. (2013) is most promising and therefore selected as a reference method in this paper. Several other studies have applied this method and found decent computation times and a lower share of false negatives compared to the original link elimination approach. Furthermore, the original labelling approach introduced by Ben-Akiva et al. (1984) is included as a reference method, because it outperforms the later proposed method of Broach et al. (2010) in terms of computation time and performs only slightly worse in terms of producing false negative errors.

3. Introducing the data-driven path identification approach (DDPI)

Due to the increased availability of (passively) collected revealed preference data and the issues associated with current choice set generation algorithms, the opportunity arises to identify choice sets using a data-driven approach. In this section, the data-driven approach coined Data-driven Path Identification (DDPI) which is introduced in Ton et al. (2017), is elaborated upon.

The DDPI approach is based on revealed preference data, like Wi-Fi, Bluetooth or GPS data of a large sample of individuals collected over a longer period. The idea behind this approach is to combine all observed routes from one origin to one destination into a single choice set at the origin-destination level (OD Pair). Using this method, the false negative error (not reproducing the observed route) is resolved. Furthermore, all routes that are included have been chosen by an individual, this means that these routes are optimised to a certain extent. Consequently, it is likely that these routes have been considered by an individual and from this set one route has been chosen. Therefore, the proposed method is expected to be less prone to false positive errors (including routes that are not considered) than choice set generation algorithms. However, because the choice set contains only chosen routes, it is possible that other routes that were considered but not chosen, are excluded, consequently potentially resulting in a choice set that is too small. A counterargument is that if data is collected over a long enough period of time, all relevant and considered routes are part of the data-driven choice set, therefore reducing this issue.

Several requirements need to be met for the DDPI approach to be applicable. First, the data should be collected over a sufficiently long period of time to allow multiple observations per OD pair. Second, it is necessary to have at least two routes per OD pair to facilitate the estimation of a route choice model. However, because of issues with

Table 1
Performance of applied deterministic choice set generation algorithms.

Deterministic category	Method	Study	Data	Mode	False negative error	Max no. alternatives	Comp. time
Link elimination method	Link elimination	Bekhor et al. (2006)	Boston, USA	Car	40%	Unknown	Medium
		Prato and Bekhor (2007)	Turin, Italy	Car	42%	10	–
		Ghanayim and Bekhor (2018)	Tel-Aviv, Israel	Bicycle	40%	10	–
	Breath-first search on link elimination	Rieser-Schussler et al. (2013)	Zurich, Switzerland	Car	37%	20	–
		Hess et al. (2015)	United Kingdom	Trucks	27%	100	–
Labelling approach	Labelling	Halldórsdóttir et al. (2014)	Copenhagen, Denmark	Bicycle	26%	15	–
		Bekhor et al. (2006)	Boston, USA	Car	34%	20	Medium
	Calibrated labelling	Prato and Bekhor (2007)	Turin, Italy	Car	28%	16	Low
		Broach et al. (2010)	Portland, USA	Bicycle	61%	3	–
		Broach et al. (2010)	Portland, USA	Bicycle	80%	9	Low
		Broach et al. (2010)	Portland, USA	Bicycle	78%	20	Medium

endogeneity, it is preferable to have more than two routes per OD pair. Because the observed routes are optimised to a certain extent by the individual, the variability of the routes is low. By including more routes, the variability of the routes increases and the issue with endogeneity will be less severe. If this is not accounted for, the estimated models will be biased. If there is an OD pair which does not meet these requirements, it needs to either be deleted or aggregated by applying a spatial clustering technique. Clustering of OD pairs can be useful in case of, for example, two neighbours heading for the same destination. It can prevent loss of data, but should be carefully addressed, because the OD pairs still need to be comparable. The impact of these requirements can be small, if they are taken into account in the design phase of the data collection.

The requirements of the method also point to the limitations of the DDPI approach. It imposes additional requirements to the data collection, because if the data is already collected and requirements are not adequately met, a (severe) loss of data and an endogeneity issue can be the result. The endogeneity is the result of including all chosen alternatives in the choice set. The issue is larger if the alternatives are more similar and there are only few. In that case, the method should not be used, as it imposes a bias in the choice model. Similar to other methods, another limitation is found in the generalisability of the results: data is collected for a certain group of people and for a certain region. Consequently, it is per definition uncertain whether the results (modelling or choice set) can be transferred to other groups of people or other regions, similarly to the generalisability issues associated with other methods.

The data collection duration (for example a week versus several months) suitable for the application of the DDPI method depends on the local network and demand properties. It is important to ensure a long enough period so that the routes observed exhibit a sufficient degree of variation.

4. Methodology for evaluating choice set specification methods

The methodology for assessing the usefulness of the DDPI approach and comparing the different choice set generation methods is presented in this section. Section 4.1 details the methodology for comparing the generated choice sets to the observed data. Furthermore, Section 4.2 discusses the evaluation methodology for estimation and validation of the route choice model. Section 4.3 then provides a synthesis of the evaluation methodology.

4.1. Evaluating the specified choice sets

The specifications of the algorithms to which the DDPI approach is compared are discussed (Section 4.1.1), and the methodology for comparing the generated choice sets to the observed routes is provided (Section 4.1.2).

4.1.1. Selected choice set generation algorithms

The BFS-LE and labelling approach have been selected for comparison. Both algorithms use calculations of the shortest path. The algorithm used to calculate the shortest path is Dijkstra (1959). The input for Dijkstra's algorithm is a (distance)matrix, which can grow very large, especially when considering bicycles. To decrease the computation time and increase the spatial diversity among routes, a topologically equivalent network reduction is adopted in this study. This means that nodes that connect only two other nodes (i.e. a node degree of two) are removed from the network and the two links are merged into one. Consequently, the network (or matrix) consists of fewer nodes and the resulting shortest path consist of fewer links, thus significantly reducing the computation time.

These choice set generation algorithms can utilise several input variables. Mostly, the algorithms are applied based on travel distance. In the bicycle route choice context, several studies have considered

alternative variables. Broach et al. (2012) used an approach that optimised criteria like percentage of designated cycle paths, subject to distance constraints. Haldórsdóttir et al. (2014) search for the shortest route in terms of road type, bicycle paths, and land use. Finally, Chen et al. (2017) used a combination of speed limits, distance, and bicycle facilities to generate routes. Due to limited data availability for the inner-city of Amsterdam (see Section 5.4), we rely largely on travel distance in the choice set generation algorithms. The two algorithms are specified below.

4.1.1.1. Breadth-first search on link elimination (BFS-LE). The BFS-LE algorithm, introduced by Rieser-Schussler et al. (2013), was developed specifically for high-density networks, e.g. urban networks. The idea behind the approach is to calculate the shortest path (in this paper we adopt calculation based on distance, like in the original study) between an origin and destination, add this path to the choice set and then remove the links of this shortest path step-by-step, starting from the origin node. In each step a new shortest path is calculated and added to the choice set, given that it is unique. A tree structure is adopted to keep track of the removed links and the resulting adapted networks, this means that in the second tree level two links are eliminated (the link that was deleted from the shortest path and the link from the new shortest path).

Maximum computation time, tree-depth, and choice set size can be used as termination measures for the BFS-LE algorithm. In this study, we applied a mix of these measures. Because an individual is not able to remember or consider many routes, we have set the maximum to 20 routes. This seems adequate given the findings from Hoogendoorn-Lanser (2005) indicating that different individuals only know seven alternatives. Since we only search for 20 unique routes, we have applied a tree-depth of one, with a random draw of 20 routes in case more routes are generated. The second level sometimes generated over 1000 routes, and induced an exponential growth in computation time. The unique routes found in tree-depth one, are added to the choice set resulting from tree-depth zero.

4.1.1.2. Labelling approach. The labelling approach proposed by Ben-Akiva et al. (1984) searches for the most optimal route based on different network-related search criteria, e.g. distance, travel time or number of left turns. This method facilitates the composition of a very diverse choice set, given the available data. The number of labels encoded, sets the maximum value of the number of alternatives included in the choice set. The input-matrix required for the Dijkstra's algorithm is adapted for each of the labels considered. In this study, we have identified three labels, resulting in a maximum choice set size of three.

The three labels are the shortest path based on distance, the highest percentage of separate cycle paths and the least amount of intersections on the route. The matrix that serves as input for the Dijkstra algorithm is node-based. Consequently, each link is presented as a connection between two nodes. The algorithm then searches in this matrix to identify the shortest path. Regarding separate cycle paths, each link that has a separate cycle path or a protected lane, has a weight of zero, all other links have a weight of one. The ideal route found by the algorithm consists of 100% separate cycle path, thus maximising the amount of cycle path. Furthermore, regarding intersections, each link is assigned with the same weight, therefore the algorithm searches for the shortest path in terms of the number of links traversed. In the absence of more detailed information, all intersections (with a node degree of at least three) are treated equally.

4.1.2. Evaluation methodology for specified choice sets

The DDPI approach directly uses the observed routes to identify the choice set, consequently there is no difference between the DDPI approach (after data preparation) and the observed routes, and it is not evaluated separately. The performance of the algorithms is evaluated

by comparing the generated choice sets to the observed routes. First, a qualitative analysis is performed, in which two OD pairs are selected and visually compared. This gives an indication on the spatial distribution of the generated routes and potential differences and similarities between the choice sets. Second, a quantitative analysis provides descriptive statistics of three network related variables, based on previous work on bicycle route choice¹: percentage of separate cycle paths, distance and number of intersections per kilometre. This analysis shows the general characteristics of the different choice sets compared to the observed routes.

Furthermore, the heterogeneity of the generated choice sets is investigated, quantitatively showing how spatially different the generated routes are. This is done by calculating the path size (PS) factor for each route in the choice set, which is an indicator for overlap between routes (Ben-Akiva and Bierlaire, 1999).

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

where PS_{in} is the path size factor, Γ_i is the set of links in route i , l_a is the length (distance) 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. This means that the PS factor depends largely on the size and composition of the choice set (i.e. including many irrelevant routes affects this factor). The path size factor ranges between zero and one, where one indicates an independent route and zero indicates complete overlap with other routes in the choice set.

The main objective of choice set generation algorithms is to reproduce all observed routes, i.e. resulting with zero false negative errors. To test to what extent the algorithm can reproduce the observed routes, the following formula for the reproduction rate is adapted from Prato and Bekhor (2007):

$$RR_r = \sum_{n=1}^N I(O_{nr} > \delta) \tag{2}$$

where RR_r is the reproduction rate for algorithm r . $I(\bullet)$ is the reproduction function, which is equal to one if the argument is true and zero otherwise; O_{nr} is the overlap rate for algorithm r for observation n , and δ is the overlap threshold, which can be set from no overlap (0%) to full overlap (100%). O_{nr} is calculated in the following way:

$$O_{nr} = \frac{L_{nr}}{L_n} \tag{3}$$

where L_{nr} is the common distance between the generated route and the observed route for algorithm r and observation n . L_n is the total distance of the observed route for observation n . The reproduction rate (Eq. (2)) yields how many observed routes are generated when allowing for a certain overlap threshold.

In addition to the reproduction rate, the behavioural consistency of both methods is assessed. The consistency index compares the algorithms to the ideal algorithm that would reproduce all the observed routes, and calculates how well the algorithms perform. The formula used to calculate this index is the following (Prato and Bekhor, 2007):

$$CI_r = \frac{\sum_{n=1}^N O_{nr,max}}{N} \tag{4}$$

where CI_r is the consistency index for algorithm r ; $O_{nr,max}$ is the maximum overlap percentage obtained for observation n using algorithm r , i.e. the best matching generated route to the observed route n ; N is the total number of observations in the sample.

4.2. Evaluating the model estimation and validation

The specifications of the route choice model that is estimated; the Path-Size Logit (PSL) model is discussed (Section 4.2.1), and the methodology to evaluate the model estimation and validation is

provided (Section 4.2.2).

4.2.1. Specification of the route choice model

A wide variety of discrete choice models, varying in computational complexity, have been developed that are suitable for route choice. Examples are Cross-Nested Logit (CNL), Paired Combinatorial Logit, C-Logit and PSL. Bliemer and Bovy (2008), Prato and Bekhor (2007) and Bekhor et al. (2006) have compared these models for route choice. They concluded that the CNL and PSL model perform best. Since the CNL model is more complex, requires specialised code and has a higher computation time, we apply the PSL model in this evaluation (Bekhor et al., 2006).

To account for potential correlation among path alternatives (e.g. route overlapping), the PSL model introduces a similarity measure in the utility function. In this study, the path size (PS) factor proposed by Ben-Akiva and Bierlaire (1999) is adopted (Eq. (1)). The probability of choosing alternative i given choice set C_n is specified as follows (Ben-Akiva and Bierlaire, 1999):

$$P(i|C_n) = \frac{e^{(\beta_d * dist_{in} + \beta_i * \frac{n_{int}}{km}_{in} + \beta_{cp} * \%sep.cyclepath_{in} + \beta_{PS} * \ln PS_{in})}}{\sum_{j \in C_n} e^{(\beta_d * dist_{jn} + \beta_j * \frac{n_{int}}{km}_{jn} + \beta_{cp} * \%sep.cyclepath_{jn} + \beta_{PS} * \ln PS_{jn})}} \tag{5}$$

where based on previous work, three explanatory variables are included per alternative i and observation n : percentage of separate cycle paths ($\%sep.cyclepath_{in}$), distance ($dist_{in}$) and number of intersections per kilometre ($\frac{n_{int}}{km}_{in}$). PS is again the path size factor calculated in Eq. (1), it ranges between zero and one, where one means no overlap and zero implies complete overlap between routes. The models are estimated using the Python Biogeme package (Bierlaire, 2016).

4.2.2. Evaluation methodology for model estimation and validation

Three route choice models are estimated and validated, using the two generated choice sets and the choice set that is identified using the DDPI approach. Because for each OD pair routes are generated using the two generation algorithms and multiple routes are observed per OD pair, a union of the observed and generated routes is created for the Labelling and BFS-LE choice sets. Fig. 1 shows this merging of observed (1.a) and generated (1.b and 1.c) routes for the BFS-LE and labelling method. All observed and generated routes for one method per OD pair are merged into one choice set (1.d and 1.e), corrected for the reproduced observed routes.

The model estimation and validation are done by splitting the data sample into two parts (80/20). The models are estimated using 80% of the observed OD pairs and validated using the remaining 20%. This way, the predictive power of the models can be tested and potential errors can be detected. The model estimation and validation is done for five random draws to test stability of the models. Note that the sampling is done on the OD pairs that result from the DDPI approach, so that the variability in the OD pair remains for the model estimation and the issue with endogeneity is less severe.

Since the models are estimated using different choice sets, a standard comparison based on log-likelihood ratio or model fit (adj. rho-square) cannot be done. The initial log-likelihood is different due to the different sizes of the choice sets. Therefore, the comparison is based on the point elasticities of the model's explanatory variables, calculated using the following formula:

$$E_{x_i}^{P_n(i)} = \frac{\partial P_n(i)}{\partial x_i} \frac{x_i}{P_n(i)} \tag{6}$$

where $P_n(i)$ is the probability that observation n chooses alternative i and x_i is an attribute (defined in Eq. (5) for alternative i). The mean elasticity is then obtained by probability weighting the elasticities for every individual n , where the probability weights relate to the probability of choosing an alternative in the choice set. In the validation phase, the probability for each alternative to be chosen is calculated for the

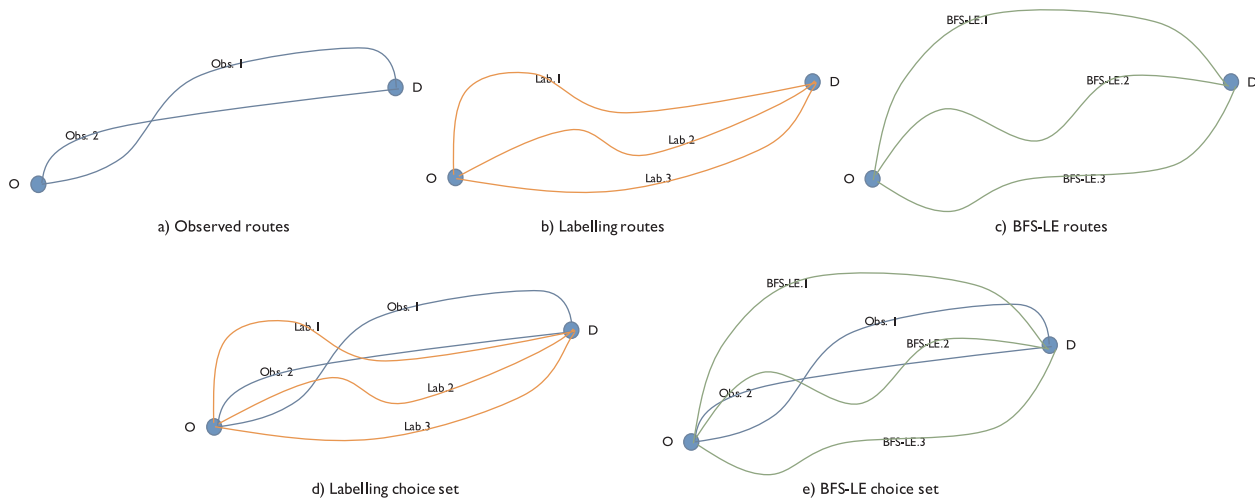


Fig. 1. Formation of choice sets for Labelling and BFS-LE algorithms.

remaining 20% OD pairs. To make a fair comparison between all models, a union of all generated and observed alternatives is generated for each OD pair (in essence a union between Fig. 1.d and e, corrected for unique routes). The union choice sets for each OD pair are used to assess the predictive power of all models, using three measures. First, the number of times the model assigns the highest utility to the chosen alternative for all observations. This gives an indication about the extent to which the model is able to predict the correct choice. Second, the RMSE value is calculated, which gives an indication of the error that arises between observed probabilities (based on observed routes) and modeled probabilities per OD pair. This value is calculated using the following formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N_{OD}} (\hat{P}_i - P_i)^2}{N_{OD}}} \quad (7)$$

where \hat{P}_i is the vector of probabilities that is predicted by the model for OD pair i and P_i is the vector of observed probabilities of OD pair i . Finally, the log-likelihood is calculated on the out-of-sample data. As a union of all generated and observed routes is used to define the choice sets, the input is the same for all models. Therefore, a comparison based on log-likelihood is possible. It is calculated using the following formula:

$$\text{Log-Likelihood} = \sum_{n=1}^N \left(\sum_{i \in C_n} y_{in} \ln P(i|C_n) \right) \quad (8)$$

where y_{in} is one if n chooses alternative i in choice set C_n , and zero otherwise, and $P(i|C_n)$ is the probability of choosing alternative i in choice set C_n .

4.3. Synthesis of the evaluation methodology

A concise overview of all the methods introduced for analysis and evaluation of the choice sets, model estimation and model validation is presented in Fig. 2.

5. Data description and preparation

The dataset that is used to assess the usefulness of the DDPI approach and benchmark the approach against the BFS-LE and labelling algorithms is a bicycle GPS dataset. This dataset was collected during a nationwide initiative in the Netherlands called the ‘Bicycle Counting Week’, which took place on 14–20 September 2015. A total of 38,000 cyclists participated using a smartphone application that tracked their cycling movements, recording more than 370,000 trips nationwide.

Additionally, a survey was distributed among the participants that used the smartphone application. Section 5.1 describes the dataset that is used in this study. Furthermore, Section 5.2 describes the map matching procedure for matching the GPS trajectory data to the network. Section 5.3 provides insights on the clustering procedure applied to the origins and destinations of all the trips made in the dataset. Finally, Section 5.4 addresses the preparations needed related to the data and network for the choice set generation methods.

5.1. GPS dataset from the inner-city of Amsterdam

In this evaluation, the focus lies on the inner-city of Amsterdam, which is a densely-built area with well-developed cycling infrastructure. The dataset was used in previous work, where the DDPI approach was applied to estimate a bicycle route choice model for this specific area (Ton et al., 2017). Fig. 3 shows the network of the inner-city of Amsterdam. In total, 3045 trips were recorded in the inner-city of Amsterdam. Not all trips could be used in this case study, as some trips were too short to be included and some could not be matched to the topologically equivalent reduced network, resulting in a total of 2819 trips. The respondents sample consists of equal shares of male and female participants. Most respondents are 31–65 years of age (80%). Most trips are made for commuting purposes (77%). Furthermore, most respondents cycle between 25 and 100 km a week (72%) (Fiets Telweek, 2015). The individual characteristics are only available on an aggregate level, due to privacy regulations, therefore it is impossible to link the GPS trajectories to individual travellers. This has two major consequences: (1) individual characteristics cannot be used in the model estimation, whereas several cycling route choice studies have identified the relevance of such variables (Hood et al., 2011; Broach et al., 2012) and (2) it is impossible to identify which trips have been made by which individuals, thus we need to treat each trip as if it was made by a unique individual and cannot therefore test for panel effects in the model estimation.

5.2. Map matching the GPS trajectory data

The map matching procedure was conducted by the organizers of the Bicycle Counting Week (van de Coevering et al., 2014). The following is an account of the procedure that has been performed. GPS data points in a trajectory have a maximum accuracy of around 5 m with respect to the infrastructure. However, outliers are observed in dense urban areas or high building areas, reducing the accuracy by up to 50 m. In urban areas, this means that the next street can be mistakenly identified. To reduce the impact of these outliers on the

	Analysing alternatives	Model estimation	Model validation
Analysis method	Identify choice set for: <ul style="list-style-type: none"> - DDPI approach - BFS-LE approach - Labelling approach 	Estimate PSL models: <ul style="list-style-type: none"> - Distance (km) - % separate cycle paths - Intersections per km 	Validate PSL models: <ul style="list-style-type: none"> - 20% out-of-sample - Union of alternatives from all choice sets
Evaluation method	<ul style="list-style-type: none"> • Qualitative analysis • Quantitative analysis • Reproduction of observed routes 	<ul style="list-style-type: none"> • Mean point elasticities 	<ul style="list-style-type: none"> • Percentage of trips correctly predicted • RMSE per OD pair • Log-Likelihood

Fig. 2. Analysis and evaluation methods for analysing the alternatives in the choice set, model estimation and model validation.

observed trajectories, van de Coevering et al. (2014) have calculated the speed between each two consecutive GPS data points and compared it to the actual GPS speed, which was determined by means of Doppler techniques. If a large discrepancy between the actual speed and the calculated speed has been identified, the outlier and two preceding and following GPS data points from the dataset were removed.

The corrected GPS trajectories can afterwards be matched to the network. The entire network is split up in nodes, after which links were divided into smaller segments to determine local differences in network speeds, which helps in determining whether a cyclist was able to cycle on a link. The map matching algorithm they applied generates all possible combinations of origin and destination points in the network, which is necessary because of the inaccuracy of the GPS data points. Routes were then plotted between all the identified combinations of origins and destinations. The goal is to minimise the distance between the GPS trajectory and the network route, which results in routes that best resemble the GPS trajectories. If a match could not be found, this may stem from missing links. In those cases, the route is partitioned and the same procedure is repeated for the sub-routes. For a more detailed

description of the map matching procedure, the reader is referred to van de Coevering et al. (2014).

5.3. Clustering of the origins and destinations of the GPS trajectories

We applied a clustering method on the observed origins and destinations, to ensure that multiple trips and routes are observed for each OD pair. A k-means clustering approach was applied which minimises the intra-cluster distance and maximises the inter-cluster distance. Different numbers of clusters were tested (150, 200, 250, and 300) to find a good balance between having enough trips per OD pair (high number of clusters) and ability to compare routes in an OD pair (low number of clusters). Finally, a total of 200 clusters provided the best results. For a more detailed description of the clustering, the reader is referred to Ton et al. (2017).

5.4. Data and network preparations for the choice set generation methods

As mentioned in Section 5.1, we cannot identify which individual

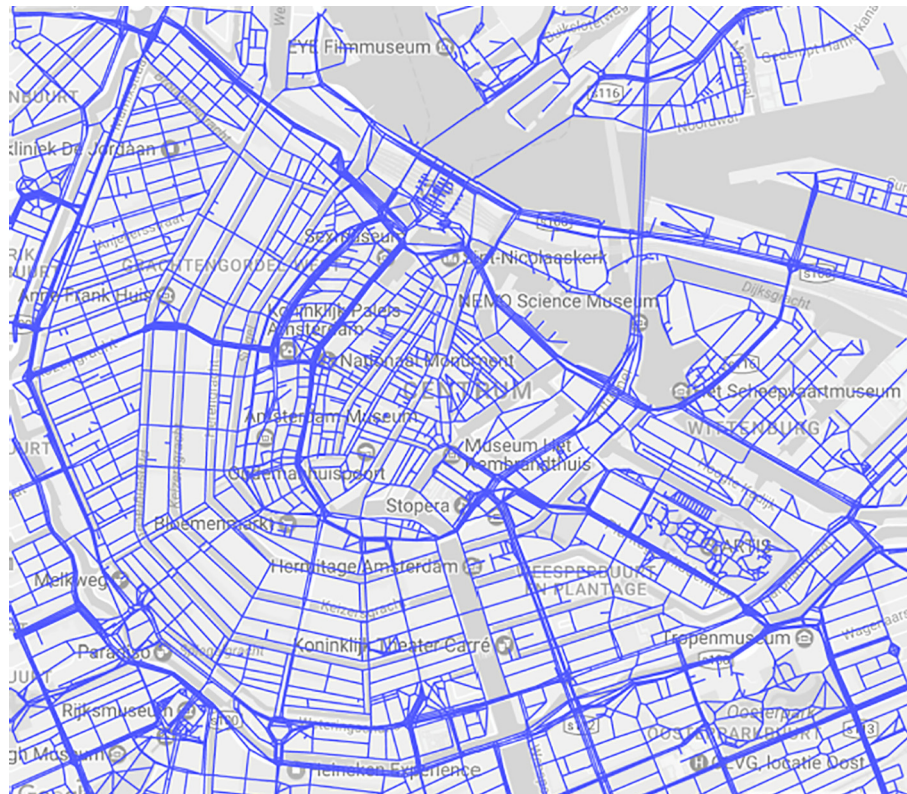


Fig. 3. Road network of the inner-city of Amsterdam.

made which trip, consequently we have to treat every trip-maker as a unique individual. Ideally, the DDPI method would have been applied per individual and OD pair. Given the mentioned restriction in the data, it is not possible to identify individual choice sets. Therefore, this study uses all trips that are observed per OD pair and combines them to form choice sets. Furthermore, data is collected over the course of one week. Consequently, we are not able to test how sensitive this dataset is with respect to the duration of data collection versus the diversity of observed routes. Data would need to be collected over a longer period of time (multiple weeks) in order to test the sensitivity of model performance to the data collection duration.

The choice set generation algorithms use the network of Amsterdam (Fig. 3) to generate the routes, therefore the network is extracted from OpenStreetMap (OSM). In the road network of OSM the two bicycle/pedestrian ferries crossing the river IJ are not included, therefore two bidirectional links are added to the network with origins and destinations at the ferry landings. Furthermore, the inner-city of Amsterdam contains many one-way streets. Tests with the choice set generation algorithms show that the generated routes contain many detours and illogical routes if these links are not considered to be bi-directional. Therefore, we have converted the entire network into a bi-directional graph. Furthermore, in the OSM network many links that are mainly used by non-motorised modes are not incorporated in the network. Tests with the choice set algorithms show that this affects many OD pairs, therefore these have been added to the network when possible. Still, many links that are used by cyclists, are not included in the network. These links could for example be shortcuts or pedestrian areas, where other modes are not allowed, both of which are not included in the network. Consequently, network-related issues could arise when generating routes. A total of 19,375 nodes is identified in the network. Due to applying topologically equivalent network reduction (as mentioned in Section 4.1.1), the number of nodes decreased to 7628 nodes (-61%) with a total of 25,135 links.

The insertion of local knowledge regarding the network, to make sure that the majority of the illogical routes will not be generated using the choice set generation methods, underscores a major advantage of the DDPI method. This method relies only on the data that is collected from observed trips and thus does not require any network-information. Consequently, local knowledge is not required for using this method for analysing alternatives, model estimation, and model prediction. Furthermore, the DDPI method can be used as a reference set in adjusting the specification of currently adopted labelling approaches. Next to that, the algorithms use the information from the network or any other data source that is available, which is especially relevant for the labelling algorithm. As mentioned before, only three labels can be identified for this study, due to the limited data availability on the network.

6. Generated choice set evaluation

The choices sets that are generated using the BFS-LE and labelling approach are compared to the observed routes according to the methodology described in Section 4.1. The qualitative analysis for two selected OD pairs is covered in Section 6.1. Section 6.2 details the quantitative analysis on the complete choice sets. Section 6.3 provides the results of the analysis on reproduction rate and behavioural consistency of the choice set algorithms. Finally, Section 6.4 concludes the choice set evaluation.

6.1. Qualitative analysis of the choice sets

The observed routes of the two selected OD pairs are plotted on the map in Fig. 4. Cyclists in the first OD pair (upper OD) travel from the west of the inner-city of Amsterdam to the north side of the central train station and cyclists in the second OD pair (lower OD) travel from the centre (Waterlooplein) to the Vondelpark in the south-west of the inner-

city.

The routes generated for the first OD pair using the BFS-LE and labelling approach are visualised in Fig. 5, together with the observed routes. The observed routes (5.1) show a diverse set of routes. The north of the station can only be reached by one of the tunnels underneath the tracks, furthermore the cyclists face the canals that form a ring around the city centre, resulting here in roughly four main routes. The BFS-LE approach (5.2) provides a set of shortest routes, showing less diversity in this case. This approach only shows spatial diversity in the city centre. It avoids following the canals, which is different from the observed behaviour. This indicates that the cyclists are not necessarily aiming for the shortest route. The labelling approach (5.3) shows a more diverse choice set, that mimics the observed behaviour better. It does not provide exact matches, but provides routes that are more spatially different and makes use of the direction of the canals. This first comparison indicates that the labelling approach mimics the observed behaviour better in terms of spatiality and behaviour.

The generated choice sets for the second OD pair are visualised in Fig. 6. The observed routes (6.1) again show a spatially diverse image. For most routes, the number of turns is minimised. The cyclists start northwards, then follow one of the ring roads and continue north, with different turning points. The BFS-LE approach (6.2) shows similar behaviour for the shortest route, however this route turns later than any of the observed routes. The northbound route that is generated is very different from the observed routes. Again, this approach generates a less spatially diverse choice set, that is unable to find all the observed routes. The labelling approach (6.3) is again more spatially diverse than the BFS-LE approach, but shows different routes than to the observed routes. Two of the three generated routes are comparable to the observed routes, in terms of turning. The third route turns often, which is very unlike the observed behaviour. The comparison of the second OD pair shows again that the labelling approach mimics the observed routes better than the BFS-LE approach, however the differences between the choice sets are still large. This qualitative analysis indicates that behaviour of cyclists is not captured based on one objective/label.

6.2. Quantitative analysis of the choice sets

In this section, the choice sets that are generated by the BFS-LE and labelling approach are compared to the observed routes based on a quantitative analysis. The descriptive statistics are calculated for distance, percentage of separate cycle path and the number intersections per kilometre. Furthermore, the path size factor (Eq.(1)) is calculated, which is an indicator for heterogeneity of the choice set. Table 2 shows the results of the quantitative analysis.

The observed routes show that the mean distance travelled is 1.9 km, whereas the entire area included in the research covers about 6 km. This indicates that the average cyclist does not cross the entire inner-city. Furthermore, the percentage of separate cycle paths encountered on the routes and the amount of intersections per kilometre (all types of intersections) are rather low, the latter was expected from the qualitative analysis. Finally, the path size factor is on average 0.67, which indicates a relatively heterogeneous set of routes, matching the results from the qualitative analysis. The routes chosen by all cyclists are spatially diverse and have a low degree of overlap.

The BFS-LE approach optimises for distance, which is reflected in the lower mean distance and standard deviation. However, the difference with respect to observed routes is negligible, which seems to imply that the cyclists prefer shorter routes. As mentioned before, several of the links, found in observed routes, are not included in the network. Inspections of the OD pairs crossing the city centre, showed that 25% of the trips cross these areas even though the network does not include these, indicating that the true shortest path cannot be found by the algorithms. It shows that the true mean distance might be lower than shown in Table 2, indicating that the preference for the shorter routes might be less straightforward than appears now. The BFS-LE approach

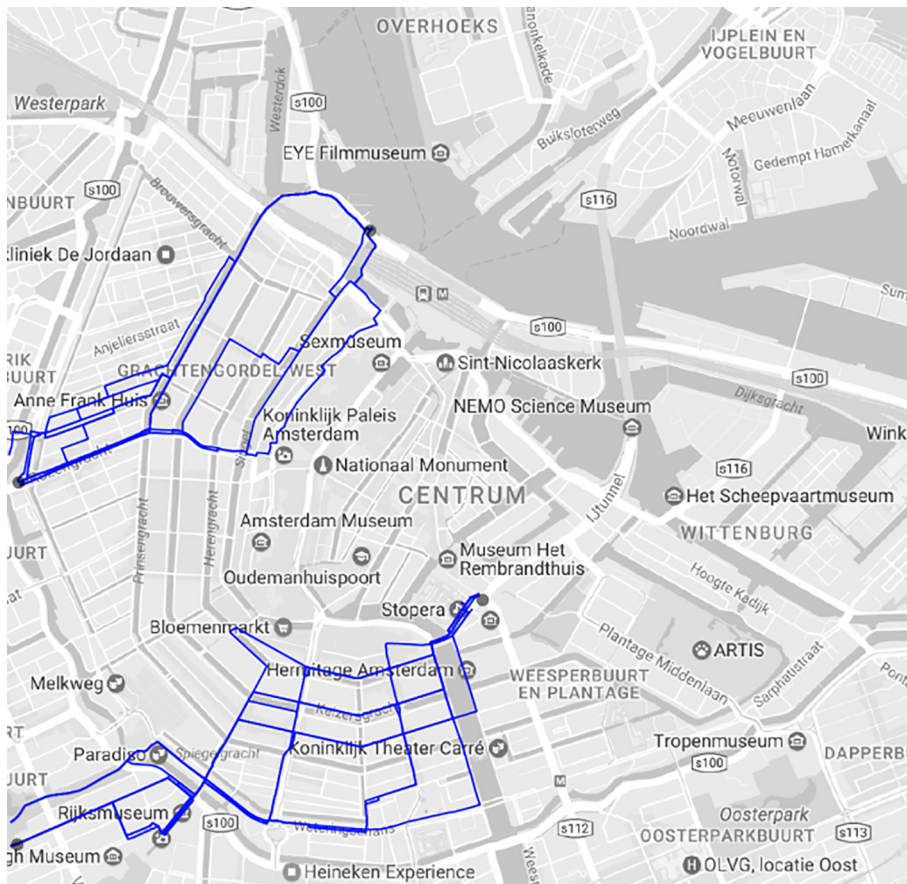


Fig. 4. Observed routes from two selected OD pairs, plotted on the map of Amsterdam.



Fig. 5. Routes generated for a given OD pair from the West of Amsterdam to the central train station, for (1) observed routes, (2) BFS-LE approach and (3) Labelling approach.

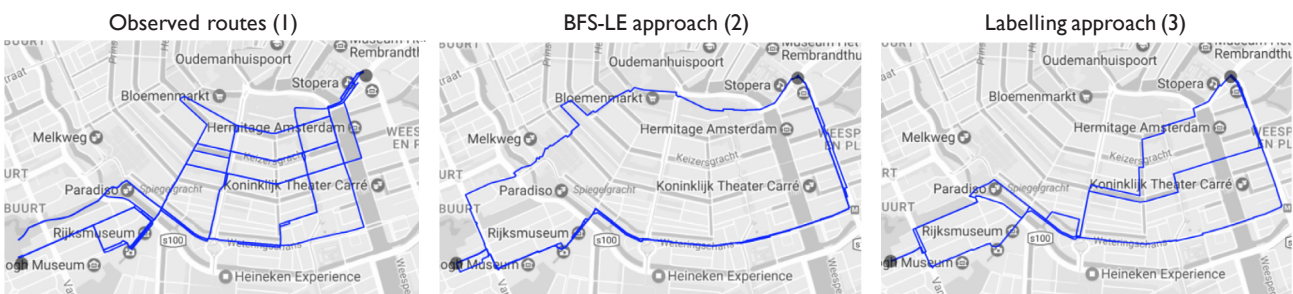


Fig. 6. Routes generated for a given OD pair from Waterloooplein to Vondelpark, for (1) observed routes, (2) BFS-LE approach and (3) Labelling approach.

Table 2

Descriptive statistics of the explanatory variables and heterogeneity indicator for each choice set identification approach.

Variable	Observed routes (N = 2819)			BFS-LE approach (N = 12,361)			Labelling approach (N = 2034)		
	Mean	Median	St.Dev	Mean	Median	St.Dev	Mean	Median	St.Dev
Distance (km)	1.93	1.85	1.01	1.92	1.85	0.78	2.82	2.47	1.66
Separate cycle path %	37.9%	34.7%	26.4%	8.3%	6.6%	8.2%	19.4%	9.7%	22.8%
Intersections per km	14.8	14.5	5.0	32.2	32.1	6.9	19.9	15.9	10.5
Path Size factor	0.671	0.704	0.232	0.135	0.090	0.126	0.833	0.864	0.136

also shows a low percentage of separate cycle paths and a high amount of intersections per kilometre compared to the observed routes. Most likely because the algorithm does not optimise for these variables. Due to the nature of the algorithm, it finds a low variety of routes, leading to a relatively homogeneous set of routes, reflected in the qualitative analysis.

The labelling approach generates a route that optimises for each variable in the descriptive statistics, therefore the standard deviations are large. The mean distance is larger than both other choice sets, whereas the percentage of separate cycle path and number of intersections per kilometre are in between the observed routes and BFS-LE algorithm. Furthermore, due to the optimisation on different variables, the choice set is very heterogeneous and spatially divers (as was also found in the qualitative analysis).

6.3. Reproduction of observed routes

This section covers the reproduction rate and behavioural consistency of both the BFS-LE and labelling approach. The reproduction rate is calculated for different levels of overlap between generated and observed routes, varying from 70% to 100%. Table 3 shows the results of these analyses.

The false negative error for both methods is about 99%, implying that the overwhelming majority of observed routes are not included in the generated choice-sets. The labelling approach is slightly better at reproducing the observed trips and has a higher behavioural consistency compared to the BFS-LE approach. The qualitative analysis showed that the labelling approach could partially reproduce the observed routes, however the overlap between the observed and generated routes is lower than 70%. The BFS-LE approach performs even worse, as was also visible in the qualitative analysis. As mentioned before, network-related issues could impact the choice set generation. This dependency of choice set algorithms on the network shows one advantage of the DDPI method, as this method does not rely on network information.

6.4. Conclusions regarding the evaluated choice sets

The choice sets resulting from the BFS-LE and labelling approach differ largely from one another, and they differ largely from the observed routes. The labelling approach is better than the BFS-LE approach in terms of mimicking the observed routes, but shows very large false negative errors (not generating the observed alternative). The quality of the network representation (topology and available label

Table 3

Number and percentage of observed routes generated by each choice set generation approach for different threshold levels.

Algorithm	100% Overlap		90% Overlap		80% Overlap		70% Overlap		CI
	# trips	% trips	# trips	% trips	# trips	% trips	# trips	% trips	
BFS-LE approach	26	0.9%	53	1.9%	92	3.3%	175	6.2%	0.2701
Labelling approach	38	1.4%	65	2.3%	110	3.9%	183	6.5%	0.3024

Note: the total number of trips is 2819.

information) that serves as input for the choice set generation methods, which is poor in the bicycle-context, influences the routes that are generated, especially when generating routes based on individual network characteristics. In this case, the observed behaviour is not captured by these characteristics. The differences indicate that cyclists optimise based on more than one network-related objective. Ehrgott et al. (2012) proposed a method for bi-objective optimisation, as they found that cyclists do not optimise based on one objective, like car drivers might do with distance or travel time. Two other methods that might be able to overcome this issue are the link-based approach introduced by Fosgerau et al. (2013) and importance sampling approaches like the Metropolis-Hastings approach (Flötteröd and Bierlaire, 2013), as they approach the choice set generation from the universal choice set.

7. Evaluation of model estimation and validation

This section covers the evaluation of the model estimation (7.1) and validation (7.2). Three route choice models are estimated using the choice sets resulting from the labelling approach, BFS-LE approach and DDPI approach (as shown in Fig. 1). The evaluation takes place according to the methodology proposed in Section 4.2.

7.1. Route choice model estimation

The most elegant way of dealing with non-generated observed routes, would be to eliminate the entire OD pair. However, in this case it would mean that only very few OD pairs would remain (approximately 1% of the trips). Therefore, in practice the observed routes that have not been generated are added to the choice set (e.g. Broach et al. (2010)). Consequently, a union of routes is created based on network characteristics and observed behaviour (like depicted in Fig. 1). This method entails that information/observed behaviour is added to the choice set, which will increase the performance of these choice sets in model estimation and consequently introduces an issue with endogeneity (by including chosen alternatives). The comparison in the model estimation is therefore skewed, due to this poor performance in terms of reproducing observed alternatives.

Five models are estimated for each choice set, every time using a different random sample of 80% of the OD pairs, to investigate the stability of the models. Table 4 shows the estimation results for one of the model runs.

The signs of distance, separate cycle path percentage and intersections per kilometre are as expected and are the same for each model.

Table 4
Estimated PSL models using the identified choice sets from data, BFS-LE and labelling.

Variables	DDPI model		BFS-LE model		Labelling model	
	Coef.	t-test	Coef.	t-test	Coef/	t-test
Distance (km)	-0.225	1.72 [*]	-0.341	2.84 ^{**}	-1.84	21.88 ^{**}
% separate cycle path	0.153	1.00	1.34	9.47 ^{**}	1.53	11.45 ^{**}
Intersections/km	-0.018	2.11 ^{**}	-0.159	23.90 ^{**}	-0.118	21.61 ^{**}
(Ln) Path Size	-0.380	3.94 ^{**}	1.03	17.11 ^{**}	0.291	3.77 ^{**}
N	2249		2249		2249	
Null log likelihood	-3059.718		-6,921.409		-4,419.422	
Final log likelihood	-3044.254		-3539.881		-3627.528	
Likelihood ratio test	30.928		6763.057		1538.788	
Adj. rho square	0.004		0.488		0.178	

* Significant on 90% confidence interval.
** Significant on 95% confidence interval.

However, the parameter and *t*-test values are different. The DDPI model has lower *t*-test values compared to the other models, which is due to the endogeneity issue that plays a role in the DDPI choice set. It has the tendency to make attributes less significant. Furthermore, the sign of the path size factor is different for the DDPI model. In this case a route that has more overlap with other routes receives a higher utility. In the context of public transport, Lam and Xie (2002) also found a negative parameter. They argue that overlapping routes can reduce uncertainty by allowing more en-route rerouting possibilities and hence contribute to the robustness of the route taken, which could also hold for the bicycle route choice situation. In case of the BFS-LE and labelling model, adding the observed routes results with a positive PS factor. The generated alternatives overlap with each other, but often the observed alternatives are very different, resulting in a higher utility for the non-overlapping routes. Consequently, the interpretation of the negative PS sign is different from the positive PS sign, showing a difference between observed and generated choice sets.

To compare these models, the average point elasticities for all explanatory variables are calculated (Table 5). The elasticity provides information on the impact of marginal changes in each of these variables on the probability of being chosen.

The interpretation of the elasticities is such that 1% increase in distance results in a decrease in the probability of being chosen of 0.29% for the DDPI model, whereas the BFS-LE model shows a 0.44% decrease and the labelling model shows a decrease of 2.58%. The relative difference between the impact of the BFS-LE model and DDPI model is 52%, but is around 790% with the labelling model. In the labelling model, the impact of marginal changes to all variables, is much higher compared to the other models. The routes generated by the labelling algorithm are very divers and optimised for different criteria, which indicates that increasing the variability in the alternatives (labelling routes plus observed route), induces a higher elasticity.

7.2. Route choice model validation

The model validation provides insight into the predictive power of the models. The 20% remaining OD pairs are used to validate the models. For the validation, the alternatives of all three choice sets are

Table 5
Mean point elasticities for each explanatory variable for all models.

Variable	DDPI model Elasticity	BFS-LE model Elasticity	Labelling model Elasticity
Distance	-0.289	-0.440	-2.577
% separate cycle path	0.042	0.350	0.426
Intersections/km	-0.188	-1.702	-1.316

Table 6
Average validation measures for all 5 estimated models per choice set.

	Correct choice predicted	RMSE OD pair	Log-likelihood
DDPI model	1.3%	0.6264	-2057.083
BFS-LE model	21.1%	0.5677	-1206.231
Labelling model	27.8%	0.4728	-1188.331

combined for each OD pair to make the comparison fair (resulting in a maximum of 41 alternatives for 695 OD pairs, which is the same input for all models). For five random draws the models are estimated and validated. Table 6 shows the results of the validation.

The DDPI model has lower parameter values compared to the other models. This means for the validation that it does not punish the less attractive alternatives as much as the other models. Consequently, the maximum utility for one alternative is low and similar for all alternatives. This results in a very low percentage of correctly predicted choices. The BFS-LE and labelling models score higher on this validation measure, and are on average able to predict at least one choice correct per OD pair. In terms of prediction per alternative, the two models that were estimated on a generated choice set that has a higher variability and includes both good (observed) routes and bad (generated) routes, perform better.

In terms of the RMSE that is weighted over the OD pairs, the models perform similar (although the BFS-LE and labelling model outperform the DDPI model). This measure gives an indication on the average error that would occur when for example predicting the flows on the network. The DDPI model assigns a rather equal probability to all alternatives, resulting in an average error that is similar to the RMSE of the two other models. These models on the other hand, provide a low probability to the worse (generated) alternatives and a very high probability to the good (observed) alternatives.

The null log-likelihood for this set of alternatives (calculated using $LL(0) = -\sum_n \ln(J_n)$, with J_n being the number of alternatives in choice set C_n) is -1740.149. The closer the final log-likelihood is to zero, the better the out-of-sample performance is. Both BFS-LE and labelling models improve significantly compared to the null log-likelihood. The DDPI models, which are estimated using only observed information, perform worse on the out-of-sample data in terms of its added value compared to providing equal probabilities to all alternatives (null log-likelihood). Consequently, we can conclude that the DDPI method should not be used for prediction purposes.

7.3. Conclusions regarding model estimation and validation

Due to the small number of matches of generated routes with observed routes, the choice sets are enriched with observed routes. Consequently, the choice sets have more information compared to purely generated choice sets, introducing endogeneity. The models that are estimated using the different choice sets differ in their parameter values, *t*-test values and elasticities. This is in line with expectations as the size and composition of choice set are known to influence the model estimation (Bovy, 2009).

The DDPI model has lower parameter values and *t*-test values due to small variability in the choice set and issues with endogeneity. Due to the inclusion of the observed alternatives in the BFS-LE and labelling choice set, where they were not generated, these models perform very well as an artefact. The large variability between alternatives (especially in the labelling choice set) and inclusion of both relevant and irrelevant alternatives (especially in the BFS-LE choice set), increases the model fit compared to only using observed routes (DDPI method). The effect of explanatory variables on route choice is higher for the labelling model compared to the other models. The BFS-LE model is a less extreme version of the labelling model, with relatively high

parameter and *t*-test values but elasticities that are more similar to those obtained using the DDPI approach. The reason for this might be the number of alternatives that is included in the BFS-LE approach, which is generally 17 more than the labelling approach.

In terms of predictive powers, the DDPI model was expected to perform less as it is data-driven and might therefore react different to out-of-sample prediction than the labelling and BFS-LE models, which was confirmed by all validation measures. The DDPI method is not suitable for out-of-sample prediction.

8. Conclusions and future research directions

This paper presents the findings of an evaluation of a data-driven approach (DDPI) for choice set identification in travel behaviour analysis, performed by comparing the DDPI method to two choice set generation methods: BFS-LE method introduced by Rieser-Schussler et al. (2013) and the labelling approach introduced by Ben-Akiva et al. (1984). Bicycle GPS data from the city of Amsterdam was used as a case study. The comparison was based on three aspects. First, an analysis of the choice sets that are identified, which was evaluated by means of a qualitative (visual) analysis, a quantitative analysis, and the reproduction of observed routes. Second, estimation of a route choice models using the three identified choice sets, which were evaluated by means of calculating elasticities. And third, validation these models on out-of-sample data, which were evaluated by means of correctly predicted choices, RMSE per OD pair and the log-likelihood.

In conclusion, the data-driven DDPI method is useful when evaluating or analysing the alternatives in the choice set and can help in understanding the preferences of individuals (using model estimation). The DDPI is not suitable for prediction on out-of-sample data.

The ability of choice set generation algorithms to reproduce observed paths largely depends on the correctness of the underlying network. In this study, the network was intended for motorised traffic (i.e. not validated for bicycle traffic), resulting in choice sets that are not suitable for analysis of the alternatives for cyclists (e.g. in terms of composition and characteristics). Generally, cyclists are allowed to cycle against one-way streets, however this is not included in the network. Furthermore, cyclists do not necessarily comply to the traffic rules in the Netherlands, as exhibited in using links in the network that are not identified for cyclists (e.g. short cuts or pedestrian areas). The first can be incorporated in the network by making all links bi-directional; however, the latter is harder to incorporate. Consequently, a discrepancy arose between the observed routes and the generated routes. The number of generated routes that could be matched to observed routes was very low, partially due to network incompleteness. However, we tested the significance of this shortcoming by removing the affected OD pairs, and found that the number of matched routes was still very low, indicating that generating routes based on single network characteristics (as is done in these algorithms) does not match with the observed behaviour. In conclusion, the choice set based on observed behaviour provides a better source for analysing the alternatives than a generated choice set based on network characteristics.

Given the differences and similarities between the estimated choice models, we conclude that the DDPI method provides useful insights into behaviour. In terms of model fit, it performed worse than the generated choice sets, mostly due to lower variability between routes and their respective attributes. However, no additional network information is required for the DDPI method. Hence, it does not rely on the quality of the underlying network for information or routes that need to be generated. Mostly because of that reason this method is a valuable addition to the existing choice set generation methods, as it does provide insights into preferences of individuals regarding attributes.

The case study analysed in this paper gives first insights into the usefulness of the data-driven DDPI approach for travel behaviour analysis. In this study the data-driven choice set has been applied to bicycle route choice. Future research can test the usefulness of the proposed

DDPI method for other types of choice set generation, for example activity scheduling and destination choice, and for route choice models of other modes, for example the car, which potentially exhibits a larger degree of diversity of routes within a shorter time period, due to congestion and traffic lights. Next to that, the model is now estimated on data from one week. It would be very useful to test on a dataset that covers a longer period of time (e.g. a month), because this potentially increases observed variability and thereof reduces the risk of endogeneity. Furthermore, the performance of choice set generation methods depends on the quality of the underlying network. Future studies may match the observed routes and links to the existing network prior to the choice set generation so that missing links can be added to the network. This will potentially result in a higher reproduction of observed routes. Also, this will provide more routes per unique OD pair, therefore reducing the need for clustering. Finally, the methods in this study were tested using random utility theory (specifically the PSL models). A direction for future research could be to apply the method within the random regret framework and test its performance.

Acknowledgements

This research was supported by the Allegro project (no. 669792) which is financed by the European Research Council and Amsterdam Institute for Advanced Metropolitan Solutions. The data from the field experiment Bicycle Counting Week, used for this research was provided by the municipality of Amsterdam.

References

- Azevedo, J.A., Santos Costa, M.E., Silvestre Madeira, J.J., Vieira Martins, E.Q., 1993. An algorithm for the ranking of shortest paths. *Eur. J. Oper. Res.* 97–106. [https://doi.org/10.1016/0377-2217\(93\)90095-5](https://doi.org/10.1016/0377-2217(93)90095-5).
- Bekhor, S., Ben-Akiva, M.E., Ramming, M.S., 2006. Evaluation of choice set generation algorithms for route choice models. *Ann. Oper. Res.* 235–247. <https://doi.org/10.1007/s10479-006-0009-8>.
- Bellman, R., Kalaba, R., 1968. On *k*-th best policies. *J. Soc. Ind. Appl. Math.* 582–588. <https://doi.org/10.1137/0108044>.
- Ben-Akiva, M.E., Bergman, M.J., Daly, A.J., Ramaswamy, R., 1984. Modeling inter-urban route choice behaviour. In: Volmuller, J., Hamerslag, R.E. (Eds.), *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory*. Kluwer Academic Publishers, Boston, pp. 299–330.
- Ben-Akiva, M., Bierlaire, M., 1999. Discrete choice methods and their applications to short term travel decisions. In: Hall, W. (Ed.), *Handbook of Transportation Science*. Kluwer, Dordrecht, The Netherlands, pp. 5–34.
- Bierlaire, M., 2016. PythonBiogeme: a short introduction. Report TRANSP-OR 160706, Series on Biogeme. Transport and Mobility Laboratory, School of Architecture, Civil and Environmental Engineering. Ecole Polytechnique Fédérale de Lausanne, Switzerland.
- Bliemer, M.C., Bovy, P.H., 2008. Impact of route choice set on route choice probabilities. *Transp. Res. Rec.* 10–19. <https://doi.org/10.3141/2076-02>.
- Bovy, P.H., 2009. On modelling route choice sets in transportation networks: a synthesis. *Transp. Rev.* 43–68. <https://doi.org/10.1080/01441640802078673>.
- Broach, J., Gliebe, J., Dill, J., 2010. Calibrated labeling method for generating bicyclist route choice sets incorporating unbiased attribute variation. *Transp. Res. Rec.* 89–97. <https://doi.org/10.3141/2197-11>.
- Broach, J., Dill, J., Gliebe, J., 2012. Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transp. Res. Part A* 1730–1740.
- Casello, J.M., Usyukov, V., 2014. Modeling cyclists' route choice based on GPS data. *Transp. Res. Rec.* 155–161.
- Chen, P., Shen, Q., Childress, S., 2017. A GPS data-based analysis of built environment influences on bicyclist route preferences. *Int. J. Sustainable Transp.* 218–231.
- Dhakar, N.S., Srinivasan, S., 2014. Route choice modeling using GPS-based travel surveys. *Trans. Res. Rec.* 65–73.
- Dial, R.B., 2000. Bicriterion traffic equilibrium: T2 model, algorithm, and software overview. *Transp. Res. Rec.* 54–62. <https://doi.org/10.3141/1725-08>.
- Dijkstra, E.W., 1959. A note on two problems in connexion with graphs. *Numer. Math.* 269–271.
- Ehrgott, M., Wang, J.Y., Raith, A., Van Houtte, C., 2012. A bi-objective cyclist route choice model. *Transp. Res. Part A* 652–663. <https://doi.org/10.1016/j.tra.2011.11.015>.
- Fiets Telweek, 2015. Regiobrochure Nationale Fiets Telweek 2015. Fietstelweek, Breda (in Dutch).
- Fiorenzo-Catalano, S., 2007. *Choice Set Generation in Multi-modal Transportation Networks*. PhD thesis. Delft University of Technology.
- Flötteröd, G., Bierlaire, M., 2013. Metropolis-hastings sampling of paths. *Transp. Res. Part B* 53–66. <https://doi.org/10.1016/j.trb.2012.11.002>.
- Fosgerau, M., Frejinger, E., Karlstrom, A., 2013. A link based network route choice model

- with unrestricted choice set. *Transp. Res. Part B* 70–80. <https://doi.org/10.1016/j.trb.2013.07.012>.
- Frejinger, E., Bierlaire, M., Ben-Akiva, M., 2009. Sampling of alternatives for route choice modeling. *Transp. Res. Part B* 984–994. <https://doi.org/10.1016/j.trb.2009.03.001>.
- Ghanayim, M., Bekhor, S., 2018. Modelling bicycle route choice using data from a GPS-assisted household survey. *Euro. J. Transp. Infrastruct. Res.* 158–177.
- Halldórsdóttir, K., Rieser-Schussler, N., Axhausen, K.W., Nielsen, O.A., Prato, C.G., 2014. Efficiency of choice set generation methods for bicycle routes. *Euro. J. Transp. Infrastruct. Res.* 332–348.
- Hess, S., Quddus, M., Rieser-Schussler, N., Daly, A., 2015. Developing advanced route choice models for heavy goods vehicles using GPS data. *Transp. Res. Part E* 29–44. <https://doi.org/10.1016/j.tre.2015.01.010>.
- Hood, J., Sall, E., Charlton, B., 2011. A GPS-based bicycle route choice model for San Francisco, California. *Transp. Lett.* 63–75. <https://doi.org/10.3328/TL.2011.03.01.63-75>.
- Hoogendoorn-Lanser, S., 2005. Modelling Travel Behaviour in Multi-modal Networks. PhD thesis. Delft University of Technology.
- Lam, S.H., Xie, F., 2002. Transit path-choice models that use revealed preference and stated preference data. *Transp. Res. Rec.* 58–65. <https://doi.org/10.3141/1799-08>.
- Li, S., Muresan, M., Fu, L., 2017. Cycling in Toronto, Ontario, Canada: Route choice behavior and implications for infrastructure planning. *Transp. Res. Rec.* 41–49.
- Mansky, C., 1977. The structure of random utility models. *Theory and Decision* 229–254.
- Menghini, G., Carrasco, N., Schussler, N., Axhausen, K., 2010. Route choice of cyclists in Zurich. *Transp. Res. Part A* 754–765. <https://doi.org/10.1016/j.tra.2010.07.008>.
- Montini, L., Antoniou, C., Axhausen, K.W., 2017. Route and mode choice models using GPS data. Proceedings of the 96th Annual Meeting of the Transportation Research Board. TRB, Board. Washington.
- OVIN, 2011. Onderzoek Verplaatsingsgedrag Nederland 1986–2012. CBS, Den Haag.
- Pan-European Programme, 2014. Fourth High-level Meeting on Transport, Health and Environment. Retrieved from Paris Declaration: World Health Organisation & United Nations, Paris.
- Prato, G., 2009. Route choice modeling: past, present and future research directions. *J. Choice Model.* 19–28. [https://doi.org/10.1016/S1755-5345\(13\)70005-8](https://doi.org/10.1016/S1755-5345(13)70005-8).
- Prato, C.G., Bekhor, S., 2007. Modeling route choice behaviour: how relevant is the composition of choice set? *Transp. Res. Rec.* 64–73. <https://doi.org/10.3141/2003-09>.
- Prato, C.G., Bekhor, S., Pronello, C., 2012. Latent variables and route choice behavior. *Transportation* 299–319. <https://doi.org/10.1007/s11116-011-9344-y>.
- Prato, C.G., Halldórsdóttir, K., Nielsen, O.A., 2018. Evaluation of land-use and transport network effects on cyclists' route choices in the Copenhagen Region in value-of-distance space. *Int. J. Sustainable Transp.* 1–12.
- Ramming, M.S., 2002. Network knowledge and route choice. PhD thesis. Massachusetts Institute of Technology.
- Rieser-Schussler, N., Balmer, M., Axhausen, K.W., 2013. Route choice for very high-resolution data. *Transportmetrica A: Transp. Sci.* 825–845. <https://doi.org/10.1080/18128602.2012.671383>.
- Skov-Petersen, H., Barkow, B., Lundhede, T., Jacobsen, J.B., 2018. How do cyclists make their way? – a GPS-based revealed preference study in Copenhagen. *Int. J. Geogr. Inf. Sci.* 1469–1484.
- Ton, D., Cats, O., Duives, D., Hoogendoorn, S., 2017. How do people cycle in Amsterdam? Estimating cyclists' route choice determinants using GPS data from an urban area. *Transp. Res. Rec.* 75–82.
- van de Coevering, P., de Kruijf, J., Bussche, D., 2014. Bikeprint: policy renewal and innovation by means of tracking technology. Colloquium Vervoersplanologisch Speurwerk, Eindhoven (in Dutch).
- Zimmerman, M., Mai, T., Frejinger, E., 2017. Bike route choice modeling using GPS data without choice sets of paths. *Transp. Res. Part C* 183–196.