

Evaluating a data-driven approach for choice set identification

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

Publication date

2017

Document Version

Final published version

Published in

Proceedings of the International Choice Modelling Conference 2017

Citation (APA)

Ton, D., Duives, D., Cats, O., & Hoogendoorn, S. (2017). Evaluating a data-driven approach for choice set identification. In *Proceedings of the International Choice Modelling Conference 2017*

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 and Serge Hoogendoorn^a

^a Delft University of Technology, Department of Transport and Planning, Stevingweg 1, 2628 CN Delft, The Netherlands

Abstract

The specification of the choice set for travel behaviour analysis is a non-trivial task, as 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 severe errors to the specified choice set: false negative (not generating observed routes) and false positive (including irrelevant alternatives) errors. Due to increased availability of revealed preference data, like GPS, it is possible to identify the choice set in different way: data-driven. The data-driven path identification approach (DDPI), introduced in this paper, combines all unique routes that are observed for one origin-destination pair into the choice set. This paper evaluates this DDPI approach, by comparing it to two choice set generation methods (breadth-first search on link elimination and labelling). The evaluation is based on three main purposes of choice sets: analysis of alternatives, model estimation and prediction. The conclusion is that the DDPI approach is a useful alternative for choice set identification. The findings indicate that in analysing alternatives, the DDPI approach is most suitable, as it is equal to the observed behaviour. For model estimation the DDPI approach provides a useful alternative to choice set generation methods, as it provides insights into the preferences of individuals. In terms of prediction, the DDPI approach is suitable on a network level, but not on the individual level. The average performance over all alternatives is similar for all choice sets, but on individual level the DDPI method does not predict well.

Keywords: data-driven choice set generation, BFS-LE approach, labelling approach, cyclists' route choice, travel behaviour analysis comparison

1. Introduction

In the context of travel behaviour, many choices have to 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 comes with several flaws.

In recent years, large improvements have been made in revealed preference data collection methods. New data sources, such as GPS, which 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 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 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 will not occur and the potential false positive error is negligible because all the routes included in the choice set have been chosen by the individual.

This paper aims to evaluate 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 alternative for choice set identification, compared to choice set generation algorithms. The data-driven approach has already been successfully applied in a bicycle route choice study (Ton et al., 2017). The evaluation of the method is done by means of a case study, where bicycle GPS data is used to identify the choice set and this choice set is used in the estimation and validation of a route choice model. In the case study the data-driven approach is compared to two choice set generation methods, in order 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).

In this paper, Section 2 reviews contemporary choice set classification and generation procedures. In Section 3, the data-driven approach is elaborated upon in terms of requirements of data, and opportunities and limitations of the method. Section 4 describes the methodology for evaluating the specified choice sets and 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 classification and generation

The specification of route choice sets is a non-trivial task and has been the topic of extensive research. This section aims at reviewing contemporary choice set classification and generation procedures in order to identify the most suitable methods for a comparison to the data-driven approach. Section 2.1 discusses the different classes of choice sets that have been identified. Section 2.2 details the different choice set generation methods that have been proposed in the state-of-the-art and provides an assessment of these methods.

2.1. Choice set classification

Choice sets may differ in size and composition, depending on who defines them and which aspects are taken into account. Hoogendoorn-Lanser & Van Nes (2004) have identified multiple choice set classes from an individuals' point of view (Figure 1). This classification differs for the individual and the researcher, but in this study we evaluate the individuals' perspective.

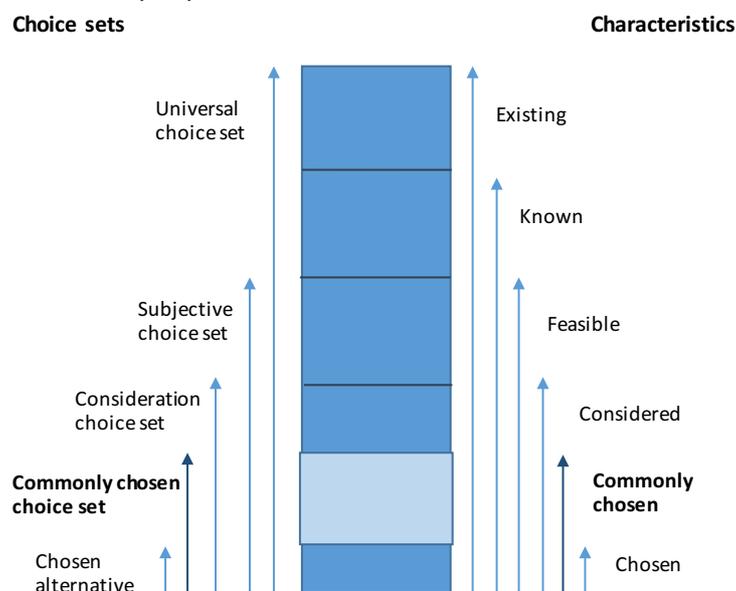


Figure 1: Classification and hierarchy of choice sets and characteristics from an individuals' perspective (adapted from Hoogendoorn-Lanser & Van Nes (2004))

The *universal choice set* contains all existing alternatives. The individual does not know all these alternatives. Hoogendoorn-Lanser (2005) found that individuals know only a limited number of alternatives; maximum seven alternatives for her study. The *subjective choice set* contains a subset of the universal set, which

the individual deems feasible (in time, money, physical ability etc.). The *consideration set* is again a subset of the previous sets, and includes only those alternatives the individual has considered for his trip. The other alternatives are not considered, due to e.g. the individual's characteristics and preferences. Hoogendoorn-Lanser & Van Nes (2004) state that the individual makes a choice from the consideration choice set. This set includes the alternatives, which the individual has weighed against one another to come to a choice, all other alternatives are left out of the decision. Consequently, the consideration set is preferred for travel behaviour analysis (Bovy, 2009).

Advancements in data collection methods have increased the amount of data available for route choice research. When using the observed trips of the individual to identify the choice set, as we propose in this paper, a new hierarchy level is introduced into the classification of Hoogendoorn-Lanser & Van Nes (2004): the *commonly chosen choice set* (see Figure 1). Combining all observed trips creates a choice set that contains alternatives that are chosen at least once by the (group of) individual(s). This new level is a subset of the consideration choice set, because not all routes that are considered will be chosen.

As mentioned before, most studies apply choice set generation methods to compose the choice set. Such methods are prone to false negative (not including the chosen alternative) and false positive (including irrelevant alternatives) errors. This mismatch is mainly attributed to the fact that the researcher and individual have different levels of knowledge of the network and the researcher is unaware of the exact preferences of the individual (Hoogendoorn-Lanser & Van Nes, 2004). Due to the occurrence of both errors, but especially false positive errors, a choice set is generated that is larger than the consideration choice set. The larger these errors in the choice set, the closer one comes to the universal choice set instead of the consideration set. On the other hand, the commonly chosen choice set generally is smaller than the consideration set. However, the occurrence of false positive errors is negligible and false negative errors are excluded by default. This indicates, that unless the choice set generation algorithm is able to minimise both errors, the commonly chosen choice set resulting from the data-driven approach will be closer to the the consideration set.

2.2. Choice set generation methods

Many different methods have been proposed for identifying route choice sets (for detailed reviews see Fiorenzo-Catalano (2007) and Ramming (2002)). Recently, two alternative approaches have been proposed that address the choice set generation differently. The first is the sampling approach (Frejinger et al., 2009; Flötteröd & 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 a choice at every node for the next link, given their destination. This approach consists of an unlimited choice set while avoiding the explicit enumeration of choice-set alternatives. These methods seem to be a valuable alternative to choice set generation methods. Both the choice set generation methods and the data-driven approach do not start with the universal or unlimited choice set, like these alternative methods. Instead they start respectively from the network and data. Therefore, we will focus on choice set generation algorithms in this paper.

An overview of the genealogy of the choice set generation methods, showing when they were first introduced in route choice modelling, is presented in Figure 2. Bovy (2009) and Prato (2009) identify four categories of choice set generation methods: deterministic methods, stochastic methods, probabilistic methods and constrained enumeration methods. The data-driven approach introduces a fifth category of choice set generation methods: the data-driven methods.

Most choice set generation methods belong to the *deterministic methods* 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. *Constrained enumeration methods* are not only based on shortest routes, but rely on 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. *Probabilistic methods* assign a probability for each alternative to be included in the choice set. A full probabilistic approach, as proposed by Manski (1977), which includes the choice set generation and selection in the utility function, is infeasible due to its computational complexity. Due to its prevalence in the literature, computational efficiency and deterministic nature, *deterministic methods* are selected as reference methods for the comparison in this study.

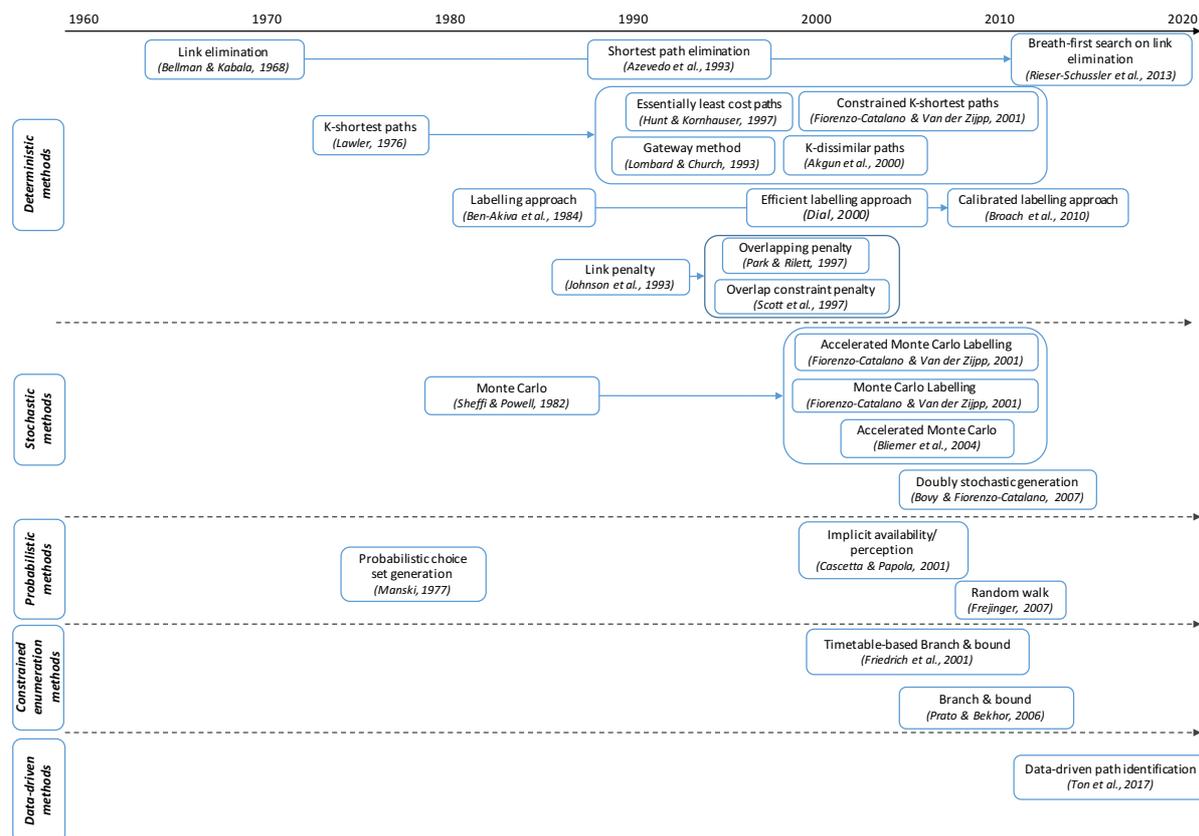


Figure 2: Choice set generation methods for route choice

Four categories of deterministic methods are identified (Figure 2): shortest paths, link elimination, labelling and link penalty. The *shortest path methods* have the lowest performance in terms of reproducing the observed routes (Bovy, 2009). Therefore, we will discuss only the better performing deterministic methods.

The *link elimination method* iteratively removes links that are on the shortest path and finds new shortest paths (Bellman & Kabala, 1968). Both Prato & Bekhor (2007) and Bekhor et al. (2006) evaluated this approach and found that in about 41% of the cases false negatives are produced (Bekhor et al., 2006; Prato & Bekhor, 2007). 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 a unrealistic choice set (e.g. large detours). Rieser-Schüssler 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-Schüssler et al., 2013). It has been successfully applied for different cases, e.g. cars (Rieser-Schüssler et al., 2013; Prato et al., 2012; Dhakar & Srinivasan, 2014; Montini et al., 2017), bicycles (Menghini et al., 2010; Halldorsdottir 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 alternatives given a certain label (e.g. distance, time, number of turns etc.). Prato & 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. Their false negative rate is 60% (Prato & Bekhor, 2007). Bekhor et al. (2006) specified and examined 16 different labels in their research. They found that each individual label generates only between 8% and 34% of the observed alternatives, however combined they are able to reproduce 72% of the observed routes. 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 (Broach et al., 2010).

Finally, the *link penalty method* penalises each of the links on the shortest route and repeats the shortest path search (Johnson et al., 1993). Prato & Bekhor (2007) found that this method outperformed the labelling approach in terms of generating observed alternatives. Bekhor et al. (2006), found contradicting results, mostly due to their specification of the labels. The link penalty method also performed less than the link elimination method, both in terms of generating observed alternatives and computation time. Two adaptations to the link penalty method have been proposed; first, a penalty that excludes the links closest to the origin and the destination, to avoid minor deviations at the start and end of the route (Park & Rilett, 1997) and second, a penalty combined with a restriction on the number of overlapping links (Scott et al., 1997).

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. Next to that, computation time of the methods is compared.

Table 1: Performance of applied deterministic choice set generation algorithms

Deterministic category	Method	Study	Mode	False negative error	Max no. alternatives	Comp. time
Link elimination method	Link elimination	(Bekhor et al., 2006)	Car	40%	?	Medium
		(Prato & Bekhor, 2007)	Car	42%	10	-
	Breath-first search on link elimination	(Rieser-Schussler et al., 2013)	Car	37%	20	-
		(Hess et al., 2015)	Trucks	26%	15	-
		(Halldorsdottir et al., 2014)	Bicycle	34%	20	Medium
Labelling approach	Labelling	(Bekhor et al., 2006)	Car	28%	16	Low
		(Prato & Bekhor, 2007)	Car	61%	3	-
		(Broach et al., 2010)	Bicycle	60%	4	-
	Calibrated labelling	(Broach et al., 2010)	Bicycle	80%	9	Low
Link penalty method	Link penalty	(Bekhor et al., 2006)	Car	78%	20	Medium
		(Bekhor et al., 2006)	Car	43%	40	High
		(Prato & Bekhor, 2007)	Car	44%	15	-
		(Prato & Bekhor, 2007)	Car	46%	15	-

The link elimination and labelling approach perform best from the deterministic methods, in terms of computation time and minimising false negatives in comparison to the number of alternatives generated (Table 1), albeit the latter remains a persistent and significant modelling issue. Furthermore, 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 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-Schüssler et al. (2013) is selected as a reference method in this paper, as 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. Next to that, 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 modelling issues associated with current choice set generation algorithms, the opportunity arises to identify choice sets using a data-driven approach. In this section a data-driven approach, named Data-Driven Path Identification (DDPI), is proposed.

The DDPI approach is based on GPS data of a large sample of individuals collected over a longer period of time. 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 all of these routes have been considered by an individual and from these 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. Therefore, it can only be asserted that the commonly chosen choice set is identified, whereas the considered choice set is preferred for travel behaviour analysis (Figure 1).

Several requirements need to be met for the DDPI approach to be applicable. First of all, 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 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. In the event that 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 can be useful if, for example, two neighbours heading for the same destination, are clustered into one OD pair. 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 issue with endogeneity can be the result. If endogeneity is a large issue, the method cannot be used, as it imposes a bias in the choice model. Another limitation is found in the generalisability of the results: as any travel behaviour study, 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. This depends on the representativeness of the individuals in the sample as well as the transferability of the regions, similarly to the generalisability issues associated with other methods. However, because the DDPI method is based on data, we want to highlight this issue.

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 discussed 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.

4.1. Evaluating the specified choice sets

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

4.1.1. Choice set generation algorithms

The BFS-LE and labelling approach are specified below.

Breadth-first search on link elimination (BFS-LE)

The BFS-LE algorithm, introduced by Rieser-Schüssler 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) 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). The algorithm used to calculate the shortest path is Dijkstra, thus requiring a distance matrix (Dijkstra, 1959).

In order 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 consists of fewer nodes and the resulting shortest path consist of fewer links, consequently significantly reducing the computation time.

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 is still high given the findings from Hoogendoorn-Lanser (2005) indicating that people only know seven alternatives. However, it will increase the probability of generating the observed route. Since we only search for 20 unique routes, we have applied a tree-depth of one. The second level sometimes generated over 1,000 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.

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. We have applied the Dijkstra algorithm for optimising each label (Dijkstra, 1959). The input-matrix required for the calculation, is adapted for each of the labels considered. The labelling approach also uses the topologically equivalent reduced network.

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 of all, 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. Then, a quantitative analysis is done, which provides descriptive statistics of three network related variables, based on previous work on bicycle route choice (Ton et al., 2017): 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 & Bierlaire, 1999).

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

Where PS_{in} is the path size factor, 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 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 ultimate objective of these algorithms is to reproduce all observed routes, i.e. resulting with zero false negative errors. In order to test to what extent the algorithm is able to reproduce the observed routes, the following formula for the reproduction rate is adapted from Prato & Bekhor (2007):

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

Where RR_r is the reproduction rate for algorithm r . $I(\cdot)$ 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 from no overlap (0%) to full overlap (100%). O_{nr} is calculated in the following way:

$$O_{nr} = \frac{L_{nr}}{L_n} \quad (3)$$

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 algorithm to the ideal algorithm, which would reproduce all of the observed

routes, and calculates how well the algorithms perform. The formula used to calculate this index is the following (Prato & Bekhor, 2007):

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

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, and the methodology to evaluate the model estimations and model validations is provided.

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 & Bovy (2008), Prato & 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.

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 & Bierlaire (1999) is adopted (Eq. 1). The probability of choosing alternative i given choice set C_n is specified as follows (Ben-Akiva & Bierlaire, 1999):

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

Where based on previous work, three explanatory variables are included per alternative i and observation n : percentage of separate cycle paths ($\% sep. cycle path_{in}$), 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.

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. 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. 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 the different sizes of the choice sets. Therefore, the comparison is based on the point elasticities of the models explanatory variables, calculated using the following formula:

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

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 remaining 20% OD pairs. In order to make a fair comparison between all models, a union of all generated and identified alternatives is generated for each OD pair. The union choice sets for each OD pair are used to assess the predictive power of all models, using two 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 modelled 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 .

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 is collected during a nationwide initiative in the Netherlands called the 'Bicycle Counting Week', which took place on 14-20 September 2015. 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 used to estimate a bicycle route choice model for this specific area (Ton et al., 2017). Figure 3 shows the network of the inner-city of Amsterdam.



Figure 3: Roadnetwork of the inner-city of Amsterdam

In total, 3,045 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 2,819 trips. The respondents sample consists of equal shares of male and female participants. The majority of the respondents is 31-65 years of age (80%). Most trips are made for commuting purposes (77%). Furthermore, most respondents cycle between 25 and 100 kilometre a week (72%) (Fiets Telweek, 2015). For more details on the collected data, the reader is referred to Ton et al. (2017).

The choice set generation algorithms use the network of Amsterdam (Figure 3) to generate the routes, therefore the network is extracted 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. However, in the city centre many pedestrian-only areas are present, which are not included in the network. Although cyclists are not allowed in these areas, they are being used by cyclists, causing some potential network-related issues. A total of 19,375 nodes is identified in the network. Due to the topologically equivalent network reduction, the number of nodes decreased to 7,628 nodes (-61%) with a total of 25,135 links.

The algorithms use the information from the network or any other data source that is available, this is especially relevant for the labelling algorithm as multiple labels can be defined. Due to the limited data availability on the network, only three labels can be identified, 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.

6. Generated choice set evaluation

The choice 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 Figure 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.

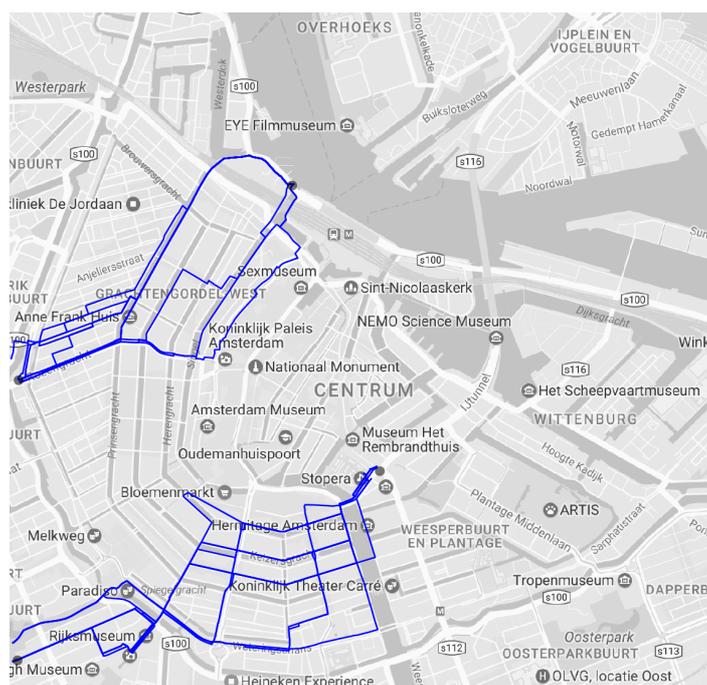


Figure 4: Observed routes from two selected OD pairs, plotted on the map of Amsterdam

The routes generated for the first OD pair using the BFS-LE and labelling approach are visualised in Figure 5, together with the observed routes. The observed routes (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 (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 (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.

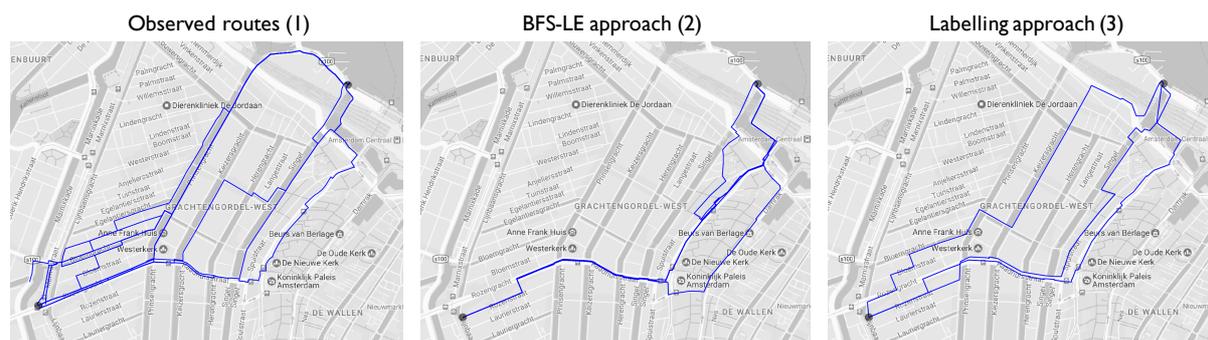


Figure 5: Routes generated for a given OD pair from the West of Amsterdam to the central train station, for (1) observed routes, (2) BFS-LE approach and (3) Labelling approach.

The generated choice sets for the second OD pair are visualised in Figure 6. The observed routes (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 (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 of the observed routes. The labelling approach (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.



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

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.

Table 2: Descriptive statistics of the explanatory variables and heterogeneity indicator for each choice set identification approach

Variable	Observed routes (N=2,819)			BFS-LE approach (N=12,361)			Labelling approach (N=2,034)		
	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

The observed routes show that the mean distance travelled is 1.9 kilometres, whereas the entire area included in the research covers about 6 kilometres. 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 are rather low, the latter was expected from the qualitative analysis.

Finally, the path size factor is on average 0.67, which indicates a rather heterogeneous set of routes, matching the results from the qualitative analysis. The routes chosen by all cyclists are spatially diverse and have a relatively low degree of overlap.

The BFS-LE approach optimises for distance, which is reflected in the lower mean distance and standard deviation compared to the observed routes. However, the difference is negligible, which seems to imply that the cyclists have a preference for shorter routes. As mentioned before, the pedestrian-only areas in the city centre 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 they are not allowed there, indicating that the true shortest path cannot be found by the algorithms. It shows that the true shortest 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 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.

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.

The false negative error for both methods is about 99%. 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 was able to partially mimic the observed routes, however the overlap between the observed and generated routes is lower than 70%, possibly due to the limited number of generated routes. The BFS-LE approach performs even worse, as was also visible in the qualitative analysis. As mentioned before, the network did not include pedestrian-only areas in the city centre where cyclists are not allowed. However, when excluding the affected trips (25%), the performance of the algorithms remains very poor (e.g. 1.8% compared to 1.4% of 100% reproduction for the labelling approach). The behaviour of the cyclists in Amsterdam can clearly not be captured by means generation of routes based on one objective.

6.4. Conclusions on 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 large false negative errors (not generating the observed alternative). When generating routes based on single network characteristics, the quality of the network influences the routes that are generated. In this case, the observed behaviour is not captured correctly. The differences indicate that cyclists optimise based on more than one network-related objective. The observed behaviour shows a combination of shorter distances, more cycle path and fewer intersections per km. 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 & 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 (including added observed routes). 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 them. However, in this case it would mean that only very few trips and routes 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. For the labelling approach, in most OD pairs 25% of the choice set (1/4) is based on observed behaviour, whereas for the BFS-LE approach this is mostly 5% (1/21). This means that information is added to the choice set, that will increase the performance of these choice sets in model estimation and also introduces an issue with endogeneity. 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. Table 4 shows the estimation results for the first 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)	-0.200	-1.67	-1.35	-3.00	-4.26	-20.13
% separate cycle path	0.150	0.98	7.57	7.37	4.63	13.13
Intersections/km	-0.016	-1.81	-0.268	-12.93	-0.244	-20.62
(Ln) Path Size	-0.395	-4.32	3.59	12.08	16.6	17.06
N	2,257		2,257		2,257	
Null log likelihood	-3,101.890		-6,507.034		-3,092.304	
Final log likelihood	-3,085.697		-142.796		-416.448	
Likelihood ratio test	32.387		12,728.475		5,251.711	
Adj. rho square	0.004		0.977		0.864	

The signs of distance, separate cycle path percentage and intersections per kilometre are as expected and are the same for each model. However, the 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. The routes chosen by the cyclists in the sample are relatively unique, but apparently the routes that bear more overlap with other routes are preferred. In case of the BFS-LE model, adding the observed routes creates a positive PS factor. The non-chosen alternatives overlap with each other, but often the chosen alternative is very different, resulting in a higher utility for the non-overlapping routes. This difference shows how the observed choice set is different from the generated choice sets (with additional observed routes).

In order to compare these models, the average point elasticities for every explanatory variable 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.

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

Variable	DDPI model Elasticity	BFS-LE model Elasticity	Labelling model Elasticity
Distance	-0.264	-0.044	-0.768
% separate cycle path	0.041	0.035	0.131
Intersections/km	-0.167	-0.167	-0.427

The interpretation of the elasticities is such that 10% increase in distance results in a decrease in the probability of being chosen of 2.64% for the DDPI model, whereas the BFS-LE model shows a 0.44% decrease and the labelling model shows a decrease of 7.68%. The relative difference between the impact of the BFS-LE model and DDPI model is 500% and is even larger 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 BFS-LE and DDPI

models are comparable for the intersection and cycle path variables. This indicates that increasing the variability in the alternatives (labelling routes plus observed route), induces a higher elasticity.

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

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

	Correct choice predicted	RMSE OD pair
DDPI model	1.3%	0.6264
BFS-LE model	23.7%	0.5842
Labelling model	23.9%	0.6432

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. 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 very low probability to the worse (generated) alternatives and a very high probability to the good (observed) alternatives. This indicates that in terms of prediction in the network the DDPI choice set is equally useful as the other methods.

7.3. Conclusions on the use of different choice sets in model estimation and validation

Due to the small number of matches of generated routes with observed routes, the choice sets are enriched with behaviour. Consequently, the choice sets have more information compared to purely generated choice sets (endogeneity). The models that are estimated using the different choice sets, differ in their parameter values, t-test values and elasticities. Which is according to 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. The labelling model has very high variability combined with both attractive and unattractive alternative, which results in high parameter, t-test and elasticity values. This model punishes an alternative more when it is slightly worse than 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 like the DDPI approach. The reason for this might be the number of alternatives that is included in the BFS-LE approach: 20 or 21, compared to three or four for the labelling approach. In the model estimated on the small choice set, the differences between alternatives are less valued than if more alternatives are added.

Due to the inclusion of the observed alternatives in the BFS-LE and labelling choice set, where they were not generated, these models perform extraordinary well. 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), makes the model fit almost perfect. If this was not the case, we expect the models to perform less well. Due to this high variability for the labelling choice set, the mean elasticities are much higher than the other two models.

In terms of predictive powers, the DDPI model was expected to perform less as it is data-based and might therefore react different to out-of-sample prediction than the labelling and BFS-LE models. This was confirmed by the percentage of correct predicted choices (a result of the issue with endogeneity in model estimation), however in terms of RMSE the DDPI model did not differ from the other two models. On average the errors are similar. Therefore, the method can still be used for prediction, however only in terms of distribution of the network.

8. Conclusions and future research directions

This paper presented the findings of an evaluation of a data-driven approach (DDPI) for choice set identification in travel behaviour analysis, done 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). The evaluation was performed by means of a case study on bicycle GPS data from the city of Amsterdam. The comparison was based on three aspects. First, an analysis of the choice sets that are identified, by means of a qualitative (visual) analysis, a quantitative analysis and, the reproduction of observed routes. Second, the estimation of a route choice model using these choice sets, by means of calculating elasticities. And third, validation these models on out-of-sample data, by means of correctly predicted choices and RMSE per OD pair.

In conclusion, the data-driven DDPI method is useful when evaluating or analysing the alternatives in the choice set and is able to help in understanding the preferences of individuals (using model estimation). The DDPI is also useful when predicting for network distribution (average error), but is not suitable for predicting the exact chosen alternative.

The choice set generation algorithms applied in our case study are not most suitable for analysis of the alternatives (in terms of composition, characteristics etc.). In this study, the bicycle network had some issues. Cyclists are not compliant to the traffic rules in the Netherlands, resulting in ignoring one-way streets and using pedestrian-only areas for cycling. In the network, we were able to make all streets bi-directional, however not all pedestrian-only areas could be added to the network. This meant that a discrepancy arose between the observed routes and the potentially generated routes, emphasising the dependency of choice set generation algorithms to the network. The number of generated routes that could be matched to observed routes was very low, partially due to the network issue. However, if the affected OD pairs were removed, the number of matched routes was still very low, indicating that generating routes based on single network characteristics 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 believe the DDPI method to be a useful method for model estimation. In terms of performance it was comparable to the other methods that had additional information (added observed alternatives). The model can for example be used when one wants to know the preferences of individuals regarding attributes.

In terms of predictive power of the models (validation), the DDPI model can be used for prediction on network level, but not on alternative level. The low parameter values, indicate that alternatives are not punished as much for being worse than other alternatives (a result of the issue with endogeneity in model estimation), resulting in a relatively equal probability of being chosen. Consequently, it does not necessarily provide the highest probability to the chosen alternative. The other methods perform better in this respect. Whereas the average error is comparable for all models, therefore introducing similar errors in the distribution over the network. This means that the method can be used for prediction on network level.

This case study gives first insights into the usefulness of the data-driven DDPI approach for travel behaviour analysis. One issue that is now encountered is the limited variability between alternatives, this issue might be overcome by introducing a hybrid approach, where the DDPI approach is combined with a choice set generation algorithm. Another way to overcome this issue is to combine the DDPI approach with the Metropolis-Hastings sampling method (Flötteröd & Bierlaire, 2013), such that the importance sampling starts with the observed routes. Consequently, a choice set might be created that has larger variability but includes the valuable knowledge from the observed routes. Both combined methods, might make the method more suitable for prediction. Finally, in this case the data-driven choice set is applied to route choice, but we believe this could also be useful for other types of choice set generation, for example activity scheduling and destination choice.

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 from the field experiment Bicycle Counting Week, used for this research was provided by the municipality of Amsterdam.

References

Akgün, A., Erkut, E., & Batta, R. (2000). On finding dissimilar paths. *European Journal of Operations Research*, 232-246.

- Azevedo, J. A., Santos Costa, M. E., Silvestre Madeira, J. J., & Vieira Martins, E. Q. (1993). An algorithm for the ranking of shortest paths. *European Journal of Operational Research*, 97-106.
- Bekhor, S., Ben-Akiva, M. E., & Ramming, M. S. (2006). Evaluation of choice set generation algorithms for route choice models. *Annals of Operations Research*, 235-247.
- Bellman, R., & Kabala, R. (1968). On k-th best policies. *Journal of the Society for Industrial and Applied Mathematics*, 582-588.
- Ben-Akiva, M. E., Bergman, M. J., Daly, A. J., & Ramaswamy, R. (1984). Modeling inter-urban route choice behaviour. In J. Volmuller, & R. E. Hamerslag (Ed.), *Proceedings of the Ninth International Symposium on Transportation and Traffic Theory* (pp. 299-330). Boston: Kluwer Academic Publishers.
- Ben-Akiva, M., & 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.
- Bliemer, M. C., & Bovy, P. H. (2008). Impact of route choice set on route choice probabilities. *Transportation Research Record*, 10-19.
- Bliemer, M. C., Versteegt, E., & Castenmiller, R. (2004). INDY, a new analytical multiclass dynamic traffic assignment model. *Proceedings of the Fifth Triennial Symposium on Transportation Analysis*. Guadeloupe: TRISTAN V Conference.
- Bovy, P. H. (2009). On modelling route choice sets in transportation networks: A synthesis. *Transport Reviews*, 43-68.
- Bovy, P. H., & Fiorenzo-Catalano, S. (2007). Stochastic route choice set generation: Behavioral and probabilistic foundations. *Transportmetrica*, 173-189.
- Broach, J., Gliebe, J., & Dill, J. (2010). Calibrated labeling method for generating bicyclist route choice sets incorporating unbiased attribute variation. *Transportation Research Record*, 89-97.
- Cascetta, E., & Papola, A. (2001). Random utility models with implicit availability perception of choice travel for the simulation of travel demand. *Transportation Research Part C*, 249-263.
- Dial, R. B. (2000). Bicriterion traffic equilibrium: T2 model, algorithm, and software overview. *Transportation Research Record*, 54-62.
- Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerical Mathematics*, 269-271.
- Ehrgott, M., Wang, J. Y., Raith, A., & Van Houtte, C. (2012). A bi-objective cyclist route choice model. *Transportation Research Part A*, 652-663.
- Fiets Telweek. (2015). *Regiobrochure Nationale Fiets Telweek 2015*. Breda: Fietstelweek, [in Dutch].
- Fiorenzo-Catalano, S. (2007). *Choice set generation in multi-modal transportation networks*. PhD Thesis: Delft University of Technology.
- Fiorenzo-Catalano, S., & Van der Zijpp, N. J. (2001). A forecasting model for inland navigation based on route enumeration. *Proceedings of the European Transport Conference*. United Kingdom: Cambridge.
- Flötteröd, G., & Bierlaire, M. (2013). Metropolis-Hastings sampling of paths. *Transportation Research Part B*, 53-66.
- Fosgerau, M., Frejinger, E., & Karlstrom, A. (2013). A link based network route choice model with unrestricted choice set. *Transportation Research Part B*, 70-80.
- Frejinger, E. (2007). Random sampling of alternatives in a route choice context. *Proceedings of the European Transport Conference*. Leeuwenhorst.
- Frejinger, E., Bierlaire, M., & Ben-Akiva, M. (2009). Sampling of alternatives for route choice modeling. *Transportation Research Part B*, 984-994.
- Friedrich, M., Hofsaess, I., & Wekeck, S. (2001). Timetable-based transit assignment using branch and bound techniques. *Transportation Research Record*, 100-107.
- Halldorsdottir, K., Rieser-Schussler, N., Axhausen, K. W., Nielsen, O. A., & Prato, C. G. (2014). Efficiency of choice set generation methods for bicycle routes. *European Journal of Transport and Infrastructure Research*, 332-348.
- Hess, S., Quddus, M., Rieser-Schussler, N., & Daly, A. (2015). Developing advanced route choice models for heavy goods vehicles using GPS data. *Transportation Research Part E*, 29-44.
- Hood, J., Sall, E., & Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California. *Transportation Letters*, 63-75.
- Hoogendoorn-Lanser, S. (2005). *Modelling travel behaviour in multi-modal networks*. PhD Thesis: Delft University of Technology.
- Hoogendoorn-Lanser, S., & Van Nes, R. (2004). Multimodal choice set composition: Analysis of reported and generated choice sets. *Transportation Research Record*, 79-86.
- Hunt, D. T., & Kornhauser, A. L. (1997). Assigning traffic over essentially-least-cost paths. *Transportation Research Record*, 1-7.
- Johnson, P. E., Joy, D. S., Clarke, D. B., & Jacobi, J. M. (1993). *Highway 3.1 - An enhanced highway routing model: program, description, methodology, and revised user's manual*. Oak Ridge: Oak Ridge National Laboratory.

- Lawler, E. L. (1967). *Combinatorial optimization: Networks and matroids*. (Holt, Ed.) New York: Rinehart & Winston.
- Lombard, K., & Church, R. L. (1993). The gateway shortest path problem: generating alternative routes for a corridor location problem. *Geographical systems*, 25-45.
- Menghini, G., Carrasco, N., Schussler, N., & Axhausen, K. (2010). Route choice of cyclists in Zurich. *Transportation Research Part A*, 754-765.
- 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*. Washington: TRB.
- Park, D., & Rilett, L. R. (1997). Identifying multiple and reasonable paths in transportation networks: a heuristic approach. *Transportation Research Record*, 31-37.
- Prato, C. G., & Bekhor, S. (2006). Applying branch and bound technique to route choice set generation. *Transportation Research Record*, 19-28.
- Prato, C. G., & Bekhor, S. (2007). Modeling route choice behaviour: How relevant is the composition of choice set? *Transportation Research Record*, 64-73.
- Prato, C. G., Bekhor, S., & Pronello, C. (2012). Latent variables and route choice behavior. *Transportation*, 299-319.
- Prato, G. (2009). Route choice modeling: Past, present and future research directions. *Journal of Choice Modelling*, 19-28.
- 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: Transport Science*, 825-845.
- Scott, K., Pabon-Jimenez, G., & Bernstein, D. (1997). Finding alternatives to the best path. *Proceedings of the 76th Annual Meeting of the Transportation Research Board*. Washington.
- Sheffi, Y., & Powell, W. B. (1982). An algorithm for the equilibrium assignment problem with random link times. *Networks*, 191-207.
- 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. *Proceedings of the 96th Transportation Research Board*. Washington: TRB.