Perception of overlap in multi-modal urban transit route choice

Dixit, Malvika; Cats, Oded; Brands, Ties; van Oort, Niels; Hoogendoorn, Serge

DOI
10.1080/23249935.2021.2005180

Publication date
2021

Document Version
Final published version

Published in
Transportmetrica A: Transport Science

Taylor & Francis
Taylor & Francis Group

Transportmetrica A: Transport Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ttra21

Perception of overlap in multi-modal urban transit route choice

Malvika Dixit, Oded Cats, Ties Brands, Niels van Oort & Serge Hoogendoorn

To cite this article: Malvika Dixit, Oded Cats, Ties Brands, Niels van Oort & Serge Hoogendoorn (2021): Perception of overlap in multi-modal urban transit route choice, Transportmetrica A: Transport Science, DOI: 10.1080/23249935.2021.2005180

To link to this article: https://doi.org/10.1080/23249935.2021.2005180

TRANSPORTMETRICA A: TRANSPORT SCIENCE
https://doi.org/10.1080/23249935.2021.2005180

Perception of overlap in multi-modal urban transit route choice

Malvika Dixit, Oded Cats, Ties Brands, Niels van Oort and Serge Hoogendoorn

Department of Transport & Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Netherlands

ABSTRACT
Capturing unobserved correlation between overlapping routes is a non-trivial problem in route choice modelling. For urban transit networks, research so far has been inconclusive on how this overlap is perceived by travellers. We estimate a series of path size correction logit (PSCL) models to account for alternative specifications of route overlap, including a new definition of overlap in terms of transfer nodes is proposed for multi-leg journeys. Our estimation is performed on smart card data from Amsterdam. The results indicate that the overlap between transit routes is valued positively when incorporated using either link-based, leg-based or transfer node-based PSC individually, with the transfer node-based PSC resulting in the best model fit. When considered simultaneously, the overlap of transfer nodes is valued positively by the travellers, but the subsequent overlap of journey legs is valued negatively, implying that travellers prefer having multiple (distinct) travel options at common transfer locations.

ARTICLE HISTORY
Received 16 November 2020
Accepted 31 October 2021

KEYWORDS
Public transport; route overlap; smart card data; path size correction logit; circuity

1. Introduction

Public transport plays an important role in making cities more sustainable and liveable. To that end, policy-makers and transit agencies are always striving to make transit more attractive to its users. Understanding how travellers choose between alternate transit routes is useful when planning and designing systems. It can help improve transport models and their predictions, and better assess interventions and network improvements, eventually leading to an increased transit usage.

Route choice models were traditionally developed for road networks, but the last decades have seen a rise in its applications to transit networks. Until recently, these models were mainly based on stated preference data sources, which although valuable in its own right, suffer from a common drawback of discrepancy between stated and actual behaviour (Yap, Cats, and van Arem 2020). Smart card provides a rich data source for analysing route choice by providing information on the actual choices made in the network, as well as the observed travel times at a high spatio-temporal resolution. Yet, only a handful of studies

CONTACT Malvika Dixit M.Dixit-1@tudelft.nl Department of Transport & Planning, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Stevinweg 1, 2628CN Delft, Netherlands

have used it on a large-scale multi-modal transit network – namely Jánošíkova et al. (2014), Kim et al. (2019), Tan et al. (2015), and Yap, Cats, and van Arem (2020). The aim of this study is to leverage the large revealed preference dataset provided by the smart card to improve transit route choice modelling by investigating in-particular the perception of travellers regarding *overlap* between routes.

Unless explicitly accounted for, the overlap between alternative routes results in correlations between the unobserved components of routes' utilities. For road networks, it is widely accepted, and has been shown empirically, that this overlap is valued negatively by travellers (Bovy, Bekhor, and Prato 2008). However, this is not necessarily true for transit networks, where the negative perception of overlapping routes may be masked by the positive utility of having an improved level-of-service (e.g. shorter waiting time on a shared corridor) or more alternatives available in case of disruptions. Overlap between transit routes has been argued to add to the robustness of the trip (Anderson, Nielsen, and Prato 2017), which is further improved when complemented with coordinated schedules (van Oort and van Nes 2009). The research on how overlap is perceived by travellers during transit route choice is inconclusive so far, with some studies reporting a positive valuation (Hoogendoorn-Lanser and Bovy 2007; Anderson, Nielsen, and Prato 2017), while others reporting a negative valuation (Yap, Cats, and van Arem 2020; Tan et al. 2015). In this paper, we investigate this issue further by analysing the *different specifications of consideration of overlap* between transit alternatives, to identify the circumstances and underlying reasons for its impact on passengers' route choice.

Similar to road networks, transit routes can have a partial overlap with one or more links being shared by multiple routes. Moreover, for routes that involve a transfer, there could be an overlap of entire journey leg(s). Hoogendoorn-Lanser et al. (2005) defined the overlap in terms of number of legs overlapped, as opposed to links overlapped for road networks. In Tan et al. (2015), the authors used link-level overlap, but proposed additional definitions in terms of travel time of overlapped links, also incorporating the frequency of overlapped routes. In the case of urban transit networks, it is not yet clear how each of these types of path overlap (link and leg) is perceived by travellers. To the best of our knowledge, these have not been compared in the literature so far.

Furthermore, research so far has defined and considered overlap exclusively in terms of paths (or path-based overlap). In this study, we propose an additional definition of overlap between routes in terms of common transfer nodes (or transfer stops). From a traveller's perspective, each transfer node is a decision point, where he/she can choose between alternative transit lines. The expectation is that there are utility benefits associated with routes that share a transfer node, because of the multiple transit options it provides to the travellers, making the overlapped routes more robust compared to independent routes. This alternative definition of overlap is compared against the usual definition based on overlapped links and legs. Further, we distinguish between the valuation of overlap of paths versus nodes by considering them separately as well as together.

The main contributions of this paper are twofold. First, it adds to the handful of empirical studies using large-scale revealed preference (smart card) data for estimating multi-modal transit route choice models. In doing so, it provides RP-based valuations of *mode-specific* travel time and transfer attributes, which to our knowledge are not available at such granularity in the literature so far. Second, it undertakes a comprehensive investigation of overlap between transit routes by defining overlap in terms of both the path (links and legs) and transfer nodes. We report results from our application of route choice models using smartcard data for the urban transit network of Amsterdam, the Netherlands.

We start with a Multi Nomial Logit (MNL) model of route choice that includes mode-specific in-vehicle and waiting times; number and type of transfers; transfer time; circuity of routes; and mode-specific constants. The base MNL model is then compared against the alternate formulations of Path Size Correction Logit (PSCL) models defining overlap in terms of links, legs, and transfer nodes. Lastly, the path-based and node-based PSCL formulations are considered together to investigate the relative contribution of each of these to the utility of overlapping routes.

The rest of the paper is organized as follows: Section 2 describes the approach used for quantifying and incorporating overlap in route choice models. In Section 3, the steps followed for processing and preparing smart card data are described including the model specifications. Section 4 presents and discusses the results of model estimation and validation, followed by the conclusions in Section 5.

2. Overlap in transit route choice

2.1. Definitions

In this study, a transit *journey* refers to the travel made by an individual from an origin transit stop to their destination transit stop, using a *route* that may involve transfer(s) within or across different transit modes, such as bus, tram or metro. A journey may contain multiple *legs*. A leg represents a part of the journey undertaken using a single transit vehicle. A *transfer node* is defined as the transit stop where the traveller transfers between multiple legs of a journey. Each leg may consist of multiple *links* which refer to the physical path connecting two consecutive transit stops on the route.

2.2. Background

Both traffic and transit route choice typically consists of overlapping route alternatives, resulting in correlation between unobserved characteristics of overlapping routes. In the case of road networks, overlap between routes results in the utility of the overlapping routes being overestimated. This is because the routes with an overlap may not be perceived as being distinct from the perspective of a traveller, and are hence less likely to be chosen compared to similar independent routes. The basic MNL formulation assumes the unobserved characteristics of alternatives to be independent, i.e. Independence of Irrelevant Alternatives (IIA) property. To incorporate the overlap between alternatives, there are two common approaches – either explicitly modelling it by making assumptions on the correlation between error terms (such as in error component logit model), or adding a deterministic term in the utility function to approximate the correlation (such as in C-logit – Cascetta et al. 1996 or path-size logit [PSL] – Ben-Akiva and Bierlaire 1999). This study follows the latter approach, which can be more directly specified and interpreted, and is commonly adopted in practice (Frejinger and Bierlaire 2007).

Approaches such as C-logit, PSL and path-size correction logit (PSCL) aim to reduce the utility assigned to overlapping routes, thus resulting in a lower probability compared to completely independent routes (Prato 2009). The reduction in utility in these models is often proportional to the length (Cascetta et al. 1996) or cost/time (Ramming 2002) of overlapping links. This is intuitive in the case of road networks – the higher the proportion of the route overlapped, the more they are expected to be considered alike by travellers. C-logit model has been found to be generally outperformed by the PSL, because of which most recent studies use path size-based models (Prato 2009). In terms of performance PSL and PSCL have been found to yield similar results (Bovy, Bekhor, and Prato 2008). We choose to use PSCL in this study owing to its stronger theoretical foundation (Bovy, Bekhor, and Prato 2008; Tan et al. 2015).

Under the PSCL model, as defined by Bovy, Bekhor, and Prato (2008), the expression for probability of a route alternative 'i' is given by

$$P_i = \frac{\exp(V_i + \beta_{PSC} PSC_i)}{\sum_{j \in C} \exp(V_j + \beta_{PSC} PSC_j)} \tag{1}$$

where V_i = deterministic utility of route alternative i,
PSC_i = path size correction term of route alternative i,
β_{PSC} = parameter for the PSC term to be estimated, and
C = choice set of all alternative routes.

The path size correction (PSC) factor in its original form is given by,

$$PSC_i = -\sum_{a \in \Gamma_i} \left(\frac{l_a}{L_i} \ln \sum_{j \in C} \delta_{aj} \right) \tag{2}$$

where l_a = length of link a within alternative route i,
L_i = total length of alternative route i,
Γ_i = set of all links for route i,
C = set of all routes between the chosen origin-destination pair, and
δ_{aj} = link-route incidence between link a belonging to alternative route j.

The PSC term has a maximum value of 0 for completely independent routes and decreases as the overlap between routes increases, with a theoretical lower bound of $-\infty$. For road networks, β_{PSC} associated with the PSC term is typically positive, resulting in a reduction of utility for overlapped routes (since PSC itself is negative for such routes).

While for road networks, there is a consensus on how the route overlap is defined and perceived by travellers, in case of transit networks the answer is not as clear. Hoogendoorn-Lanser, van Nes, and Bovy (2005) were the first to incorporate overlap in case of transit route choice. They defined overlap in terms of number of journey legs, travel time, and distance on those legs, and found that the overlap is valued negatively for all of these definitions (i.e. overlapped routes are less likely to be chosen). Contrastingly, Hoogendoorn-Lanser and Bovy (2007) found that the overlap in the train-leg of the multi-modal inter-urban journey was valued positively by the travellers, unlike the access and egress parts which were valued negatively. For urban transit networks also, there is evidence of a positive valuation of overlap (Anderson, Nielsen, and Prato 2017). As argued by Hoogendoorn-Lanser, van Nes, and Bovy (2005), the negative perception of overlapping routes in case of transit networks may be compensated by their contribution to robustness of routes in case of disruptions.

One of the important questions in case of transit networks is how the overlap should be defined and formalized mathematically. The formulation in Equation (2) was developed for road networks, and is often directly adopted for transit networks by defining overlap in terms of common physical links and their properties (for example in Anderson, Nielsen, and Prato 2017). Tan et al. (2015) proposed a formulation based on travel time on the overlapping links and frequency of services, rather than the link length. Following a different approach, Hoogendoorn-Lanser, van Nes, and Bovy (2005) defined overlap in terms of common trip legs (as opposed to links) for inter-urban multi-modal transit routes. They found that the overlap defined in terms of number of trip legs explained the observed choices better, as opposed to travel times or distances on those legs. None of the studies so far have compared the alternative ways for defining and quantifying path overlap (link and leg) in the context of urban transit route choice.

Further, as per Hoogendoorn-Lanser, van Nes, and Bovy (2005), apart from physical path, overlap between routes can also be defined in terms of nodes, services, runs, or modes. However, applications so far have been limited to path overlap only. We argue that in case of transit route choice, decision points are pertaining to transfer nodes where you may interchange, as opposed to links in case of road networks, where each intersection is a decision point. Hence, in this study, we include both path (link and leg) overlap and transfer node (decision point) overlap.

Based on the literature reviewed, we conclude that the following two questions remain unanswered regarding the perception of overlap in transit route choice:

- Is the overlap between alternate transit routes perceived positively or negatively by the travellers?
- Which way of defining overlap – link, leg or transfer node – best captures the perception of travellers for urban multi-modal transit route choice?

In the next sub-section we describe the approach adopted in this study for addressing the abovementioned questions.

2.3. Research approach

We now discuss the different possibilities of such overlap in case of urban multi-modal route choice, and our approach for addressing overlap in the form of common links, journey legs, and/or transfer nodes. We start with journeys without transfers (i.e. single leg journeys), and subsequently extend the approach to journeys with transfers. Figure 1 shows the possible overlaps for transit route alternatives without a transfer (single leg journeys), which could either be a partial or complete overlap of physical paths of the transit lines. Routes A and B/C have a partial overlap of physical paths with only two of the links overlapping, whereas Routes B & C have a complete overlap.

In Amsterdam, where we perform our case study, a map showing physical path of transit lines is displayed at the transit stops, providing travellers with information to choose an alternate overlapping transit line. Moreover, real-time passenger information systems are provided at the majority of stops, showing the next arriving transit vehicle(s). Hence, in this study, we assume that completely overlapping transit lines using the same mode are perceived as being the same by travellers. Accordingly, for such lines, the effective waiting time at the origin stop is calculated based on the combined observed headway of overlapped lines, as derived from the Automatic Vehicle Location (AVL) data. For the other case when

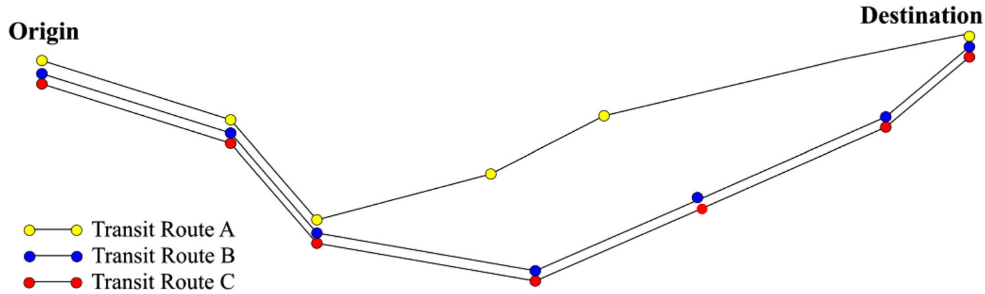


Figure 1. Overlap between transit routes without a transfer.

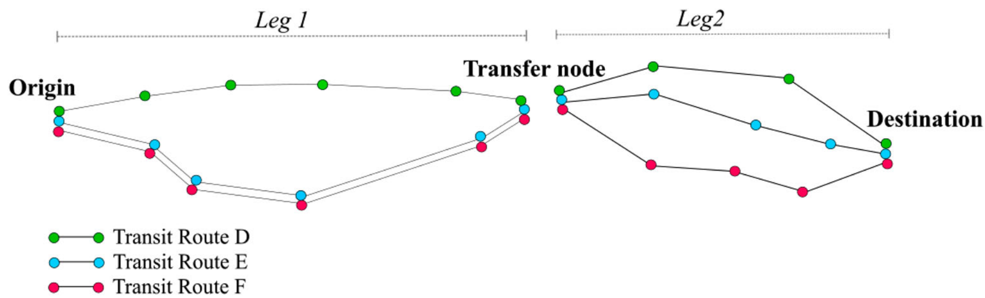


Figure 2. Overlap between transit routes with one transfer.

there is a partial overlap of transit lines (Routes A and B/C), a link-based overlap is considered. In some cases, there may be a complete overlap of physical path but different modes are used – bus and tram in our case. For such cases, a leg-based overlap (and no link-based overlap) is considered.

Next, we consider routes with multiple legs. For such routes, in addition to link and leg-based overlap, there could be a common (or overlapping) transfer node amongst the alternatives. Figure 2 shows an example of leg and transfer node overlap for routes with one transfer. Alternative E & F share overlapped leg 1, and all alternatives share an overlapped transfer node. Note that it is possible for routes to have no overlap of path (link or legs), but still have an overlap of transfer node(s) (alternatives D and E/F in Figure 2).

For all such possibilities of overlap, the unobserved characteristics of overlapping routes may be correlated. Hence, the utility is modified to take into consideration such overlap. We specify and analyse four PSC formulations for such routes – one each for link and node-based overlap, and two for leg-based overlap:

(1) **Link-based PSC:** This follows the traditional definition of PSC, as presented in Equation (2), and is based on the length of overlapping links as a proportion of the total route length.
(2) **Leg-based PSC - number of overlapped legs:** The hypothesis here is that travellers perceive the overlap in terms of number of overlapped legs, rather than the travel times or distance on those legs. In Hoogendoorn-Lanser, van Nes, and Bovy (2005), a similar definition was used for calculating the path size logit term for inter-urban transit travel.

The path-size correction term in this case is given by:

$$PSC_i^L = -\sum_{l\in\Gamma_i}\left(\frac{1}{N_i}\ln\sum_{j\epsilon C}\delta_{lj}\right) \tag{3}$$

where N_i = Number of journey legs in route i,
Γ_i = set of all legs for route i,
C = set of all routes between the chosen origin-destination pair, and
δ_{lj} = leg-route incidence between leg l belonging to alternative route j.

(3) **Leg-based PSC – travel times on overlapped legs:** The PSC term in this case is calculated based on the travel time on the overlapping leg as a proportion of total travel time of the route, as proposed by Tan et al. (2015):

$$PSC_i^T = -\sum_{l\in\Gamma_i}\left(\frac{t_l}{T_i}\ln\sum_{j\epsilon C}\delta_{lj}\right) \tag{4}$$

where t_l = travel time for journey leg l in route i,
T_i = total travel time for route i,
Γ_i = set of all legs for route i,
C = set of all routes between the chosen origin-destination pair, and
δ_{lj} = leg-route incidence between leg l belonging to alternative route j.

(4) Transfer Node-based PSC: This factor captures the overlap in terms of number of decision points for multi-leg journeys, and is given by:

$$PSC_i^X = -\sum_{n\in K_i}\left(\frac{1}{X_i}\ln\sum_{j\epsilon C}\delta_{nj}\right) \tag{5}$$

where X_i = Number of transfer nodes in route i,
K_i = set of all nodes for route i,
C = set of all routes between the chosen origin-destination pair, and
δ_{nj} = node-route incidence between node n belonging to alternative route j.

The models incorporating the above PSC factors are tested against the MNL model to establish how the addition of each PSC impacts the model fit. Table 1 summarizes the different possibilities of path and node overlap in our data, and the PSC formulations applied in each case.

3. Data preparation

We perform our analysis on the urban transit network in Amsterdam, consisting of bus, tram, and metro modes. The time period of analysis is 28 May–29 June 2018, during which 41 bus lines, 15 tram lines, and 4 metro lines were operational in the network. Figure 3 shows a map of the transit network during our analysis period. The metro network forms a part-ring structure around the city centre, with two of the lines providing direct connections from the south-eastern and southern peripheries of the city to the city centre. The tram lines have a dense network in the city centre, while also serving as feeders to the metro network. The bus network mainly serves the outskirts of the city with relatively lower density areas,

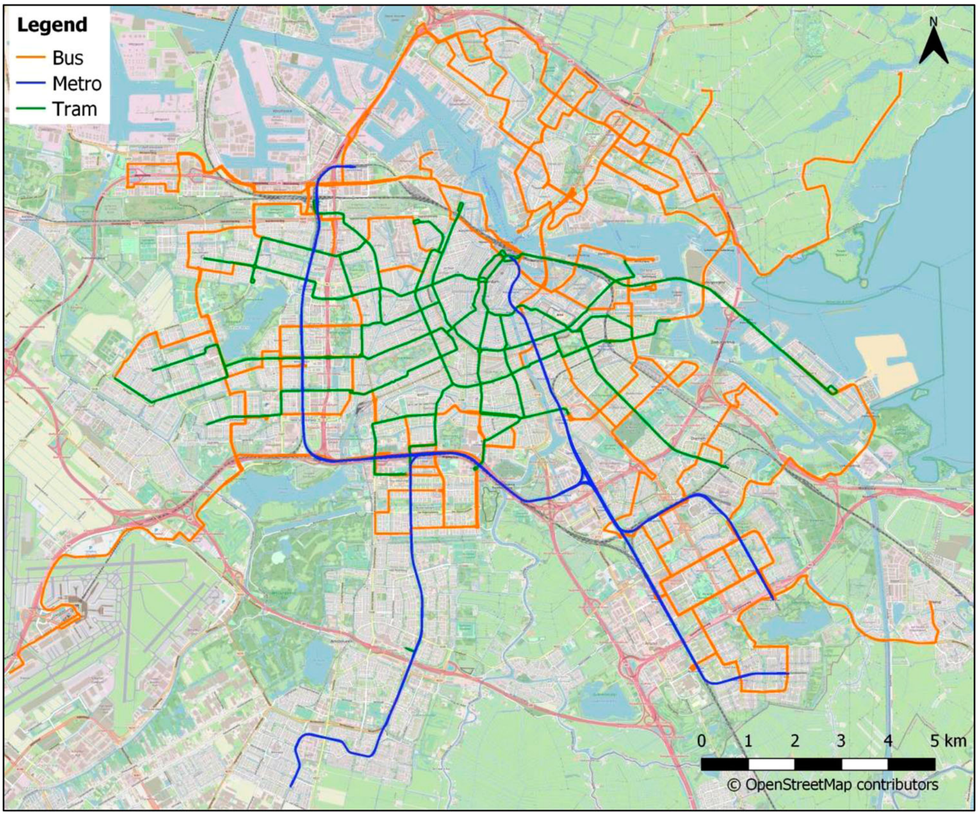


Figure 3. Amsterdam transit network.

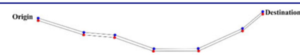
but provides some important connections, especially from the areas in the North to the city centre.

We use (anonymized) smart card data for our analysis, which captures (nearly) all journeys made in the network. On an average day during our study period, over 675,000 smart card transactions were recorded in the network. We restrict our analysis to weekday AM peak period (7 am to 10 am), in order to maximize the proportion of commuters and regular travellers in the data, who are expected to be more familiar with the route options, thereby making an informed route choice decision. The following subsections describe the steps undertaken to convert the raw smart card data to the required format for route choice analysis.

3.1. Trips to journeys

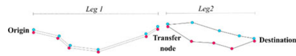
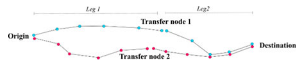
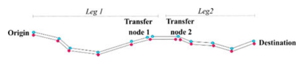
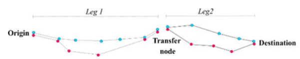
The Dutch smart card system provides information on both boarding and alighting transit stops and times (for an overview of the Dutch smart card system see van Oort, Brands, and de Romph 2015). Each transaction in the raw smart card data represents a check-in and check-out made by a passenger. For the route choice analysis, it is required to trace the entire journey of the travellers from their origin transit stop until their destination transit stop. For this, we combine individual smart card transactions to form passenger journeys

Table 1. Types of overlap considered.

Journey type	Type of overlap	Example	Link-based PSC	Leg-based PSC	Transfer node-based PSC
No transfer (single leg)	Complete overlap of path using the same mode		Assumed to be perceived as the same alternative		
	Complete overlap of path using different modes	–	✓	–	
	Partial overlap of path using same or different modes		✓	–	–

(*continued*).

Table 1. Continued.

Journey type	Type of overlap	Example	Link-based PSC	Leg-based PSC	Transfer node-based PSC
With transfer (multi-leg)	Complete overlap of one or more legs using same or different modes		–	✓	✓
	Partial overlap of one or more legs using same or different modes		✓	–	–
	Different transfer nodes but using the same route		✓	–	–
	Same transfer node but different/partially overlapped routes		if applicable	–	✓

by identifying transfers. For Amsterdam, a maximum time threshold of 35 min is applied by the operator to classify consecutive trips by an individual as transfers. However, we apply additional criteria to ensure that trip-generating activities conducted within the 35-min criteria are not wrongly classified as transfers. We do this by fusing the smart card data with AVL data, and applying the transfer inference algorithm as proposed in Gordon et al. (2013) and Yap et al. (2017). For more details on how the algorithm is applied to our case study network, the readers are referred to Dixit et al. (2019). The transfers made within the metro network are not directly available from the smart card data, since the travellers do not need to check-in and out within the metro system. Hence, we have inferred the number of transfers using a shortest path (time-wise) approach for such journeys. The Amsterdam metro network consists of only four lines with no loops, and all origin-destination pairs within the network can be reached with a maximum of one transfer. In case of origin-destination pairs where more than one transfer stop is possible (due to parallel lines), we have assigned the transfer stop corresponding to the shortest travel time (calculated as the sum of in-vehicle and expected transfer waiting times).

After performing transfer inference, the journeys were filtered to remove those with any missing route attribute (such as observed headway from the AVL data). Further, in Amsterdam's network, almost all origin-destination pairs can be reached with a maximum of two transfers. Hence, for our analysis, we have excluded journeys with more than two transfers ($<$ 0.01% of all journeys), to minimize irrational traveller behaviour in our data. This resulted in a dataset of 2.9 million journeys for the whole study period (weekday AM peak) after applying all the filters.

3.2. Aggregating transit stops

Smart card data does not provide any information on where passengers actually began or ended their journey – only the origin and destination transit stops are known. Further, the smart card data used for this study is anonymized, and the users cannot be tracked across multiple days. Hence, no information is available from the smart card data on which transit stops were considered by the traveller while making their route choice decision. A simplistic assumption is to restrict a traveller's route choice set to all the route alternatives available at his/her boarding stop and alighting stop only. However, such an assumption greatly reduces the number, and diversity of alternatives considered in the choice set. More importantly, it is an unrealistic assumption for a city like Amsterdam, where the median walking feeder distance for bus stops is more than 300 m (Brand et al. 2017), while the distance between neighbouring transit stops may be as low as 100 m in transit dense areas. Therefore, assumptions are needed regarding passenger's access/egress stop choice set. For example, Kim et al. (2019) aggregated all transit stops at an intersection into a node. We follow a similar approach of clustering neighbouring stops together, by means of agglomerative hierarchical clustering. In this method, starting with each transit stop forming its own cluster, the closest clusters are merged until a maximum distance threshold of 500 m between any two stops within a cluster is achieved. The 500 m distance threshold is chosen to achieve compact clusters, as measured by the silhouette score. 651 transit stops resulted in 279 stop clusters. The resulting clusters are shown in Figure 4 where each dot corresponds to a transit stop and the transit stops belonging to the same cluster are shown in the

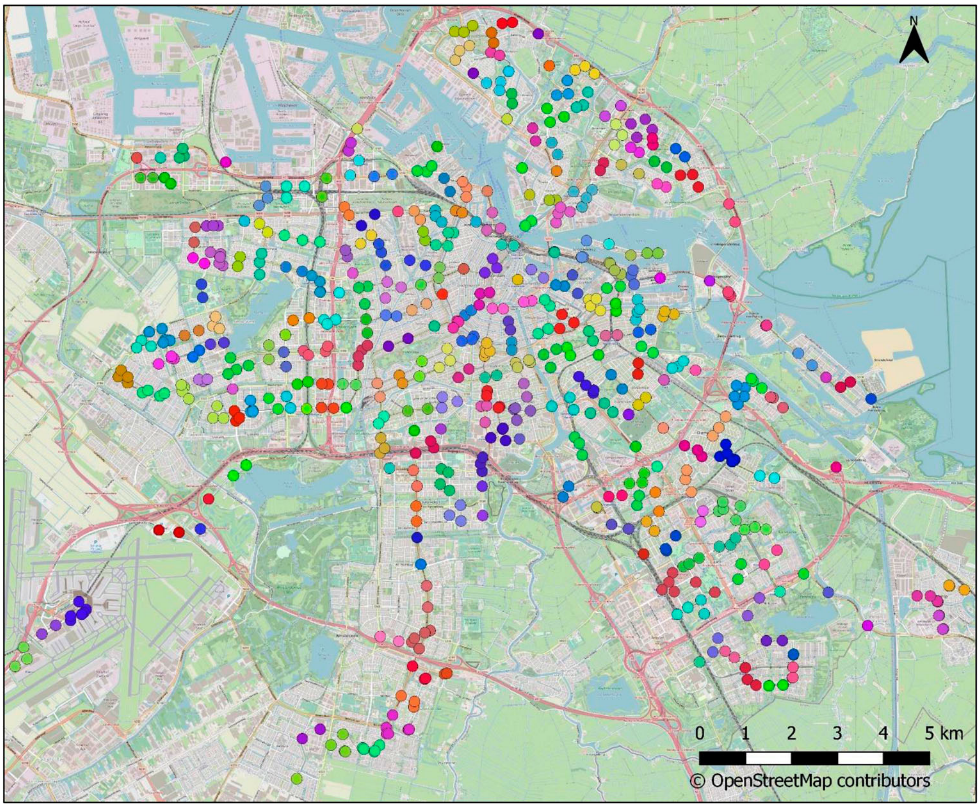


Figure 4. Clustering of transit stops in Amsterdam.

same colour. Once the transit stops are aggregated, all the routes connecting the traveller's origin and destination stop clusters can be included in his/her choice set.

3.3. Route definition and choice set preparation

A route between an origin and destination stop cluster (further on referred to as O-D) is defined as a combination of the physical path(s) followed by the transit line(s), and the transfer stop cluster(s) used for each leg of the journey. The choice set in this study is based on the observed routes taken by the travellers in the data. This eliminates the need for assumption on the feasibility of non-observed routes. Further, it allows us to calculate the route attributes (such as in-vehicle and transfer times) purely based on the observations from the smart card data.

To make a realistic representation of the available route choices, we split the study period (weekday AM peak) into six half-hour time slices and calculated the expected value of route attributes during each time slice. For example, a traveller starting their journey anytime between 7:00 and 7:30 am is assumed to have in their choice set the route alternatives that were used between that time window only. Accordingly, the expected value of route attributes is calculated by taking the median over all observations using the O-D route during the chosen time slice. Although seemingly large, the half-hour time slice is intended to maximize the number of observed journeys to have reliable estimates of route attributes.

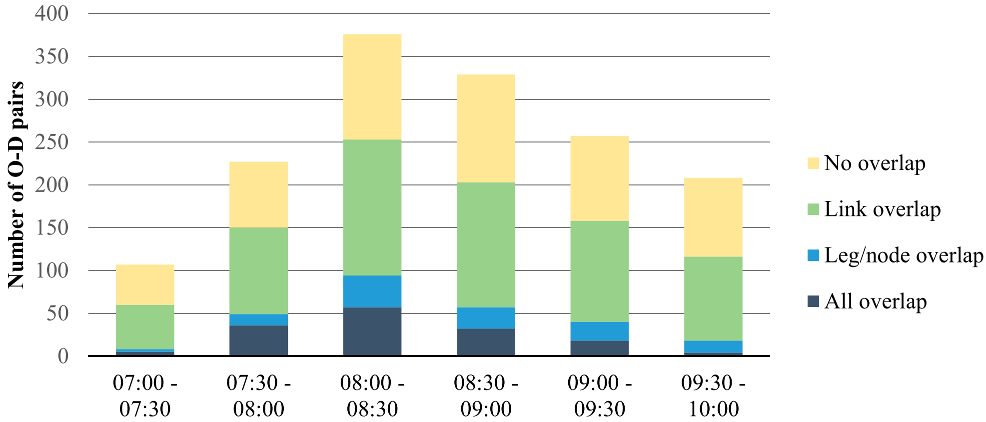


Figure 5. Number of unique O-D pairs in the data by type of overlap.

For that, only those routes with a minimum of 20 journeys in the time slice (over all days) were used for the analysis. Further, only those O-D pairs with a minimum of two route alternatives were included. Majority (91%) of O-D pairs had only two route alternatives, with a maximum of four alternatives for any O-D pair. Overall, the busiest time period within the morning peak was between 08:00 and 08:30 am, capturing 25% of all journeys. Figure 5 shows the number of distinct OD pairs used in the data for each time slice, and of those how many include an overlap in at least of the alternative routes. The proportion of OD pairs with at least some type of overlap varies between 56% and 67% for each time slice. Of these, majority have an overlap of link (only), which ranges between 42% and 49% of all O-D pairs for each time slice.

3.4. Route attributes and model specification

The way travel time is measured is different for bus and tram versus metro in Amsterdam. For buses and trams, the smart card is tapped in/out inside the vehicle, whereas for metro, this happens at the station. Because of that, the in-vehicle travel time for metro journeys is not directly available from smart card data. Hence, for this study, we derive the in-vehicle, waiting, and transfer times (if applicable) for metro journeys based on the AVL data. While the average in-vehicle time between all O-D pairs is directly available from the AVL data, the effective waiting and transfer times are derived based on the observed headway at each origin and transfer stations.

The following route attributes are populated for each O-D route and time-slice combination, which are subsequently used in our choice model specification:

- In-vehicle time by bus and tram (IVT_{bus} and IVT_{tram}): These correspond to the total in-vehicle time by bus and tram modes summed over all legs in a journey.
- Expected waiting time for bus and trams (WT_{bt}): This is calculated for the first leg of the journeys that start with bus or tram. For common lines, this is calculated based on their combined headway.

- Metro time (TT_{metro}): This includes in-vehicle time, and effective waiting time at the origin metro station.
- Number and type of transfers ($Trans_{bt}$, $Trans_{btm}$ and $Trans_m$): Transfers made are distinguished as being within bus/tram network (which includes bus-bus, tram-tram and bus-tram transfers); transfers between metro and bus/tram; and transfers within the metro network.
- Transfer time (TrT): This includes the transfer time for all types of transfers.
- Circuity ($Circ$): This quantifies the detours made in the route, and is calculated as the ratio of network to Euclidean distance of the route. It ranges from a minimum of 1 (for very short routes) to approximately 4 in our dataset.
- Mode-specific constants for bus, tram, and metro (MSC_{bus}, MSC_{tram} and MSC_{metro}): These incorporate the preference for a particular mode that is not captured by any of the above attributes.

A total of six route choice models are estimated. We start with an MNL model, the systematic utility of which is specified in Equation (6). All β represent the coefficients of the attributes which are estimated.

$$\begin{aligned} V^{MNL} = &\beta_{ivt_{bus}} * IVT_{bus} + \beta_{ivt_{tram}} * IVT_{tram} + \beta_{wait_{bt}} * WT_{bt} + \beta_{tt_{metro}} * TT_{metro} \\ &+ \beta_{trans_{bt}} * Trans_{bt} + \beta_{trans_{bt}} * Trans_{btm} + \beta_{trans_m} * Trans_m + \beta_{TrT} * TrT \\ &+ \beta_{Circ} * Circ + MSC_{Bus} + MSC_{Tram} + MSC_{Metro} \end{aligned} \tag{6}$$

Taking the above utility function as a basis, we add the PSC term to the utility function (as shown in Equation (1) earlier) and estimate five PSCL models. PSCL Models 1–4 define the overlap based on the link (Equation (2)), number of legs (Equation (3)), travel time on legs (Equation (4)), and number of transfer nodes (Equation (5)), respectively. Lastly, we combine the leg-based and node-based models to account for the contribution of each of those elements to the perception of route overlap. Different combinations of such model specifications were tried, and the best one is presented as PSCL Model 5.

4. Results and discussion

4.1. MNL model without overlap

The five PSCL models along with the MNL model as described in the previous section were estimated using the BIOGEME estimation package (Bierlaire 2020). Table 2 shows the results of the estimation. For all models, we present the final log-likelihood, rho-square-bar, and likelihood ratio statistic (LRS) with respect to MNL for comparison.

All parameters are found to be significant at $p < 0.01$ level, and the signs are as expected. All observed travel attributes being the same, there is a preference amongst travellers for using routes with metro over tram and bus. This is expected owing to the higher reliability the metro lines provide. Additionally, the simplicity of the metro network, weather protection at the stations, and a more comfortable waiting environment could be some other factors contributing to this preference. After metro, tram is found to be preferred by travellers over bus. Moreover, the in-vehicle time of bus is valued more negatively compared to tram, in line with findings from other studies reporting a ‘tram bonus’ in the Netherlands

Table 2. Model estimation results.

		Link-based	Leg-based		Node-based	Combined
Description	MNL Model	PSCL Model 1 – links	PSCL Model 2 – number of legs	PSCL Model 3 – travel time	PSCL Model 4 – transfer node	PSCL Model 5 – travel time + transfer node
Number of observations	382,295	382,295	382,295	382,295	382,295	382,295
Estimated parameters	11	12	12	12	12	13
Final log likelihood	−233,892	−233,767	−233,714	−233,764	−233,513	−233,473
Rho-square-bar	0.178	0.178	0.178	0.178	0.179	0.179
Likelihood Ratio Statistic (compared to MNL)	–	249.6	355.6	256.2	759.0	838.6
			Parameter estimates∗			
Mode-specific constant for bus (fixed)	0.00	0.00	0.00	0.00	0.00	0.00
Mode-specific constant for tram	0.49	0.48	0.50	0.50	0.51	0.50
Mode-specific constant for metro	0.84	0.84	0.86	0.88	0.90	0.90
Bus in-vehicle time (min)	−0.11	−0.11	−0.10	−0.11	−0.11	−0.11
Tram in-vehicle time (min)	−0.09	−0.09	−0.09	−0.09	−0.09	−0.09
Effective wait time bus/trams (min)	−0.19	−0.19	−0.19	−0.19	−0.19	−0.19
Metro time[a] (min)	−0.09	−0.09	−0.09	−0.09	−0.10	−0.10
Number of transfers between bus and tram[b]	−1.24	−1.24	−1.18	−1.26	−1.37	−1.41
Number of transfers between metro and bus/tram	−2.38	−2.40	−2.42	−2.47	−2.62	−2.65
Number of transfers within metro	−1.50	−1.50	−1.42	−1.44	−1.47	−1.51
Transfer time (min)	−0.25	−0.25	−0.25	−0.24	−0.23	−0.23
Circuity	−0.43	−0.41	−0.43	−0.42	−0.39	−0.39
Path size correction – link	–	−0.53	–	–	–	–
Path size correction – leg	–	–	−0.90	−0.63	–	0.52
Path size correction – transfer nodes	–	–	–	–	−1.02	−1.39

$*p < 0.01$ for all estimates.
[a]Includes in-vehicle time and origin waiting times.
[b]Includes bus-bus, tram-tram and bus-tram transfers.

(Bunschoten, Molin, and van Nes 2013). For buses and trams, one minute of waiting time at the origin stop is valued as much as 1.8 min of bus in-vehicle time, which is comparable with the values reported by Yap, Cats, and van Arem (2020) for The Hague, the Netherlands (1.5–1.6 min).

The pure transfer penalty between bus/tram modes, which is valued at 11.5 (−1.24/−0.11) minutes of bus in-vehicle time in the MNL model, is found to be much higher than the one reported by Yap, Cats, and van Arem (2020) for the buses and trams for The Hague (3.8–5.2 min), perhaps owing to Amsterdam's larger network with relatively longer transfer distances. Amsterdam also has a higher share of tourists that may prefer direct routes. Although on the higher end, our transfer penalties are found to be comparable to other values reported in the literature, such as those observed by Garcia-Martinez et al. (2018) for the multi-modal transit network of Madrid, Spain (15.2–17.7 min), or by Anderson, Nielsen, and Prato (2017) for the regional multi-modal transit network of Greater Copenhagen Area (14.1–17.9 min). Between the different types of transfers made in the network, the ones between metro and bus/tram are the least preferred. This is expected, since in case of Amsterdam, transferring to metro from bus/tram typically involves walking a longer distance and climbing (deep) stairs. In comparison, the transfers within the metro network are more convenient as metro stations are relatively small and transfers are often cross-platform and do not involve long passageways or multiple level changes. The transfer waiting times are also more reliable in case of metro, because many of these transfers are synchronized during operation. However, the most preferred transfers are within the bus/tram network which is usually at the same level.

In addition to the pure transfer penalty, travellers also have a strong dis-preference for transfer time compared to the corresponding in-vehicle time. Detour of a route, as measured by the circuity, is also found to play a role in explaining the route choice of travellers, with a trade-off of 4 min of bus in-vehicle travel time for 1 unit change of circuity. The part-ring structure of the metro network results in some O-D pairs having high circuity. Moreover, having a distance-based fare system implies that higher circuity also results in higher fares being paid. Despite that, our parameter value for circuity is found to be much lower than that reported by Kim et al. (2019) for the transit network of Seoul, which to our knowledge is the only other study that includes circuity in multi-modal transit route choice. They report a value of about 22 min of IVT for 1 unit change of circuity, possibly due to the larger scale of the network.

4.2. *Incorporating overlap*

Next, we examine how travellers perceive different aspects of overlap between alternate transit routes, looking at each type of overlap individually (PSCL Models 1–4). Firstly, all PSCL models offer an improvement of model fit compared to the MNL model, as demonstrated by the LRS = 249.6 for the worst performing PSCL model (PSCL Model 1), exceeding the critical χ^2 value of 6.6 at 1% significance level ($df = 1$).

Secondly, the sign of the PSC parameter for all models considering overlap individually (PSCL models 1–4) is significant (at $p < 0.01$ level) and negative, implying that the overlap between transit routes is perceived positively by the travellers in general, and the utility of overlapping routes is underestimated by the MNL model. This means that the travellers prefer routes having an overlap between alternatives, over completely distinct transit routes – be it an overlap of links, legs, or transfer nodes. This result contradicts with the findings from analysis of route overlap in road networks (such as in Bovy, Bekhor, and Prato 2008), as mentioned earlier in Section 2.2. However, it is in line with some studies for transit route choice (Anderson, Nielsen, and Prato 2017; Hoogendoorn-Lanser and Bovy 2007), and is perhaps explained by the claim of Anderson, Nielsen, and Prato (2017) that the PSC 'can be seen as a measure of robustness of the trip by the traveller'.

From the three path-based PSCL models (PSCL Models 1–3), using a PSCL based on number of journey legs (PSCL Model 2) explains the travellers' route choice better than the link-based or travel time-based ones, as demonstrated by the final log-likelihood and LRS (with the same number of parameters in the three models). However, out of the four proposed PSC formulations, the one based on transfer nodes (PSCL Model 4) is found to have the highest final log-likelihood, meaning that it best explains the observed data, significantly better than any of the path-based PSCL models. As with the path-based PSC parameters, travellers are more likely to choose a route that includes a transfer stop that is shared by other routes for the OD pairs, implying more options of travelling to their destination stop, which could also be considered to be more robust. It is also noted that once the node-based overlap is added, the transfer penalty is found to increase steeply for transfers within bus/tram and from bus/tram to metro. This implies that when such overlap is ignored, the positive utility derived for routes with a transfer node overlap is captured by the transfer penalty instead, leading to an underestimation of the disutility induced by the number of transfers. Simultaneously, the transfer time parameter is found to decrease marginally, further implying that travellers dislike transferring irrespective of the transfer time.

4.3. Combining path and node-based overlaps

As shown in Table 1, the notions of overlap in terms of transfer nodes and in terms of journey legs are distinctive, i.e. there can be routes where there is an overlap of transfer nodes, but no overlap of journey legs and vice-versa. Notwithstanding, often the overlap of transfer nodes and legs occurs simultaneously, leading to a high correlation between them (correlation ~ 0.7 in our case). In that case, omitting one of these factors may result in endogeniety. To understand and isolate the valuation of overlap of transfer nodes versus overlap of journey legs in more depth, we include them simultaneously. Different combinations of model formulations were estimated, and the best combination of parameters is presented (PSCL Model 5).

Estimating a model that includes both the leg-based and node-based PSC results in a slightly better model fit compared to the model with node-based PSC alone. In the combined model, the PSC parameters for overlap of transfer node are found to follow similar signs as in the individual model. However, surprisingly, the results indicate that once the overlap of transfer nodes is accounted for, a subsequent overlap of journey legs is valued negatively. This means that travellers ideally prefer routes which have an overlap of transfer nodes, i.e. decision points, rather than of journey legs per se. The latter is perceived negatively once overlap in transfer nodes is accounted for. The main argument explaining the positive perception of route overlap for transit networks is the availability of alternate travel options in case of disruptions. From that perspective, having a common transfer point but distinct journey legs between routes meets the objective. On the other hand, having the overlap in journey legs, i.e. the same travel option before or after transfer for the two overlapping routes, does not help in case of disruptions, and is hence found to reduce the attractiveness of overlapped routes, compared to completely distinct routes. This could also be interpreted as the routes with a complete overlap of legs being considered similar (rather than as two distinct routes).

Conversely, the partial overlap of journey legs (in the form of links), is found to be valued positively implying that routes with partial overlap, i.e. some but not all common links, are preferred over completely independent journey legs. This is expected as routes with some common links could provide more options for at least a part of the journey leg(s) in case of disruptions. Moreover, the routes with partial overlap are more likely to be situated in transit-dense city centre areas as opposed to outskirts. This could contribute to the preference towards them, everything else being the same, due to feelings of familiarity to the route, safety, preferred surroundings (especially for tourists in the case of Amsterdam), and perhaps a better level of service. The overall impact of route overlap depends thus on the relative values of link, leg and node overlap.

4.4. Cross-validation and sensitivity analysis

We undertook out-of-sample validation for each of our models using a cross-validation approach. Each model was re-estimated on (randomly selected) 80% of the data (305,836 observations), and the remaining 20% of the data (76,459 observations) was used as a validation dataset. This process was repeated five times (with replacement) for each model, and the average probability of chosen routes across all validation runs is presented in Table 3. In terms of prediction performance also, the node-based overlap model is found to perform better than the link or leg-based models for our data. The best performance of all is found to be for the combined node and leg-based model, marginally better than the node-based model.

Next, we perform a sensitivity analysis of our results to the choice set size. In this study, we use the 'observed' choice set as opposed to a synthetic choice set generation method. To ensure reliable estimates of travel attributes, only those routes with a minimum of 20 journeys in the half-hour time slice (over all days) were used for the analysis. However, this results in many less frequently used routes being excluded from our analysis, reducing our choice set size. In case of choice models, it is well known that choice set size and composition may greatly impact results (Prato and Bekhor 2007). For models that include route overlap (such as the PSCL), including IIA has been found to bias the results, hence it is advised to include the attractive routes only in the choice sets (Bliemer and Bovy 2008; Bovy, Bekhor, and Prato 2008). However, one could argue that routes with less than 20 observed journeys in the time period of our analysis (specially in the outskirts of the city) could still be considered 'relevant'. Hence, we test the sensitivity of our conclusions to choice set size by reducing the threshold of minimum journeys needed to include a route in our analysis. As we lower the threshold on the number of journeys, more 'less preferred' routes are included in the data set. In general, these 'less preferred' routes have more overlap between them – with 74% of OD pairs having some type of overlap with a threshold of 5 journeys, as compared to 63% for a threshold of 20 journeys.

Table 3. Cross-validation results.

Description	MNL Model	Link-based	Leg-based		Node-based	Combined
		PSCL Model 1 – links	PSCL Model 2 – number of legs	PSCL Model 3 – travel time	PSCL Model 4 – transfer node	PSCL Model 5 – travel time + transfer node
Average log-likelihood of validation data	−46,778	−46,753	−46,743	−46,753	−46,703	−46,695
Average probability of chosen route for validation data	58.89%	58.91%	58.92%	58.91%	58.96%	58.97%

Table 4. Model sensitivity to choice set size.

Description	Minimum 20 journeys	Minimum 10 journeys	Minimum 5 journeys
Number of observations	382,295	538,696	756,467
Maximum number of alternatives	4	6	7
Estimated parameters	13	13	13
Likelihood Ratio Test (compared to respective MNL)	838.6	987.2	988.6
Rho-square-bar	0.179	0.227	0.304
Parameter estimates*			
PSC – travel time	0.52	0.48	0.16
PSC – transfer nodes	−1.39	−1.10	−0.61

$*p < 0.01$ for all estimates.

Table 4 shows the model estimation statistics and path size factors for our best performing model (PSCL Model 5) for different journey thresholds. As expected, the model fit statistics improve as the number of observations increase. The PSC parameters are still found to be significant and with the same signs irrespective of the sample size, although their magnitudes decrease with reducing threshold. Amongst other parameters, the transfer-related parameters and circuity were found to be sensitive to the composition of choice set, whereas the in-vehicle and waiting times were observed to be relatively stable. Similar to Ton et al. (2018), this study has undertaken a data-driven choice-set generation approach. The impact of choice-set size when using this approach on model estimates remains a topic for further research.

4.5. Discussion

Our findings highlight the importance of transfer hubs for passenger route choice decisions. The perception of overlap is found to refer to decision points such as interchange locations. Having routes with common transfer locations that offer distinctive travel options to and from transfer locations is ideal from the perspective of travellers. Network topology analysis has demonstrated that having multiple (back-up) links increases the robustness of a transit network in case of disruptions (Jenelius and Cats 2015). Our findings imply that this also translates into increased attractiveness of overlapping routes compared to independent routes due to their contribution to journey-level robustness.

The analysis performed in this study can be extended to access stop choice. The models used in this study attempt to capture the value of robustness of routes with overlapping links or transfer nodes. More generally, this preference for more robust routes may also be reflected in case of transit travellers' access stop choice. However, in most existing models of route or access stop choice, this impact is not considered. The dataset used in this study does not provide information on door-to-door journeys, hence it is not possible to observe the preferences of travellers on their choice of origin transit stop. However, it is an interesting research direction to check for this using a data set (such as travel diary) which will allow for such an examination.

Lastly, our results show that depending on the types of modes between which the transfers are occurring, some transfers are preferred over others, in line with the findings from Garcia-Martinez et al. (2018) for intermodal transfers. Transfer penalty is expected to be a function of the transfer environment, such as level difference, number of crossings, shelters, availability of information, etc. In the case of Amsterdam, as with many other transit networks, many bus and tram lines are intended to serve as access/egress modes for the metro which is limited to major corridors. However, the higher transfer penalty for such transfers (as opposed to transfers within the bus/tram network) indicates that more attention should be given to making such intermodal transfers seamless, thereby reducing the associated transfer penalty and making such journeys more attractive.

Finally, our study is subject to three main limitations. First, crowding was not included as an attribute in any of our models. Even though the Amsterdam transit network is not very crowded, in some contexts crowding may have an impact on other attributes. Second, while the PSCL models used in this study allowed for capturing the correlation between alternatives, they do not capture the heterogeneity amongst travellers or the correlation due to the panel characteristic of the data. While the former could be explored using other model structures such as the mixed-logit, the latter is not possible given the characteristics of the dataset which does not contain panel information. Lastly, this study compared the alternate specifications of route overlap, including the new node-based overlap, using PSCL model only. It might be interesting to explore how these alternate definitions of overlap perform under other model structures used to incorporate overlap, such as the C-Logit, the PSL, or more complex ones like the Paired Combinatorial Logit or Cross Nested Logit.

5. Conclusion

The main contribution of this study is that for the first time, we provide insight into how travellers perceive different types of overlap between routes while making route choice decisions in the context of multi-modal urban transit networks. An empirical analysis by means of choice modelling was conducted for the transit network of Amsterdam using smart card data. We defined route overlap in terms of overlapping links, journey legs, and transfer nodes. Overall, incorporating route overlap resulted in a significant improvement in model fit compared to the MNL model.

Our findings support the argument of Anderson, Nielsen, and Prato (2017) and Hoogendoorn-Lanser and Bovy (2007) that having multiple options of travel enhances the attractiveness of routes that have an overlap. On the one hand, our results show that the partial overlap of routes with some links overlapping is found to be preferred by travellers, presumably because it provides more options for travel in case of disruptions. Further, travellers ideally prefer routes that have common transfer locations, but not completely overlapping journey legs. This is intuitive, as having multiple (distinct) travel options at a transfer location adds to the robustness of their route choice decision. On the other hand, completely overlapping journey legs does not add any value in terms of robustness, and is hence found to reduce the attractiveness of overlapped routes, compared to distinct ones.

The majority of studies in the literature that consider transit route overlap measure it in terms of path only. In this study, not considering the overlap in terms of transfer nodes led to the contrasting conclusion of a positive valuation of overlapping legs by travellers. Hence, a key take-away from our results is that for transit route choice, it is important to define overlap in terms of both path and transfer nodes.

Overall, this study contributed to advancing the understanding of travellers' perception of overlap during transit route choice. It also added to the limited studies that empirically analyse route choice behaviour for large-scale multi-modal transit networks using smart card data. The results show differences in perceptions of travel times and transfer penalties by mode(s) used. The trade-off values between different route attributes obtained in this study also provide behavioural insights to transit planners and policy-makers. Moreover, the methodology proposed to incorporate route overlap could be adopted for other transit networks to improve the performance and accuracy of route choice models, leading to better predictions.

Acknowledgments

This research was funded by the municipality of Amsterdam, Vervoerregio Amsterdam and the AMS Institute. We thank GVB for providing data for this research. We also sincerely thank the three anonymous reviewers whose comments and suggestions helped improve this manuscript substantially.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Malvika Dixit http://orcid.org/0000-0003-4483-7438
Oded Cats http://orcid.org/0000-0002-4506-0459
Niels van Oort http://orcid.org/0000-0002-4519-2013
Serge Hoogendoorn http://orcid.org/0000-0002-1579-1939

References

Anderson, Marie Karen, Otto Anker Nielsen, and Carlo Giacomo Prato. 2017. "Multimodal Route Choice Models of Public Transport Passengers in the Greater Copenhagen Area." *EURO Journal on Transportation and Logistics. Springer Berlin Heidelberg*, 221–245. doi:10.1007/s13676-014-0063-3.

Ben-Akiva, Moshe, and Michel Bierlaire. 1999. "Discrete Choice Methods and Their Applications to Short Term Travel Decisions." In *Handbook of Transportation Science*, edited by R. W. Hall. Boston, MA: Springer. doi:10.1007/978-1-4615-5203-1_2.

Bierlaire, Michel. 2020. "A Short Introduction to PandasBiogeme. Report TRANSP-OR 200605." Lausanne, Switzerland.

Bliemer, Michiel C.J., and Piet H.L. Bovy. 2008. "Impact of Route Choice Set on Route Choice Probabilities." *Transportation Research Record* 2076: 10–19. doi:10.3141/2076-02.

Bovy, Piet H.L., Shlomo Bekhor, and Carlo Giacomo Prato. 2008. "The Factor of Revisited Path Size: Alternative Derivation." *Transportation Research Record* 2076 (1): 132–140. doi:10.3141/2076-15.

Brand, Judith, Serge Hoogendoorn, Niels Van Oort, and Bart Schalkwijk. 2017. "Modelling Multimodal Transit Networks Integration of Bus Networks with Walking and Cycling." In 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2017 - Proceedings. doi:10.1109/MTITS.2017.8005612.

Bunschoten, Tim, Eric Molin, and Rob van Nes. 2013. "Tram or Bus: Does the Tram Bonus Exist?" In European Transport Conference 2013. doi:10.1002/kin.10070.

Cascetta, Ennio, Agostino Nuzzolo, Francesco Russo, and Antonino Vitetta. 1996. "A Modified Logit Route Choice Model Overcoming Path Overlapping Problems: Specification and Some Calibration Results for Interurban Networks." In Transportation and Traffic Theory: Proc., 13th International Symposium on Transportation and Traffic Theory. Lyon, France.

Dixit, Malvika, Ties Brands, Niels van Oort, Oded Cats, and Serge Hoogendoorn. 2019. "Passenger Travel Time Reliability for Multimodal Public Transport Journeys." *Transportation Research Record* 2673 (2): 149–160. doi:10.1177/0361198118825459.

Frejinger, E., and M. Bierlaire. 2007. "Capturing Correlation with Subnetworks in Route Choice Models." *Transportation Research Part B: Methodological* 41 (3): 363–378. doi:10.1016/j.trb.2006.06.003.

Garcia-Martinez, Andres, Rocio Cascajo, Sergio R. Jara-Diaz, Subeh Chowdhury, and Andres Monzon. 2018. "Transfer Penalties in Multimodal Public Transport Networks." *Transportation Research Part A: Policy and Practice* 114 (August): 52–66. Pergamon. doi:10.1016/J.TRA.2018.01.016.

Gordon, Jason, Harilaos Koutsopoulos, Nigel Wilson, and John Attanucci. 2013. "Automated Inference of Linked Transit Journeys in London Using Fare-Transaction and Vehicle Location Data." *Transportation Research Record* 2343: 17–24. doi:10.3141/2343-03.

Hoogendoorn-Lanser, Sascha, and Piet Bovy. 2007. "Modeling Overlap in Multimodal Route Choice by Including Trip Part-Specific Path Size Factors." *Transportation Research Record* 2003 (1): 74–83. doi:10.3141/2003-10.

Hoogendoorn-Lanser, Sascha, Rob van Nes, and Piet Bovy. 2005. "Path Size Modeling in Multimodal Route Choice Analysis." *Transportation Research Record* 1921 (1): 27–34. doi:10.3141/1921-04.

Jánošíkova, L'udmila, Jiří Slavík, and Michal Kohání. 2014. "Estimation of a Route Choice Model for Urban Public Transport Using Smart Card Data." *Transportation Planning and Technology* 37 (7): 638–648. doi:10.1080/03081060.2014.935570.

Jenelius, Erik, and Oded Cats. 2015. "The Value of New Public Transport Links for Network Robustness and Redundancy." *Transportmetrica A: Transport Science* 11 (9): 819–835. doi:10.1080/23249935.2015.1087232.

Kim, Ikki, Hyoung-Chul Kim, Dong-Jeong Seo, and Jung In Kim. 2019. "Calibration of a Transit Route Choice Model Using Revealed Population Data of Smartcard in a Multimodal Transit Network." *Transportation*, doi:10.1007/s11116-019-10008-8.

Oort, Niels van, Ties Brands, and Erik de Romph. 2015. "Short-Term Prediction of Ridership on Public Transport with Smart Card Data." *Transportation Research Record* 2535: 105–111. doi:10.3141/2535-12.

Oort, Niels van, and Rob van Nes. 2009. "Regularity Analysis for Optimizing Urban Transit Network Design." *Public Transport* 1: 155–168. doi:10.1007/s12469-009-0012-y.

Prato, Carlo Giacomo. 2009. "Route Choice Modeling: Past, Present and Future Research Directions." *Journal of Choice Modelling* 2 (1): 65–100. Elsevier. doi:10.1016/S1755-5345(13)70005-8.

Prato, Carlo Giacomo, and Shlomo Bekhor. 2007. "Modeling Route Choice Behavior: How Relevant Is the Composition of Choice Set?" *Transportation Research Record* 2003 (1): 64–73. doi:10.3141/2003-09.

Ramming, Michael Scott. 2002. "Network Knowledge and Route Choice." Ph .D. Thesis, Massachusetts Institute of Technology.

Tan, Rui, Muhammad Adnan, Der Horng Lee, and Moshe E. Ben-Akiva. 2015. "New Path Size Formulation in Path Size Logit for Route Choice Modeling in Public Transport Networks." *Transportation Research Record* 2538: 11–18. doi:10.3141/2538-02.

Ton, Danique, Dorine Duives, Oded Cats, and Serge Hoogendoorn. 2018. "Evaluating a Data-Driven Approach for Choice Set Identification Using GPS Bicycle Route Choice Data from Amsterdam." *Travel Behaviour and Society* 13 (October): 105–117. doi:10.1016/j.tbs.2018.07.001.

Yap, M. D., O. Cats, N. Van Oort, and S. P. Hoogendoorn. 2017. "A Robust Transfer Inference Algorithm for Public Transport Journeys During Disruptions." *Transportation Research Procedia* 27 (September): 1042–1049. doi:10.1016/j.trpro.2017.12.099.

Yap, Menno, Oded Cats, and Bart van Arem. 2020. "Crowding Valuation in Urban Tram and Bus Transportation Based on Smart Card Data." *Transportmetrica A: Transport Science* 16 (1): 23–42. Taylor and Francis Ltd. doi:10.1080/23249935.2018.1537319.