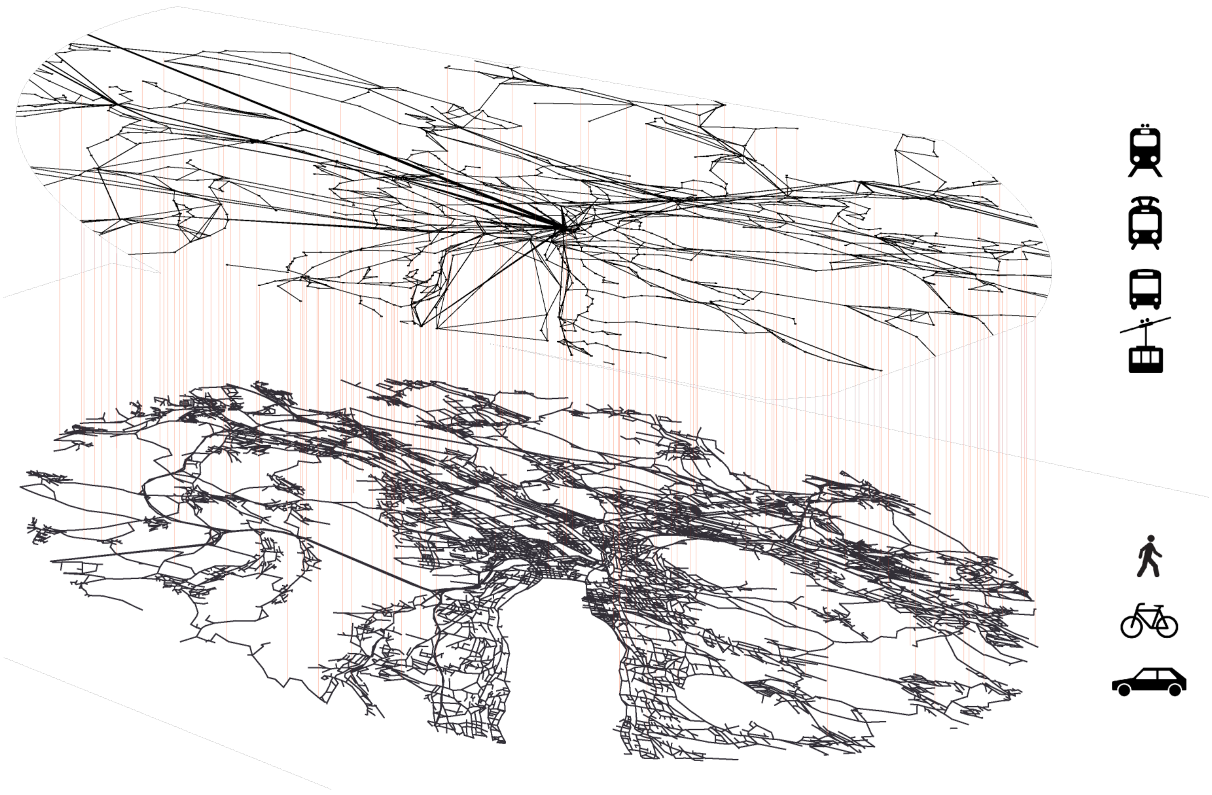


A recursive logit multimodal route choice model

Lucas Meyer de Freitas

Master thesis

Institute for Transport Planning and Systems

January 2018

IVT Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

ETH Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Acknowledgements

First, I would like to thank my parents, Rita and Caco, for their motivation and incentive since the beginning of my graduation at UFRJ. For their support and mentoring during this thesis I thank my supervisors, Prof. Kay W. Axhausen and Henrik Becker. For their previous work on routers used in this thesis and their support I would also like to thank Milos Balac, Sebastian Hörl and Joseph Molloy. The help in discrete choice modelling issues from Felix Becker and Basil Schmid were also always helpful. I would also like to thank Maëlle Zimmerman for providing the recursive logit estimation code and for her help with questions regarding this work. Felix and Igor's companionship during late working hours at the IVT made it easier to work on this project. To all my friends and study colleagues who have accompanied me on this journey, I would also like to express my gratitude, especially to Katherine for her comments on this work.

Table of contents

List of tables

List of figures

Glossary

API – Application programming interface

ARE – Federal Office for Spatial Development (Bundesamt für Raumentwicklung)

BFGS – Broyden-Fletcher-Goldfarb-Shanno algorithm

BFS – Federal Statistical Office (Bundesamt für Statistik)

EBA – Elimination by aspects

GEV – Generalized extreme value

GTFS – Generalized Transit Feed Specification

IIA – Independence of irrelevant alternatives

MNL – Multinomial Logit Model

NL – Nested Logit Model

OSM – Open Street Maps

R5 – Conveyal R5 router

RL – Recursive logit model

RP – Revealed preference

SP – Stated preference

VoT – Value of travel time savings

Masterthesis MSc Spatial Development and Infrastructure Systems

A recursive logit multimodal route choice model

Lucas Meyer de Freitas
mlucas@student.ethz.ch

March 2018

Abstract

This thesis shows the first implementation of a multimodal recursive logit route choice model. A recursive logit model is a link choice model which corresponds to a path choice model with an infinite choice set, therefore eliminating the need to generate a route choice set. A methodology for generating a static transit network based on GTFS data is introduced. A multimodal supernetwork is constructed by connecting this network to a street network. Different multimodal transit route choice models with walk as access and egress mode are estimated using data from the last Swiss mobility Microcensus in Zurich. First model results show that the recursive logit framework is a promising tool for evaluating multimodal trip making. Nevertheless, some parameter estimates indicate that further model specifications should be tested to evaluate the correlation between different modes

Keywords

Recursive logit, multimodal, GTFS, route choice, mode choice

Preferred citation style

Meyer de Freitas, L. (2018) A recursive logit multimodal route choice model. Master thesis. Institute for Transport Planning and Systems. ETH Zurich, Zurich.

Masterarbeit MSc Raumentwicklung und Infrastruktursysteme

A recursive logit multimodal route choice model

Lucas Meyer de Freitas
mlucas@student.ethz.ch

March 2018

Kurzfassung

Diese Arbeit zeigt die erste Implementierung eines multimodalen Recursive Logit Routenwahlmodell. Solch einen Modell ist ein Kantenwahlmodell, das äquivalent zu einem Routenwahlmodell mit unendlich grossen Alternativen ist. Dieser Ansatz beseitigt die Notwendigkeit nach der Generierung eines Routenwahlsatzes für die Modellschätzung. Ein statisches multimodales öV Netzwerk wird anhand GTFS Daten konstruiert und dann mit einen Strassennetwerk verbunden um ein multimodales Supernetzwerk zu erstellen. Verschiedene multimodale Routenwahlmodelle werden anhand der Daten des Mikrozensus Verkehr 2015 für öV-Routen in Zürich geschätzt. Erste Modellergebnisse zeigen, dass der Recursive Logit Ansatz ein vielversprechendes Werkzeug für die Analyse multimodaler Routen ist. Die Ergebnisse zeigen auch, dass weitere Modellformulierungen geprüft werden müssen, um mögliche Korrelationen zwischen Verkehrsmittel zu berücksichtigen.

Schlagworte

Recursive logit, multimodal, GTFS, Routenwahl, Verkehrsmittelwahl

Zitierungsvorschlag

Meyer de Freitas, L. (2018) A recursive logit multimodal route choice model. Master thesis. Institute for Transport Planning and Systems. ETH Zurich, Zurich

1 Introduction

The Swiss transport system faces many challenges in the years to come. As in other West-European countries, the mode share of the automobile has increased over the 20^{th} century (Newman and Kenworthy, 1999). Data from the latest mobility Microcensus shows that only 13% of Swiss households did not possess a car. Concurrently, the mode share of the automobile lies at 53.7%. For transit, the share is 12.7% and for slow modes, it is 32.1% (other modes account for 1.5%). Given that a series of causes led to this present situation, one can highlight the cheap availability of gasoline on the supply side and a sprawling of urban structure impacting the demand side. These trends are expected to turn, or at least to decrease in intensity, in the decades to come. While oil prices are expected to revert their downward tendency, a rapidly growing population challenges planners and officials to look for brown field instead of green field development options, creating denser agglomerations and a higher pressure on existing transport infrastructure. Such developments pose a challenge to the transport system, which must adapt to a new reality.

To propose solutions, a good understanding of the mobility behaviour of the population is needed. This entails not only the choice of mode but also the choice of route. While mode choice modelling not only has a large body of literature but is also commonly employed by practitioners, route choice modelling is a more infant field. A metadata search on web of science shows there are more than double the amount of publications in mode choice modelling than in route choice modelling (Web of Science, 2018). This infancy is due to two factors: route choice modelling is a much more complex task, and, until recently, it was too expensive to collect the data needed for estimating such models. The goal of this thesis is to tackle the first problem by applying a recursive logit model to a multimodal network. A multimodal choice framework increases the realism of the choice alternatives and allows for evaluating trade-offs among different modes. The increasing importance of bike and car sharing systems, the autonomous driving revolution and new transport services such as MaaS (mobility as a service) and other on-demand services are necessary steps towards an increased multimodality and to a more sustainable transportation system. Multimodal route choice models provide a methodology to evaluate the behaviour of travellers in multimodal trips. This thesis' goal is to implement and evaluate the use of the recursive logit framework for such a purpose.

At this point, it is important to delimit the scope of this thesis. This thesis is focused on showing a methodology to build a multimodal recursive logit model for land travel on urban networks. Although transport mode parameters are used in model estimation making it, therefore, possible to distinguish modal preferences of travellers, the model constructed here is not a simultaneous route and mode choice model, but a route choice model with different mode options. This thesis also does not review or apply methods for the estimation of route choice models based on stated preference data. Only models based on revealed preference data are reviewed and estimated. Furthermore, the work presented here does not extend to the application of models for travel demand prediction, limiting itself to model estimation.

Firstly, this thesis examines current methods in mode choice, route choice and multimodal route choice models (chapter 2). Subsequently, using revealed preference data from the latest Swiss Microcensus, the thesis will describe the steps taken for constructing and estimating a multimodal network (chapter 3). Finally, this report concludes by evaluating the results and discussing future developments in multimodal route choice modelling (chapters 4 and 5).

2 Choice modelling

2.1 Discrete choice

Discrete choice models are based on the random utility maximization theory, formalized by McFadden (1975), which aims at describing the choice process of an alternative among a set of alternatives. According to Train (2009), the following assumptions are made in a discrete choice framework:

- The decision-maker i takes mutually exclusive alternatives m_i into account, making up a choice set I^i;
- The decision-maker associates each alternative j with a perceived utility U_j^i and chooses the alternative which maximizes this utility;
- The utility associated with each alternative depends on the number of measurable attributes of each alternative and of the decision-maker;
- The decision-maker chooses only one alternative.

The perceived utility of an alternative for the decision maker is a function of a vector of attributes X_j^i and:

$$U_j^i = U^i(X_j^i) \tag{1}$$

This utility is the econometrical interpretation of what the psychologist Thurstone (1927) recognized as differences in stimulus, or excitement, generated by the properties of an alternative. Human stimulus and action, however, are not only driven by observable and directly measurable factors. At the individual level, humans are not expected to act rationally or deterministically when facing decisions (Rabin, 1998). As a result, discrete choice models are not deterministic, but stochastic models. Another reason for the stochasticity of discrete choice models is to account for the influence of non-observed factors. For this reason, an error term is part to the utility function. According to Cascetta (2009), this error term has its source on measurement errors and imperfect information regarding the alternatives. From the assumptions on the statistical distribution of the error term, different model formulations have been developed, the main ones being logit (extreme value distribution) and probit (normal distribution). These assumptions on the distributions of the error term are a necessary restraint posed on the unknown influence factors for the utility so that what for psychologists as Thurstone (1927) are an ordinal measure of welfare can be translated into a cardinal measure and therefore quantitatively modelled (Batley, 2008).

$$U_j^i = V_j^i + \varepsilon_j^i \tag{2}$$

While such models are not effective at explaining individual behavior, when data is collected from a statistically significant sample of a population, one can assume the model will correctly represent the aggregate behavior of the group (Axhausen, 2017). The most commonly used and simplest formulation of a choice model based on discrete choice is the multinomial logit model (MNL). The error is assumed to be Gumbel distributed and the alternatives are assumed to be independently and identically distributed (IID). The resulting probability of choosing an alternative P_j, given a choice set I is:

$$P_j = \frac{e^{V_j}}{\sum_{i=1}^{I} e^{V_i}} \tag{3}$$

Other modelling frameworks always follow the same principle of evaluating the exponential of the observed part of utility against all other alternatives. However, different assumptions on the error term can lead to other model formulations. The generalized extreme value distribution is also widely popular because it relaxes the IIA property, which is not always a realistic behavioral assumption.

2.2 Mode choice

The first systematic econometric studies for transportation planning began in the early 1970's in the United States. Researches were primarily interested in the question of modal choice and in willingness to pay measures, particularly the valuation of travel times. The raise in demand for such studies was driven by a changing paradigm in transport planning the US and the UK, from a highway oriented planning, towards a transport system management and public transport investments (Boyce & Williams, 2016). This shift called for models which were more disaggregated than zone-based regression models and models that were useful for evaluating policies. The models formalized by McFadden (1975) were put in practice at the same time in the evaluation of the Bay Area Rapid Transit System (BART) in San Francisco. The validation of the new approach came when the forecasts made with MNL mode choice models in 1972 for 1975 were much more accurate than official predictions based on aggregate models (Boyce & Williams, 2016).

Since then it has been common practice both in the US and in Europe to make use of models of mode choice models for predicting future mode shares, the guidance of transport policies and investments in transport infrastructure. The MNL structure is still widely used due to its simplicity and behavioural soundness (Cascetta, 2009). The nested logit model is also a widely-employed model because it relaxes the IIA assumption of the MNL. Through this, it can account for correlation between similar modes as evidenced by the famous blue bus/ red bus paradox (Train, 2009).

More recent developments have tried to extend the explanatory power of discrete choice models beyond mode specific attributes. Taste and preference heterogeneity are important attributes which play a role in a choice context. MNL and NL models do not explicitly account for such heterogeneities, simply assuming that these are part of the error term. Mixed-logit models do that by allowing for different parameters in the estimation function for each individual (Train, 2009). A distribution function is used to mimic the distribution of coefficients in the population. Random draws are taken to mimic this distribution. Besides random taste variation, mixed-logit models can also account for unrestricted substitution patterns and correlation of unobserved factors over time (Train, 2009). Mixed logit models are very cumbersome to be used in practice though. Fiebig et al. (2010) show that S-MNL or scale heterogeneity multinomial logit model can capture group heterogeneities better than the mixed logit model while having a much simpler formulation. In this model, the scale heterogeneity is represented by a scale multiplier which multiplies the vector of parameters. Scale effects are particularly significant when the observations in the model being estimated stem from different sources (Hess et al., 2007). Keane (2006) proposed combining the S-MNL model with the mixed logit model, forming the generalized multinomial logit model (G-MNL). G-MNL models are more of a class of models, since their formulations can vary depending on how the scale parameter of the S-MNL and the mixing component of the mixed logit are combined (Wittink, 2011).

The advances made possible by new model formulations above do account for taste heterogeneity but at they do not propose methods to include decision-maker specific parameters in the model. Latent models can account for this by adding attitudinal variables to the utility function of the model which are selected through factor analysis. The inclusion of such latent classes in the utility function specification are expected to increase the explanatory power of models by taking more account for the psychological factors behind the choice process (Ben-Akiva et al., 2002).

Further innovations beyond the standard MNL formulation are in the formulation of the utility function itself. Mackie et al. (2003) show that travel cost should be made sensible to the income of the decision-maker. This is done through the multiplication of cost variables with the relative income of an individual in comparison to the population, coined the Mackie interaction term. Models estimated on stated preference data from the Swiss Microcensus 2015 (ARE, 2017) widely employ such interaction terms, for travel cost and travel distance variables, therefore making the valuation of an attribute relative to individual characteristics.

2.3 RP and SP data

The quality of a model is at best as good as the data used to estimate it. Data can be collected in different ways and in different settings. These different settings are the object of this chapter. The behaviour of consumers in a market can be observed principally in two ways. Their choices can be either directly observed or consumers can be asked about hypothetical decisions in a fictive market. The first data type is called revealed preference (RP) while the second is called stated preference (SP). RP data has the advantage of being an observation of actual behaviour in a real environment. Still, not all econometric questions can be answered with RP data. Choices in non-existing markets and future demands might be difficult to evaluate with RP techniques. Furthermore, Louviere et al. (2000) states that explanatory variables stemming from RP data might have little variability and are therefore highly collinear. These facts might pose severe numerical issues and restrictions when estimating models. The reason for that is often, that consumers are not always aware of all the alternatives available for their choices. This is particularly true in the case of route choice as discussed in chapter 2.4. A very important reason for the use of stated preference methods in the transportation field is the fact that transport services are often public goods and therefore not paid for directly at each use by a traveller (this is not true for every country and every transport system though). When making investments decisions or policy comparisons, policy makers need to compare alternatives in the form of a cost-benefit analysis and need valuation parameters for doing so.

SP data provides a method to infer such values by asking the respondent to compare alternatives and make trade-offs between them. In this kind of data, the researcher can control the correlation structure of the variables used, eliminating numerical estimation issues while at the same time focus on trade-offs between relevant attributes (Boyce & Williams, 2016). This higher control of the researcher over the choice situation also comes with the downside of the framing of the choice situation being decisive for the choices and trade-offs made by the decision makers as exemplified by Beck et al. (2017) for the case of travel time savings. The initial suspicion of researchers and economists towards the hypothetical SP data was slowly overcome by the realisation that SP data could enhance discrete choice estimations when combined with RP data (Hensher, 1994). The Swiss value of time study (ARE, 2017) makes use of such a methodology, although SP data is predominant, representing 90% of all observations. Cost also plays an important role in the predominance of SP observations, since collecting RP data with the required variance needed for discrete choice estimation can become very expensive.

In the end of the day, both data sources are needed. RP data gives an insight into actual behaviour and preferences which is invaluable for understanding human decision-making. At the same time the clear trade-offs provided by SP methods complement RP data to make researchers able to consistently estimate willingness to pay in markets where public goods are mostly available or where the market is non-existent. In route choice modelling, some studies have used SP data to evaluate preferences (Vrtic, 2010, Fosgerau et al., 2007). Route choice studies based on SP can be used to evaluate policies but are to be viewed with caution when predicting future behaviour. In the case of route choice, even RP data face several problems, since in most available methods for estimation, parameters are biased by the definition of the choice set. The advantage of a recursive logit model relies precisely on the fact that this bias is removed by excluding the necessity to construct a virtual choice set to estimate a route choice model.

2.4 Route choice

Most choice situations for which discrete choice modelling is employed are situations in which there is a finite number of alternatives. This condition, for example, applies to mode choice modelling. The route choice process is a more complex one, especially for modelling. This complexity arises not only from practical modelling complications but also from the underlying mental processes that individuals go through when selecting a route in a network. The challenges of route choice models are summarised by Bovy (2009):

1. The actual choice set is vast and mostly unknown by the decision maker;
2. Large subset of feasible and/or attractive routes;
3. High heterogeneity of route types (for individual modes) and service types (for transit);
4. Multidimensionality of choices (different modes in a multimodal network and time dimension in a network with loadings);
5. Complex loading patterns in the road network (queue formation, congestion, bottlenecks, spillovers) which play a role in the choice process but are difficult to incorporate in the models;

6. Strong individual preferences and behaviour due to differences in experience and cognitive processes.

Two main issues can be highlighted in the route choice modelling process: insufficient observed parameters which play a role in the decision-making process, and a large mismatch between network knowledge of the individual and the actual network options. While objective aspects such as travel time play an important role for route choice, subjective aspects also have palpable weight but are difficult to measure (Hoongedoorn-Lanser, 2005). Aspects such as felt safety, aesthetics, and generic feelings towards a locality belong to this category. The great variation among individuals when evaluating such aspects only adds to the complexity. Furthermore, another important factor is the different evaluation of objective and subjective route attributes depending on the trip purpose.

Most importantly, the choice set considered by the individual as possible routes is small and variable. Empirical findings show that 95% of travellers do not take more than 5 routes into account in their choice set (Stern and Leiser, 1988). Hoongedoorn-Lanser (2005) reports that most travellers do not even mention route alternatives other than the chosen one, and when they do, the route of the main mode does not change. Often individuals base their choice on heuristics such as minimization of travel time, or the minimization of travel cost. Further, these heuristics are based on incomplete information and on the experience of the traveller on the network. In fact, route choice is a trial-and-error process determined by the information on the routes accumulated over time (Richardson, 1982). The route choice set of an individual which has lived in a city for a month is likely to be different from this same individual's choice set after a year in the same place.

Individuals, therefore, read and navigate differently through a network, partly as a result of the different knowledge they hold of the networks they navigate on. While there are different route choice processes, which are subsequently discussed in this chapter, it is important to highlight that individuals usually do not realise all of the alternatives they have. This is also due to heuristics in their choice process. A person that does not enjoy biking through congested streets will not consider biking through them, although from the modeller's perspective it is part of the choice set. Individual preferences and heuristics play an important role for the traveller when choosing a route. This results in a difficulty to observe all the variables which play a role in the decision-making process and to an exceedingly small number of actual routes being considered for the choice.

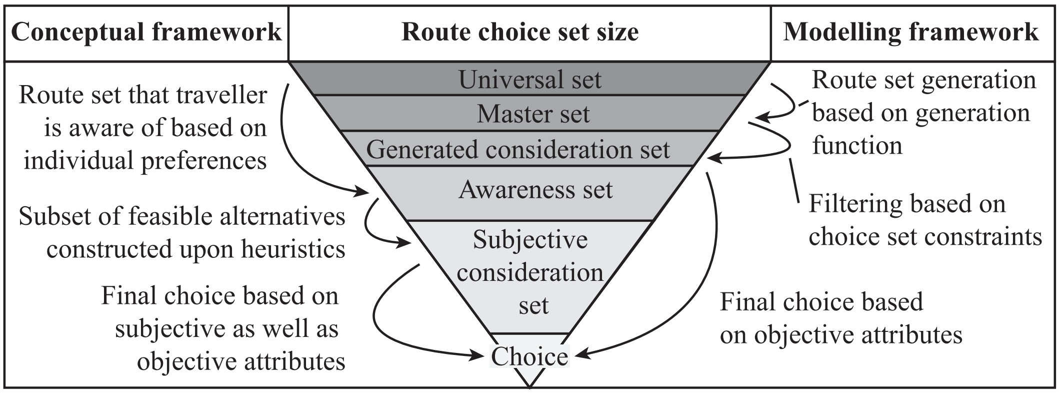
The way in which routes are chosen by travellers can also vary. The choice can be sequential, i.e. edge after edge (Marzano & Papola, 2004), simultaneous, when the choice is made once for a trip, or strategic, which represents the adaptive behaviour of a traveller based on network loading conditions (Gao & Chabini, 2006). Jansen & Den Adel (1985) reported that 75% of travellers in the Netherlands followed a simultaneous decision-making process, while the rest followed sequential procedures. The simultaneous process can follow a strategic or a hierarchical behaviour, the latter meaning that a route is chosen based on habitus of always choosing this route. This is mostly common for commuting trips.

These different choice procedures are split into two parts. First, individual constraints and eliminatory aspects (such as the exclusion of routes through a certain city district) define the choice set in a non-compensatory process. This process is followed by a compensatory process in which trade-offs are made among interest attributes. The literature is unanimous in the fact that the division of the route choice process into these two separate processes is the best approach to human decision making and is, therefore, the most employed one in the estimation of route choice models (Ben-Akiva et al., 2004, Bovy, 2009, Papinski et al., 2009, Bekhor et al., 2006a). This division is applied to most models used for estimation, with exception of the recursive logit (Fosgerau et al., 2013) and the nested recursive logit (Mai et al., 2015a) approaches. Below, emphasis will be made firstly on the generation of choice sets, and then on the route choice models.

The practical modelling of route choice models also faces several difficulties. These arise from the choice set construction. As Anderson et al. (2017) posit, it is a non-trivial challenge to generate a choice set and model route choice behaviour while accounting for similarities across alternatives and heterogeneity across travellers. Traditional route choice models take a series of route-based variables as an input. From a numeric perspective, these alternatives cannot be too correlated, else it will be impossible to estimate a stochastic model. Modellers have to artificially produce a large set of alternatives (most studies use between 30 and 250 alternatives) in order to estimate model parameters, while travellers usually only evaluate between one and four alternatives (Hoongedoorn-Lanser, 2005). This leads traditional route choice models to be extremely sensitive to the choice set generation. The algorithm used for the creation of alternative routes as well as the size of the choice set significantly changes the model's results.

A comparison between the modeller's and the individual's route choice set process is presented in Figure 1. The universal choice set is an infinite choice set, containing all the possible routes from A to B. The master set is the choice set after the application of a choice set generation algorithm on an origin-destination pair on a network. Since such algorithms can produce a prohibitively large set or unfeasible alternatives, to obtain the consideration set, feasibility and logical constraints are generally applied on such master sets. On the individual's side, the awareness set is smaller than the consideration set. This set consists of the alternatives that an individual is aware of as possible and is dependent on his or her network knowledge. A preference and heuristics filter is applied, which then generates the routes de facto considered by the individual for the route choice. The differences between the modelling and the conceptual framework in Figure 1 reflect the differences between network knowledge at the individual level and the actual supply of transport infrastructure (Hoongedoorn-Lanser, 2005).

Figure 1 Generic scheme of the route choice process



Adapted from Bovy (2009).

2.4.1 Choice set generation

The generation of choice sets is the first step for the estimation of a route choice model. From a behavioural perspective, the generation of a choice set is a non-compensatory process in which the selection of routes belonging to it is made through a process named 'elimination by aspects' (EBA) (Tversky, 1972). Deterministic or stochastic processes can be used by researchers to mimic this mental process (Ramming, 2002). The suitability of choice set generation algorithms and route choice models here is only valid for model estimation, since other choice set generation algorithms might be more suited for prediction than for estimation for the same choice set (Prato, 2009). Generally, prediction requires more relevant routes in a choice set than model estimation. According to Ben-Akiva & Lerman (1985) even small, well-sampled choice sets can produce satisfactory results. This contrasts with findings by Prato (2006), Prato & Bekhor (2007), Bliemer & Bovy (2008) and Hoongedoorn-Lanser (2005), which find that choice set size and composition do affect model estimation.

As already stated, route choice modelling differs substantially from other choices because most alternatives are not identifiable. In an urban context, there might exist hundreds of feasible routes between an OD pair (Prato, 2009). I highlight the word "feasible" here as, in theory, there is an endless possibility of possible routes for an origin-destination pair. This set of endless possibilities is called the 'universal set' and refers to the entire network. Rieser-Schlüssler et al. (2013) define a feasible route as a continuous, loop-less, and low-cost route. This is logical from a behavioural standpoint since travellers do not take unnecessary actions (Hoongedoorn-Lanser, 2005). Choice set construction usually follows three procedures (Figure 1). The construction of a master set corresponds to the process of constructing a network model of the real network which includes all possible routes between an origin and destination. After this step, algorithms for choice set generation are used to construct a set of feasible sets between the origin and destination. This modelling framework mimics the conceptual framework which is based on an individual's decision-making regarding route choice. The filtering conducted by the algorithm constructs a route choice set with the goal of removing overlapping routes, illogical route structures, non-observation of an alternative to the observed individual (exclusion of bike routes if a bike is not available for example) and utility levels below a certain threshold (Bovy, 2009). One main contradiction between observed behaviour and reality is that, while alternatives in a choice set should not present dominance, this is often the case in reality, i.e. one route is always much better in all parameters than all the others.

Deterministic choice set generation methods

According to Prato (2009), deterministic methods are the most common ones in route choice set generation. The simplicity of these methods, in addition to the ease with which their methodology is comprehended and implemented, make them more popular than stochastic methods. The drawback of these methods is, compared to stochastic methods, they are inferior in their representation of chosen routes (Vrtic, 2003). This is expected as the discussion above shows that the route choice set generation framework is both a highly ambiguous one - with widely varying preferences between individuals - and not a deterministic one. The main deterministic route choice set generation methods are presented below.

Shortest path methods

These methods are an extension of known algorithms to find the shortest path in a network such as the Dijkstra algorithm (Dijkstra, 1959). These methods are called K-shortest path methods and the analyst generates K routes between an OD pair within a minimum and a maximum of an attribute (eg. travel time). After the generation of the shortest route on a network, the procedure is to either increase the impedance on edges in the shortest route so that *K* new routes can be generated, or to remove edges in the shortest route, so that detours are forced (Bekhor et al., 2006). The former method is also called edge penalty method, while the latter is named edge elimination method. According to Prato (2009), these methods run the risk of often generating unrealistic and cyclical routes as well as too similar routes, all properties which are not desired in a route choice set generation. Prato & Bekhor (2006) apply a modified version of the edge elimination method with a branching algorithm which has a higher success rate. The edge penalty approach also has different implementation frameworks, with varying success (Bekhor et al., 2006, De la Barra et al., 1993, Park & Rilett, 1997). Generally, these are not able to achieve the same success rates as stochastic methods (Prato, 2009).

Labelling methods

The labelling method presented by Ben-Akiva et al. (1984) considers the variations in traveller preferences. It acknowledges that travellers have different preferences (Prato, 2009). Multiple edge attributes are evaluated in this method, to formulate a generalized cost function of a edge, based on different labels (attributes) such as travel time, travel cost, and travel distance). Labelling does not appear to reproduce chosen routes with much fidelity as well. The coverage of chosen routes is highly dependent on the definition of labels by the analyst, which conversely is dependent on prior knowledge of traveller's preferences (Prato, 2009). Bekhor et al. (2006), find that the combination of labelling methods with simulation provided the best representation of chosen routes among several methods.

Stochastic choice set generation methods

Stochastic methods apply random utility theory to the generation of route choice sets. One main difference to deterministic choice set generation methods is the behavioural assumption regarding how route choice sets are constructed. While deterministic methods assume a sequential decision-making procedure (i.e. edge to edge), stochastic methods assume a simultaneous decision-making, in which all OD pairs are evaluated simultaneously (Prato, 2009). The alternatives are constructed from a master choice set (*G*). The methods presented here account for the construction of a choice set *C*, on which the traveller chooses a route *i* (Equation 4). Stochastic choice set generation methods treat an individual choice set as a latent class, since it is impossible to derive unchosen alternatives from observations and individuals never recognize all possible feasible alternatives in stated preference surveys (Ramming, 2002).

$$P_n(i) = \sum_{C \in G} P_n(i|C) P_n(C) \tag{4}$$

Simulation methods

Simulation methods produce feasible routes by constructing edge impedances from probability distributions (Bekhor et al., 2006). Ramming (2002) proposes a method that generates different edge costs based on generalized cost probability distributions. In this method, iterations are performed in an all-or-nothing assignment on the network. After each iteration, the shortest route is stored and after a fixed number of iterations, a set of routes is obtained. Different authors have studied variations of this method with different draws and different probability distributions (eg. Bierlaire & Frejinger, 2005, Nielsen, 2000). Simulation methods have the advantage of rendering generated routes not only feasible but attractive (Prato, 2009). When using these methods, modellers are confronted with the issue that while a normal distribution is the best theoretical representation of actual costs, it is also not the most suited one for computational modelling. The models implemented for simulation methods also require a correction term to account for inequalities between alternatives (Prato, 2009). This is necessary as in route choice models, due to the countless alternatives in the choice set, it is impossible to include alternative specific constants (Frejinger et al., 2009). Rieser-Schüssler et al. (2013) and Dugge (2006) find that for large networks with small edge sizes, common characteristics to real navigation networks, it can be extremely time-consuming to generate a set with a significant number of edges.

Doubly stochastic

Another approach is the doubly stochastic generation function (Nielsen, 2000). Its main assumption is that not only are travellers mistaken in perceiving costs of a route, they also have a different perception. Therefore, a random term is added to the generation function to describe taste heterogeneity for the traveller. This method has the advantage of producing a higher taste heterogeneity within the generated alternatives.

Monte Carlo methods

Monte Carlo simulation methods have also been used to construct route choice sets but are far less present in the literature than the methods above. First, the attributes of the edges on the network are randomized and then the Monte Carlo algorithm searches the shortest path in this randomized network (Zantema et al., 2007). This method can also be coupled with edge labelling to construct a search algorithm that is based on a variation of the search criteria. The advantage of this method is that it finds a route choice set with a high spatial variation of routes in the choice set.

Constrained enumeration methods

Constrained programming is the programming technique used to solve optimization problems through the construction, and then solution, of a search tree of possible solutions (Soto et al., 2015). Under this modelling framework, it is assumed that travellers make decisions based on other aspects than cost minimization (Prato, 2009). Prato & Bekhor (2006) propose a branch and bound algorithm with a good representation of actual routes. The branch and bound algorithm constructs branches at every new node reached. These branches are then bounded depending on the constraints set by the modeller. In the study by Prato & Bekhor (2006) the constraints were: removal of routes with a high degree of edge overlap; the removal of routes with too many U-turns; routes with too long travel times. Rieser-Schüssler et al. (2013) find that the algorithm, although providing good results, takes too much time to be computed for routes with more than 30 edges. Hoongedoorn-Lanser (2005) also employs a modified version of this algorithm for choice set generation in transit networks.

Probabilistic methods

Probabilistic methods are characterized by the addition of a generation probability to each route on a network. The Implicit Availability/Perception (IAP) model proposed by Cascetta & Papola (2001) includes the probability of a route being part of a choice set. This probability is a function of attitudinal and perceptual aspects of the traveller, which are in practice obtained from socio-economic variables (Ramming, 2002). However, this author proves the model does not produce satisfactory results.

Other methods

Constrained random walk

Frejinger (2007) develops a random sampling of choice set alternatives which uses a probability for each edge of being included based on the edge distance to the shortest route, allowing for this methodology to consider the universal set as the master set. The approach combines a stochastic constrained random walk algorithm with a probabilistic approach as well as a correction for route overlap (an expanded path size approach). The constraint in the random walk algorithm biases the random walks towards the shortest route. The author chooses a Kumaraswamy distribution with its parameters defining the bias of the algorithm towards the shortest path. Frejinger et al. (2009) find that this procedure, despite not being behaviourally realistic, significantly improves the model fitness when compared to models without corrections and with standard path size. The method is applied on a relatively small simplified network of Börlange, a small Swedish town. Rieser-Schüssler et al. (2013) find that the method was unable to produce suitable route choice sets for a detailed network where route lengths had large variations.

Breadth first search with edge elimination

Rieser-Schüssler et al. (2013) propose a route set generation algorithm that combines edge elimination with a breadth first search algorithm. This algorithm calculates repeated least-cost paths for an OD pair with the A-Star Landmarks shortest path algorithm, which is faster than the Dijkstra algorithm. The method builds trees of routes between an OD pair in a network based on a path-tree in which sub-networks are constructed from each node reached. The advantage of the method is that the tree-building procedure between an OD pair continues until a pre-determined number of routes in an OD pair is constructed. The method also guarantees that both the shortest path route is included in the set and that unique, realistic routes are generated. A main advantage of the method is its faster computation time, being at least 10 times faster than the second fastest method tested in the study (a stochastic method).

Metropolis-Hastings sampling

Flötteröd & Bierlaire (2013) propose a probabilistic method that can sample routes in a network with any probability distribution. The algorithm is an application of the Metropolis-Hastings algorithm to generate routes in a network. The algorithm generates a Markov chain based on a predefined statistical distribution, which is not necessarily normal. The algorithm randomly modifies a shortest route by varying a point in it and leaving other points fixed. Frejinger's (2007) random walks distribution, for example, can thus be replicated in this framework. The main advantage of the method is that it is able to sample paths for any application and it is more easily expandable to other travel modes than car. The main feature of the algorithm is that it does not provide a normalizing constant and therefore route enumeration is avoided, allowing for a path sampling to be conditioned by a given probabilistic distribution. However, the downside of the Metropolis-Hastings algorithm is that it is time-consuming, making it too costly for large networks.

Choice set generation for multimodal networks

The generation of route choice sets for mixed networks is still in its infant stage and has only been developed from the early 2000's. Hoongedoorn-Lanser (2005) extend the branch and bound transit route choice set generation method introduced by Friedrich et al. (2001) to be used in multi-modal networks. This method has the advantage of being easily reproduced since it is not a stochastic process. Friedrich's method (Friedrich et al., 2001) also has the advantage of presenting only a small number of alternatives, since in his study only a maximum of five alternatives are generated. This is based on the number of alternatives that are presented to real travelers based on the train online trip information system of the German railways.

The main challenge in building a choice set for multimodal networks is the fact that the transport supply system consists of a multi-layered combination of both discontinuous (transit) and continuous (all other) services. This introduces several practical aspects to model such a supernetwork (Fiorenzo-Catalano, 2007). In the routing within a multi-modal, or PT network, a special attention has to be given to transfers.

Fiorenzo-Catalano (2007) states that the following methods have been applied to multi-modal networks:

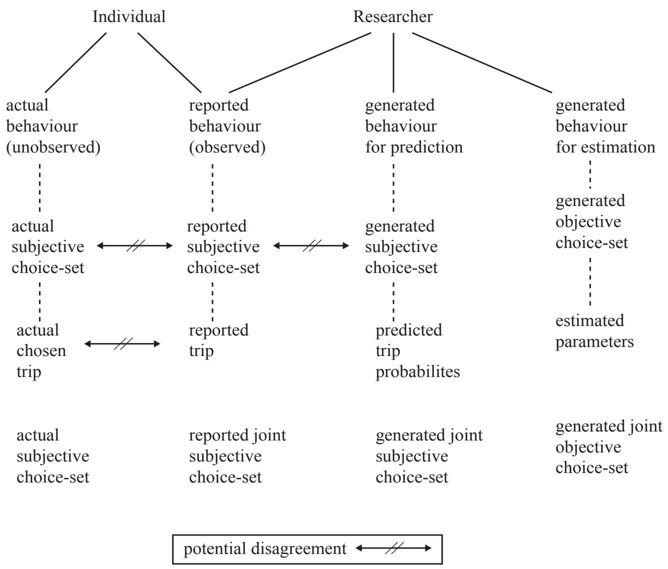
- The k shortest path approach
- Shortest path with composition rules
- Multi-objective shortest path search
- Labelling approach
- Simulation method

Hoongedoorn-Lanser's (2005) model is a branch and bound algorithm which poses a series of constraints to the choice set construction. These constraints are posed so that the route choice generation model mimics the actual choice set behavior, such as posing a constraint on unnecessary transfers, eliminating loops or avoiding travel in the opposite direction of the destination.

Overview of route choice set generation methods

While deterministic methods have been the state of the art for many years, since the 2000's much effort has been made in order to develop stochastic and probabilistic methods which effectively provide a better route coverage. This is extremely important for route choice modelling, since the omission of relevant routes in the choice set can lead to biased parameter estimation (Schüssler & Axhausen, 2009). These efforts have led to the development of more complex models, which also add stochastic elements to a deterministic framework or combine it with sampling probabilities. Still, there is no single route choice set generation method that is able to replicate all of the relevant routes for the decision maker in one choice set. Parameters of route choice models constructed with the traditional two step route choice modelling framework (choice set construction and then route choice) will always be somewhat biased. Bovy (2009) states that constrained enumeration methods are so far the best since such a method has been able to achieve a 97,9% reproduction rate of observed routes. Ramming (2002) finds values in the range of 40-60% for deterministic methods and only when combined with stochastic ones, a maximum reproduction rate of 84% was achieved (he did not evaluate constrained enumeration). Rieser-Schüssler et al. (2013) on the other hand find a coverage rate of up to 73% of chosen routes with their method. This rate was higher than the stochastic method evaluated in their study and the constrained enumeration method evaluated. One of the main issues with the studies which apply algorithms is that they are either based on a synthetic network or on a macro-scale simplified network of a city. Rieser-Schüssler et al. (2013) draw attention to the fact that some models are not well suited for choice set generation in detailed navigation networks which are the closest to the real-world networks. The availability of GPS data sets demands such detailed networks and a representative choice set generation and, to meet such demands, the adaptation of methods is needed. Furthermore, the increasing complexity of the modelling framework of route choice set generation models does not always generate better choice sets. This is due to a great mismatch between the modelling framework and the conceptual framework at the individual level (Figure 2). Kazagli et al. (2016) propose a new approach, in which the methodology is simplified in order to match the simple mental representation conducted by travellers when making route decisions. The drawback of this model is that it has to be adapted on a case by case basis. Fosgerau et al. (2013) propose a different approach which considers the master choice set as a consideration set, breaking off from the commonly accepted conceptual framework for a traveller's behaviour, but allowing the model to represent all possible routes in the choice set. This approach led to the development of the recursive logit model, which is reviewed after the presentation of common route choice models.

Figure 2 Relationship between actual, observed, estimated and predicted behaviour and corresponding choice set.



Source: Hoongedoorn-Lanser (2005)

2.4.2 Route choice models

Route choice models are divided into two main families, based on the assumption made on the probability distribution of the error term. Simple deterministic shortest path models such as the Dijkstra algorithm will not be discussed here.

Logit models

The choice of a route has many properties that distinguish it from other choice situations, such as mode choice models. These properties make the two main logit family models, namely Multinomial Logit and Nested Logit, not suitable to be used in a route choice context. The former because it does not allow for correlation among alternatives, and the latter because it assumes that edges belong exclusively to one nest (i.e. route) which is not a realistic assumption (Prato, 2009). The two Logit family models presented below are based on an MNL model but add a correction term to account for correlations between routes.

C-Logit

The C-Logit model is one of the most common for route choice. It was developed by Cascetta et al. (1996) and has the following form:

$$P_n(i|C_n) = \frac{e^{V_{in} - CF_{in}}}{\sum_{j \in C_n} e^{V_{jn} - CF_{jn}}} \tag{5}$$

Where: V_{in} is the utility of path i for person n.

C_n is the path set for person n.

CF_{in} is the commonality factor of path i for person n.

The commonality factor is the degree of similarity between path i and the other paths in C_n.There are at least four different formulations of this factor (Prato, 2009). For the sake of simplicity only the first one will be presented here based on Cascetta (1996). This formulation expresses correlation based on the common length between routes.

$$CF_{in} = \beta_0 \ln \sum_{j \in C_n} \left(\frac{L_{ij}}{\sqrt{L_i L_j}} \right)^{\gamma} \tag{6}$$

Where: L_{ij} is the length of edges common to path i and j.

L_i and L_j are the length of paths i and j respectively.

β_0 and γ are the parameters to be estimated.

The advantage of C-Logit consists in the robustness of the parameters (Prato & Bekhor, 2007). The disadvantage is that the commonality factor only captures the correlation based on one attribute (Prato, 2009).

Path Size Logit

Path Size Logit models outperform C-Logit ones (Prato, 2009). Its formulation is as simple as the C-Logit one but is based on a different theoretical framework (Ben-Akiva et al., 2004):

$$P_n(i|C_n) = \frac{e^{V_{in} + \ln PS_{in}}}{\sum_{j \in C_n} e^{V_{in} + \ln PS_{in}}} \tag{7}$$

Where: V_{in} is the utility of path i for person n.

C_n is the path set for person n.

PS_{in} is size of path i for person n.

The path size PS is a measure of how many edges a route PS shares with other routes. The values of PS are bounded between 0 and 1, 1 representing a route that shares no edge with any other route. The path size is defined as follows:

$$PS_{in} = \sum_{a \in \Gamma_i} \left(\frac{l_a}{L_i}\right) \frac{1}{\sum_{j \in C_n} \frac{L_i^{\gamma}}{L_j^{\gamma}} \delta_{aj}} \tag{8}$$

Where: l_a is the length of edge a;

L_i is the length of path i;

δ_{aj} is 1 if edge a is in path j and 0 otherwise;

Γ_i is the set of edges of path i;

γ is a parameter to be calibrated.

Higher values of γ are associated with a better fit of the model (Prato, 2009). The definition of the path size above seems to be problematic since it produces counter-intuitive results for these high values (Frejinger, 2007) and requires non-linear utility functions (Prato, 2009). Further, the behavioural interpretation of the γ parameter has proven difficult (Prato, 2009). A limitation shared by the Path Size Logit and the C-Logit model is that both models capture only a part of the correlation among routes.

Generalized Extreme Value models

Paired Combinatorial Logit

The application of the Paired Combinatorial Logit model was presented in a route choice context by Prashker & Bekhor (1998). In this model, route alternatives are paired for a choice.

$$P_k\,(i) = \textstyle\sum_{k \neq j} P(kl)P(k|kl) \tag{9}$$

Where $P(kl)$ is the marginal probability of choosing the pair *(k,l)* among all possible pairs and $P(k|kl)$ is the conditional probability of choosing route k of the pair.

The choice probabilities are highly dependent on the correlation between routes k and l. This correlation parameter is identical to summation term of the C-Logit correlation term and therefore also dependant on route overlaps. However, the Paired Combinatorial Logit approach has not been thoroughly studied yet.

Cross Nested Logit

Another way to account for the correlation between alternatives is to use a nested logit structure. The Cross Nested Logit model overcomes the limitation of the Nested Logit approach which only allows edges to belong to separate routes. The first application of this model to route choice was done by Prashker & Bekhor (1998). The choice probabilities are as follows:

$$P(i) = \sum_m P(m)P(i|m) \tag{10}$$

m being the nest (which represents edges) and *i* a route passing through this edge. The conditional and marginal probabilities are respectively:

$$P(i|m) = \frac{(\alpha_{im} e^{-V_i})^{1/\mu}}{\sum_j (\alpha_{jm} e^{-V_j})^{1/\mu}} \tag{11}$$

$$P(m) = \frac{\left\{\sum_i (\alpha_{im}\, e^{-V_i})^{1/\mu}\right\}^{\mu}}{\sum_b \left\{\sum_j (\alpha_{jb}\, e^{-V_j})^{1/\mu}\right\}^{\mu}} \tag{12}$$

Where:

$\alpha = \frac{l_a}{L_i}\delta_{ai}$ is the correlation coefficient. l_a is the length of the edge in the nest and L_i the length of the route. δ_{ai} is a dummy which is equal to 1 if the route *i* uses the edge *a* and zero otherwise.

μ is the nesting coefficient.

Applications of the model often result in nesting coefficient values close to 1, meaning that the model tends to collapse to a MNL model (Prato, 2009). This also means that the correlation coefficient does not capture route similarity properly. The degree of suitability of the model is not superior to the models presented above (Prato, 2009).

Bekhor and Prashker (2001) suggest a generalization of the Cross Nested Logit to a Generalized Nested Logit model. This is achieved by changing the nested coefficient towards a parametrized average of the inclusion coefficients:

$$\mu = \left(1 - \frac{\sum_{i \in C_n} \alpha_{ai}}{\sum_{i \in C_n} \delta_{ai}}\right)^{\gamma} \tag{13}$$

The results from this model are not significantly different from the Cross Nested Logit model presented by Prashker & Bekhor (1998) and it also tends to collapse to a MNL model.

Non-GEV models

Multinomial Probit

The Multinomial Probit model was proposed by Daganzo & Sheffi (1977). It assumes a normal distribution for the error term. However, the cumulative distribution function does not have a closed form, thus the probability of choosing one alternative over the other is not predetermined. The model is not often employed because of the high efforts needed to estimate it (Prato, 2009).

Mixed Logit

The mixed logit model, also called Logit Kernel in the context of route choice modelling has the main characteristic that its error term can be decomposed into two parts: one that contains heteroscedasticity and one that is Gumbel distributed (Ben-Akiva et al., 2004). The first term captures the correlations among alternatives with factor analysis. The utility function in this model takes the following form:

$$U = \beta X + F\xi + v \tag{14}$$

Where: β is the vector of unknown parameters;

X the explanatory variables;

F are the factor loadings;

ξ the vector of multivariate distributed latent factors;

v is the Gumbel error term.

The probability of choosing a route i is dependent on the number of simulation draws *D:*

$$P(i) = \frac{1}{D}\sum_{d=1}^{D} \Lambda(i|\zeta^{d}) \tag{15}$$

Where: $\Lambda(i|\zeta^d)$ is the probability of choosing alternative i given the vector ζ^d of random variables.

$$\Lambda(i|\zeta^d) = \frac{e^{\mu(X_i\beta + F_i T\zeta)}}{\sum_j e^{\mu(X_j\beta + F_j T\zeta)}} \tag{16}$$

Where: $T = \frac{\xi}{\zeta}$ are the unknown parameters.

Ben-Akiva et al. (2004) show that the Logit Kernel model has a considerably better fit than all of the previously presented models. Moreover, they also add a path size term to account for route overlap, further improving model fit. Frejinger & Bierlaire (2007) introduce the concept of subnetworks, aggregates route choice sets into meaningful sets of alternatives to the traveller based on route characteristics (such as main highways). They use the Logit Kernel model in order to model the correlation of routes in the subnetwork and assume that the routes are correlated even if they do not share any edge. Although the model appears to provide a better fit than the other ones reviewed so far, Ramming (2002) shows that the model estimates can be instable even with large numbers of draws. Prato (2009) also mentions that it might be difficult to obtain significant estimates.

2.5 The recursive logit model

The recursive logit model was first presented by Fosgerau et al. (2013) in a context of growing interest in route choice modelling due to the ever-cheaper availability of GPS data feeds, allowing the analyst to directly observe chosen routes by travellers. It is the most significant development in route choice modelling in the recent years because it eliminates the most significant caveat of its predecessors: the need for generating a choice set. The recursive logit model is a dynamic discrete choice model with an infinite choice set, and the choice process is described as a sequence of edge choices (Fosgerau et al., 2013). The model is dynamic because it makes use of dynamic programming algorithms to maximize the utilities through sequential edge choices. The model is therefore consistent with Bellman's principle of optimality which states that *"an optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision"* (Bellman, 1957). In the dynamic programming terminology this optimality principle is also known as the optimal substructure property (Lew & Mauch, 2007). The interpretation for route choice models of this statement is that optimal routes also have optimal subroutes. This is the theoretical basis for the recursive logit framework, namely that the choice of a route can be broken down to the sequential decision of edges within that route.

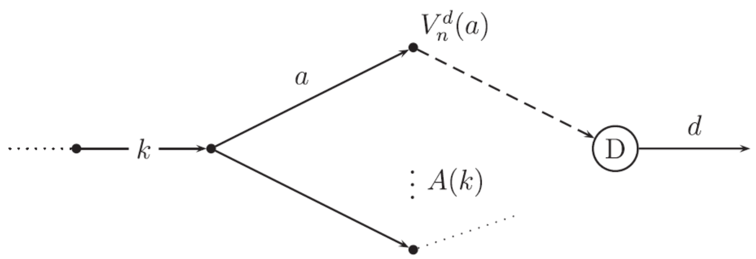
Edge utilities are defined for edge pairs and consist of two terms, an instantaneous utility of the sum of utilities and an expected maximum utility for a destination. The recursive logit framework can consistently estimate parameters, assigning probabilities of choosing paths given their observed attributes and a rational behaviour of individuals. The only inputs needed for the model is a network with edge attributes and observations in the form of a sequence of edge ID's followed within a route. Further on, Fosgerau et al. (2013) present a simple and straightforward way to predict edge flows based on the observations.

While the standard route choice models presented above are route based, assuming that the decision-making process followed by the traveller is equivalent to a simultaneous choice of all edges in the route, the recursive logit model formulates a route choice as a sequential choice of edges. Such an on-the-fly sequential edge choice is, as discussed in the beginning of this chapter, not a better but only a different psychological choice process. The advantage of the recursive logit model is, therefore, not a better understanding of the choice process of individual travellers, but rather a consistent and simplified route choice modelling framework. However, as will be demonstrated below, while the transferring of the choice context from a path-based to a edge-based one eliminates several issues, other practical problems arise.

2.5.1 Theoretical framework

Given a graph $G = (A, V)$ where A is the set of edges and V the set of nodes representing a static or dynamic transport network. $k, a \in A$ are edges where k is the state edge and a is a potentially chosen edge at edge k. Further on, $A(k)$ is the set of edges outgoing from edge k and d is a non-existent, dummy edge to represent a destination (Figure 3). This dummy edge is needed since the model is edge based, and to ensure that the destination node is visited. Being A the set of edges in the network and D the set of destinations, the set of all edges is $\hat{A} = A \cup D$.

Figure 3 Notation illustration



Source: Fosgerau et al. (2013)

When an edge k has been visited, the choice of the next edge a is made, so that both the instantaneous edge utility $u(a|k) = v(a|k) + \mu\varepsilon(a)$ and the downstream utility given by the value function $V_a^d(a)$ are maximized. The superscript d indicates that the value function is destination specific. The random term $\varepsilon(a)$ is assumed i.i.d. extreme value type 1 with zero mean. The deterministic part of the instantaneous utility $v(a|k)$ is a function of the observed characteristics of the current state, that is, the current edge k and the following chosen edge a. The choice of a edge a, given a state (edge k) is made so that the instantaneous utility at edge a as well as the expected downstream utility, until the destination k is maximized. The decision of the next edge k is in theory a Markov decision process, but with the important difference from classic Markov chains that the subsequential choice is not stochastic but deterministic. The utility of choosing an edge a given a state k is a maximization problem which is most easily solved by means of the Bellman equation (Bellman, 1957) as shown in Equation 17. This equation is the value function and it represents the expected maximum utility from edge k to the destination in the path.

$$\frac{1}{\mu} V_n^d(k) = \boldsymbol{E}\left[max_{a \in A(k)}\Big(v_n(a|k) + V_n^d(a) + \varepsilon(a)\Big)\right] \ \forall k \in A \qquad (17)$$

Since the error term $\varepsilon(a)$ is i.i.d. extreme value type I, Williams (1977) shows that the value function $V_n^d(k)$ can be written recursively by the following logsum:

$$V_n^d(k) = \mu \ ln \sum_{a \in A} \delta(a|k) e^{\frac{1}{\mu}\left(v_n(a|k) + V_n^d(a)\right)} \qquad (18)$$

Where: $\delta(a|k) = \begin{cases} 1 & a \in A(k) \\ 0 & a = d \end{cases}$

The econometric interpretation of this logsum is a consumer surplus or a gain in utility from making a choice. It can be rewritten as:

$$e^{\frac{1}{\mu}V(k)} = \begin{cases} \sum_{a \in A} \delta(a|k) e^{\frac{1}{\mu}\left(v_n(a|k) + V_n^d(a)\right)} & \forall k \in A \\ 1 & k = d \end{cases} \qquad (19)$$

For solving the system above, the equation is written in matrix form so that a system of linear equations can be solved. For this, it is assumed that **M** ($|\hat{A}|\ x\ |\hat{A}|$) is an incidence matrix which defines instantaneous utilities for all edges in the network.

$$M_{ka} = \begin{cases} e^{\frac{1}{\mu} v_n(a|k)} & a \in A(k) \\ 0 & otherwise \end{cases} \tag{20}$$

M has a zero row for each k=d. Let **z** ($|\hat{A}|\ x\ 1$) be a vector with elements $z_k \ = e^{\frac{1}{\mu} V(k)}$ and let **b** ($|\hat{A}|\ x\ 1$) be a vector with elements $b_k = 0, k \neq d$ and $b_d = 1$ for $k = d$. The system of linear equations becomes:

$$(\mathbf{I} - \mathbf{M})\mathbf{z} = \mathbf{b} \tag{21}$$

This system allows us to calculate the value function $V(k)$ for each destination using only instantaneous utilities for all edges in the network. Where **I** is the identity matrix. If **I** – **M** is invertible, the system has a solution. This is not always the case as noted by Fosgerau et al. (2013). To estimate the recursive logit model, such a linear system must be solved for each destination at each iteration of the Log Likelihood maximization algorithm. The calculations associated to this can become very expensive computationally. Furthermore, the computation time increases exponentially for larger networks since large inverse matrices have to be calculated. Mai et al. (2015b) introduces a decomposition method that allows to solve only one system of linear equations for all destinations at each iteration, allowing for significant time savings during computation. For common destinations, the probability of choosing a edge can now be calculated, independently of the origin:

$$\mathbf{P}_{\mathrm{k}} = \frac{\mathbf{M_k} \cdot \mathbf{z^T}}{\mathbf{M_k}\ \mathbf{z}} \tag{22}$$

Where · represents an element-by-element product.

Given the utility function in Equation 18, the probability of choosing an edge a given a state k and the choice set *A(k)* is:

$$P_n^d(a|k) = \frac{e^{\frac{1}{\mu}\left(v_n(a|k)+V_n^d(a)\right)}}{\sum_{a' \in A(k)} e^{\frac{1}{\mu}\left(v_n(a'|k)+V_n^d(a')\right)}} = e^{\frac{1}{\mu}(v_n(a|k)\,+V(a)-V(k)} \tag{23}$$

Which corresponds to a MNL choice probability of edge a given edge k. Considering that a path σ is a sequence of edges $(k_0 \ldots k_I)$ with $k_{i+1} \in A(k)$ the probability of choosing path σ is a MNL as given below:

$$P_n^d(\sigma) = \prod_{i=0}^{I-1} P\left(k_{i+1}\middle|k_j\right) = \frac{e^{\frac{1}{\mu}v(\sigma)}}{e^{\frac{1}{\mu}V(k_0)}} \tag{24}$$

Where $v(\sigma) = \sum_i^{I-1} v\left(k_{i+1}\middle|k_j\right)$ is the deterministic part of the utility. $V(k_0)$ corresponds to the value function representing all of the possible paths from an origin k_0 to a destination d. Denoting this set of paths as $\Omega(k_0)$ we have that

$$e^{\frac{1}{\mu}V(k_0)} = \sum_{\sigma \in \Omega(k_0)} e^{\frac{1}{\mu}v(k_0)} \tag{25}$$

And therefore, the choice probability for a single path collapses to a path based MNL model with an infinite but discrete choice set of all paths originating from k_0, $\Omega(k_0)$:

$$P_n^d(\sigma) = \frac{e^{\frac{1}{\mu}v(\sigma)}}{\sum_{\sigma \in \Omega(k_0)} e^{\frac{1}{\mu}v(k_0)}} \tag{26}$$

For estimating a recursive logit model, maximum likelihood estimation is used. The MATLAB estimation algorithm made available by Maëlle Zimmerman uses the BFGS algorithm, which is in many aspects the best algorithm for maximum likelihood estimation available (Train, 2009). For each iteration and for each observation the system of linear equations (Equation 21) has to be solved. This poses a potential problem, because this system does not have a solution for all possible parameter values, making the problem constrained (Fosgerau et al., 2013). If already one observation does not have a solution, then the log-likelihood function is not defined since (**I** - **M**) cannot be inverted. Fosgerau et al. (2013) deal with this issue by being conservative in the algorithm step size. The log-likelihood function for a vector of parameters β is defined for observations $n = 1, \ldots, N$ as (Fosgerau et al., 2013):

$$LL(\beta) = ln \prod_{n=1}^{N} P(\sigma_n) = \frac{1}{\mu} \sum_{n=1}^{N} \sum_{i=0}^{I-1} v(k_{i+1}|k_i) - V(k_0) \tag{27}$$

2.5.2 Further developments

Fosgerau et al. (2013) also introduced a edge size attribute analogous to the path size attribute in Path Size Logit route choice models. Such a parameter is important to account for the correlation of utilities of paths from different observations which overlap. As the same autor notes, *"ignoring this correlation may result in erroneous path probabilities and substitution patterns"* (Fosgerau et al., 2013). The main difference between the correction in the path size logit model and the edge size attribute in the recursive logit model is that while the first is based on the degree of overlap of different routes, the second is based on expected edge flows. Denoting the demand for trips originating at edge a and ending at destination d as $G(a)$ and the expected flow on edge a as $F(a)$, the latter can be analytically determined as:

$$F(a) = G(a) + \sum\nolimits_{k \in A} P(a|k)F(k) \tag{28}$$

The equation above can also be translated into a system of linear equations (Fosgerau et al., 2013):

$$(\mathbf{I} - \mathbf{P}^{\mathbf{T}})\mathbf{F} = \mathbf{G} \tag{29}$$

If $(\mathbf{I} - \mathbf{P^T})$ is invertible then the matrix $\mathbf{F}$ representing edge flows can be calculated.

The edge size parameter for a edge is the expected edge flow for the instantaneous vector of parameters $\tilde{\beta}$ at each iteration.

$$\mathbf{LS} = \mathbf{F}(\tilde{\beta}) \tag{30}$$

Fosgerau et al. (2013), Mai et al. (2015a), Mai et al. (2015b) and Zimmerman et al. (2017a) show that the inclusion of the parameter improves the model and assigns more realistic choice probabilities for different paths. This comes at the expense of significantly longer computational times though.

Mai et al., (2015a) introduce the nested recursive logit model, therefore relaxing the IID assumption for the MNL based recursive logit model. This model assumes that the scale of the error terms in the instantaneous utility is different for different random terms. This assumption results in the system of linear equations (Equation 21) to be non-linear since the scale parameters μ are not equal. This significantly increases the computational cost to estimate the model

2.6 Measuring route utilities

Route choice is governed by the knowledge of a network and the purpose of a trip. More significantly, however, the choice of a route is governed by cost and travel time (Raveau et al., 2016). The caveat here is that the difference between true cost and perceived trip cost by the decision maker are often significantly different (Henley et al., 1981). This ignorance by the modeler and decision maker is also one of the main characteristics of stochastic choice models (Vrtic, 2003). The true cost of automobile travel, for example, includes fuel cost, maintenance, and taxes, which differs from each automobile and driver. The perceived cost can be a fraction of the actual costs (Ivehammar & Holmgren, 2015). The cost awareness can also change significantly during periods of sharp price change (Henley et al., 1981).

For individual motorized travel, the number of traffic lights on the route, the number of left and right turns as well as U-turns are important parameters which are used in several route choice studies (Fosgerau et al., 2013, Ramming, 2002, Bekhor & Prashker 2001, Zantema et al., 2007 to name a few). It is usually expected that left turns provide a greater disutility since opposite traffic lanes are usually crossed when making left turns, while right turns generally do not cross concurring traffic flows. For active modes, the ascending slope in the route should also play a significant role and should also be part of the utility equation (Zimmerman et al., 2017a).

For transit, the cost structure is usually more complex. Important parameters to be estimated are access, waiting, egress and transfer times, headway times and, of course, the trip costs. In some countries, especially those with open-transit systems, flat fares are set for zones independently of mode (as in most central European countries). For others, the fare is distance-based and varies for different modes. For some developing countries, public transport cost is not unitary and might be dependent on direct negotiation between the provider and the traveler. In addition, many exemption structures are often available for groups with lower a purchasing power such as students, younger individuals or retired senior ones. Furthermore, fares might change depending on the time of the day. The exact estimation of the fare cost requires detailed information on the individual. When fares are not paid upfront, that is, when a monthly or yearly travel card is acquired, the travel cost is not trip related. Economically speaking, they are sunk costs and cannot be easily reproduced as part of a single trip. However, they should be accounted for in some way.

This raises the question of whether the decision-making framework for mode and route choice should be limited to an individual trip choice. Indeed, experience with SP surveys shows that framing effects have a significant impact on the chosen alternative (Biswas, 2009). Beck et al. (2017) show that the decision framework for mode choice models has a significant impact on model results. The inclusion of long term influences can particularly change key parameters, such as the values of travel time savings by a magnitude of 10. Such influences are, for example, the choice of owning an automobile, location choice, and budget allocation for different activities.

2.7 Multimodal route choice

Multimodal trips are trips composed of two or more stages in which at least two different transport modes are used. All transit trips are therefore multimodal, since another mode has to be used to get to and from a transit station. While most research has focused on route choice in road networks, not many efforts have been directed to the route choice problem in transit networks and even fewer efforts have focused on multi-modal networks. A transit network poses a series of difficulties to work with in a route choice framework, as is presented in the next chapter. The first works dedicated exclusively to multimodal route choice model estimation based on RP data were two doctoral theses at TU Delft, namely from Hoongedoorn-Lanser (2005) and Fiorenzo-Catalano (2007), the latter focusing on the choice-set generation problem for multimodal networks. Vrtic (2003) worked on a simultaneous mode and route choice model for long distance travel.

Hoongedoorn-Lanser's (2005) research is based on in-train interviews with travelers. She uses new choice set generation approaches, which are compared to subjective choice sets reported by the respondents to validate them. A diachronic supernetwork is used to perform a run-based branch and bound choice set generation algorithm, producing an average of 37 alternatives per chosen path. The generated alternatives are kept or deleted based on the evaluation of logic, feasibility and traveler's preference conditions. Hoongedoorn-Lanser (2005) shows two approaches for modelling multimodal route choice: direction-free models and directional models. Direction-free models distinguish trips by two types: home-end or activity-end. Such models have the purpose of describing the role of availability of transport modes for the traveler as well as network knowledge. An important behavioral assumption of these models is that choices made for the outbound trip will determine the return trip. Directional models are estimated to evaluate transfer-related aspects of multimodal trip decision. The behavioral assumption in these models is that the decisions made regarding the return trip are not dependent on the outbound trip and vice-versa. The author then estimates MNL, nested logit and path size logit models.

Fiorenzo-Catalano's (2007) doctoral thesis is instead more focused on the development and evaluation of different route choice set generation methods for multimodal networks. She too uses a supernetwork for the same area as in Hoongedoorn-Lanser's (2005) work. One main contribution of Fiorenzo-Catalano's thesis is the development of the doubly stochastic choice set generation algorithm, which generates individualized choice sets based on the traveler's preferences. These seem to play a more significant role in generating choice good quality choice sets than objective network attributes, thus showing how individual attitudes are important for a good representation of route choice behavior.

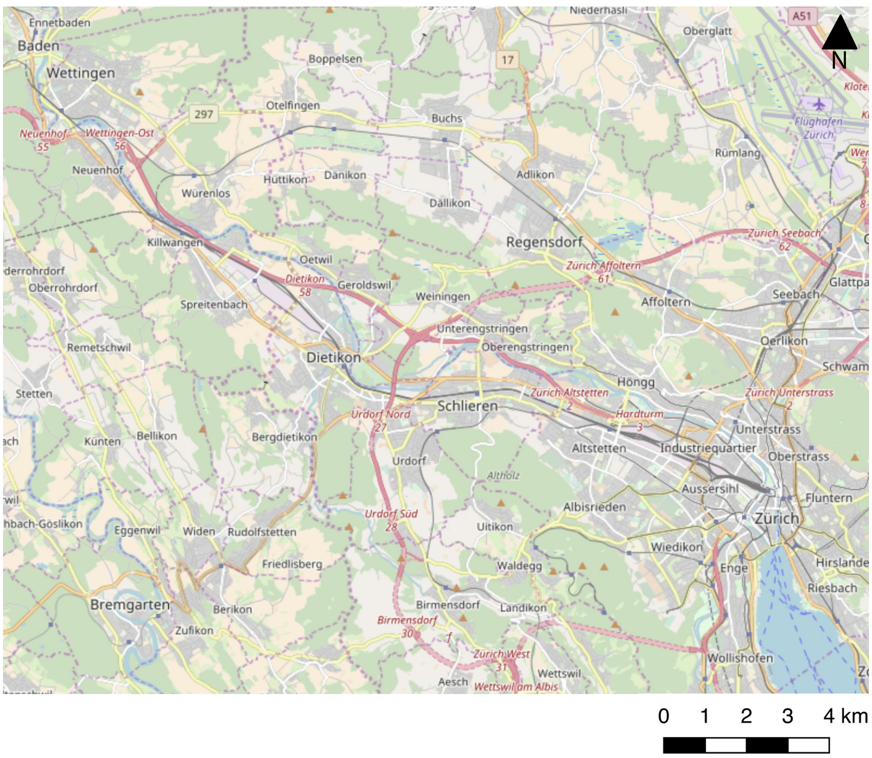
Anderson et al. (2017) use information on over 5'000 transit travelers in Copenhagen to estimate a mixed path size logit model on a choice set generated with a doubly stochastic algorithm. While Hoongedoorn-Lanser (2005) and Vrtic (2003) performed their studies on long-distance corridors in, respectively, the Netherlands and in Switzerland, Anderson et al. (2017) estimate their model on an urban network. The data of their model is a revealed preference data stemming from telephone and online interviews. Finally, Montini et al. (2017) are the first to estimate a multimodal route choice model on a real network using detailed revealed preference data stemming from GPS observations. These authors used a Breadth First Search on Edge Elimination (BFS-LE) algorithm (Rieser-Schüssler et al., 2013) to generate route alternatives for car, bike and walk modes and the *basic* and *via* choice set generation algorithm for transit stages (Rieser-Schüssler et al., 2014). The estimation of the model is made using an MNL formulation.

This work is the first to apply the recursive logit framework to a multimodal network. Zimmerman et al. (2017b) showed the application of the recursive logit model on a dynamic transit network of Zurich based on GPS data. In this study, on the other hand, a static transit network is constructed and then connected to a street network. The steps followed in this network design are detailed below.

3 Methods and data

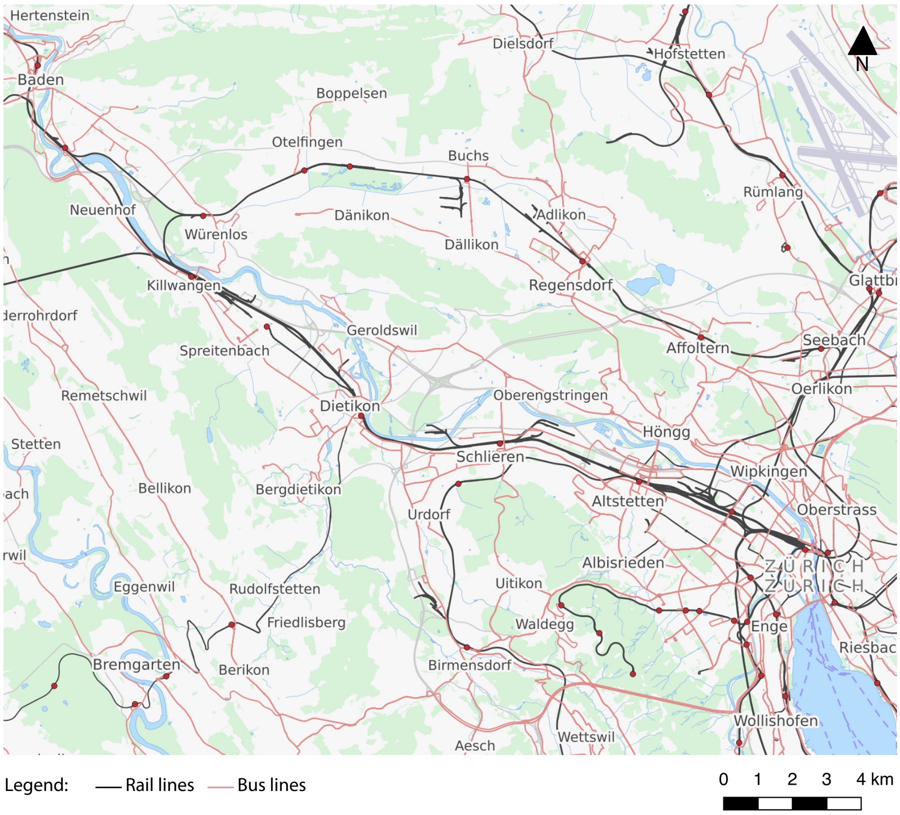
First, a study area had to be delimited, since it is would be computationally too expensive to estimate recursive logit models for all the reported paths in the Microcensus. An extended area of the city of Zurich was chosen for such purposes (Figure 4) The study area was delimited by the area with a 10km radius around Hardplatz in Zurich for the car network. For the transit network the area was delimited by the following bounding box: Latitude [47.2960, 47.4940], Longitude [8.3019, 8.6541]. The generation of the networks and the routing of observations in them is explained after the methodology for estimating the mode choice models is described.

Figure 4 Study area: Car network



Source: Open Street Maps (2018)

Figure 5 Study area: Transit network



Source: Open Street Maps (2018)

3.1 Mode choice models

For the estimation of the mode choice models, the Microcensus data had to be routed to obtain the variables needed for model estimation. For this purpose, different routers were tested. The routers used were the Google Directions API, the Graphhopper Directions API as well as the Conveyal R5 router. The routes were first estimated for path observations. Tables 1 and 2 present qualitative and quantitative aspects of the routers.

Table 1 Qualitative router evaluation

	Google Directions	GraphHopper Directions	Conveyal R5
Can route all modes	Yes	No transit	Yes
Provides alternative routes (transit)	Yes	No	Yes
Provides alternative routes (other modes)	Yes	Yes	No
Standalone version available	No	Yes	Yes
Possibility to extend standalone version	No	Yes	No
Cost-free use of API	No	No	-

Table 2 Quantitative router evaluation

	Google Directions	GraphHopper Directions	Conveyal R5
Max. alternatives car and slow modes	3	>20	1
Max. alternatives transit	4	-	>20
Avg. generated alternatives car	2.1	6.3	-
Avg. generated alternatives walk	2.5	2.8	-
Avg. generated alternatives bike	2.2	1.66	-
Avg. generated alternatives transit	3.5	-	2.5
Routing success car	89.5%	99.9%	-
Routing success walk	99.3%	99.9%	-
Routing success bike	99.7%	99.9%	-
Routing success transit	77.7%	-	99.0%

The ease of use and the fact that it is possible to route all the routes with it, made the Google Directions API the natural first choice for estimating mode choice models. As the used router produces a bias in the model estimation, it also makes sense to use the same router for all modes for the sake of consistency. The router did not exactly mimic the routes present in the Microcensus (Figures 5 and 6) but the results are to be interpreted as acceptable since the chosen routes in the Microcensus do not stem from a direct observation of route choice, but also from routed paths. In the phone interview, respondents are only asked about the start and end points of a trip and a few waypoints for car trips with over 5km of length. Therefore, the route lengths presented by the Federal Statistical Office (BFS, 2017) are also routed and represent an indirect observation. As can be observed the routed paths are shorter than the reported routes and show a shorter travel time. This bias is expected to stem from differences between the Google Directions API and the router employed by the Federal Statistical Office (BFS, 2017) to obtain the route variables.

Figure 6 Routed travel times/ Microcensus travel times

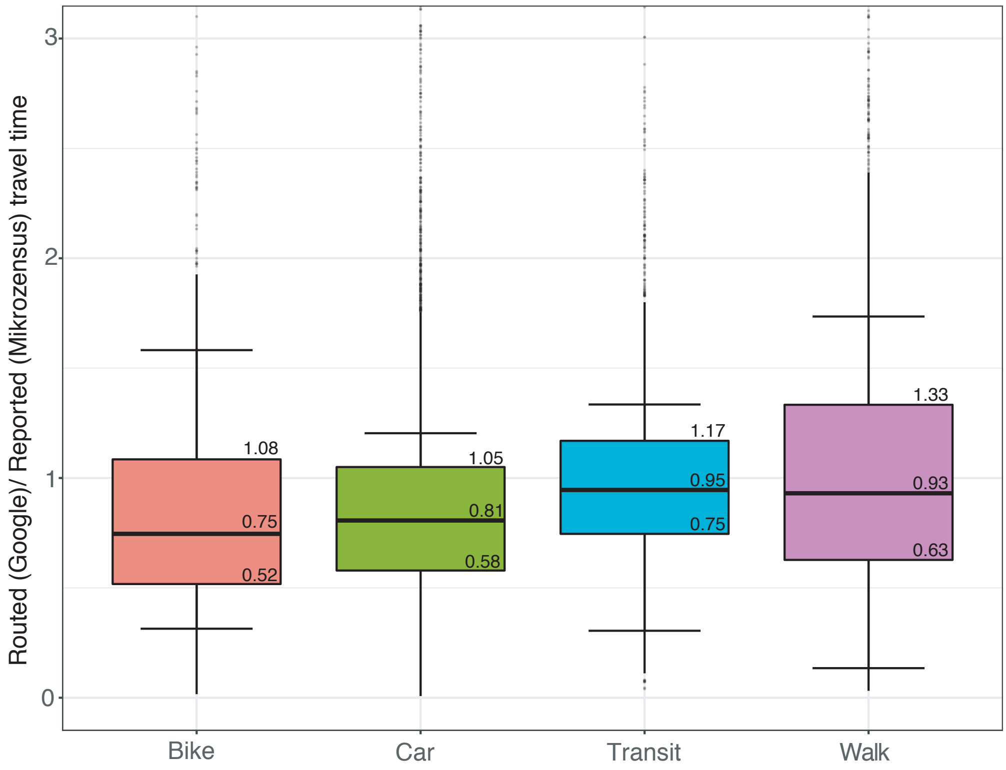
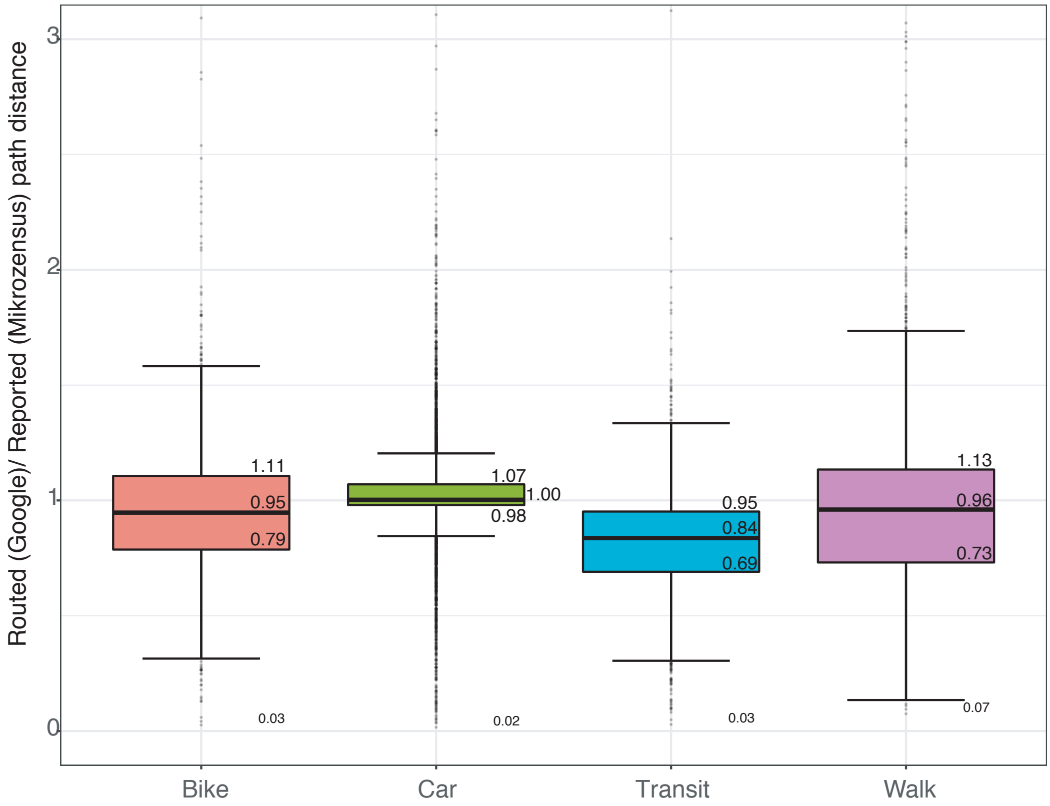


Figure 7 Routed travel distances/Microcensus travel distances



For each chosen path, unchosen alternatives had to be routed to construct mode alternatives for estimating a mode choice model. To obtain these, a route for the OD pair in each path observation is routed to obtain parameters for other modes.

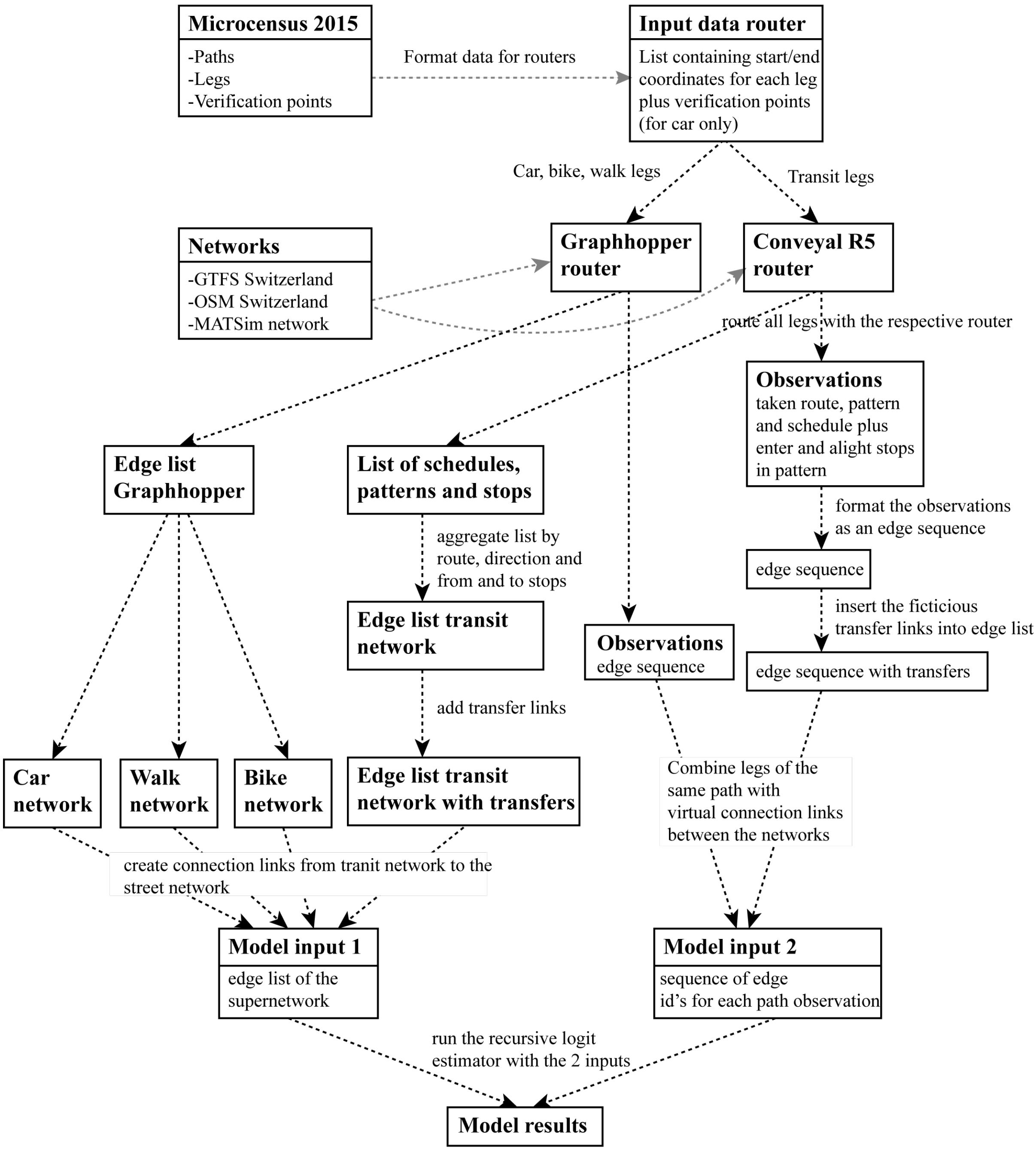
For the estimation of the models the following variables were extracted from the router on a path basis: travel times (for all modes), access/egress times (transit), waiting time (transit) and cost. The travel cost is not directly observed, but inferred. To calculate the travel costs for car, a linear function based on an average fuel price per km is used. The value used was of 0.12 CHF/km and draws on values used by Schmutz (2015) for mode choice modelling based on Microcensus data. The transit trip cost was also based on a linear function based on the national km price for tickets (VöV, 2017). To better represent the pricing scheme, a flat cost of 4.40 CHF was applied for all trips where the linear function provided a cost below this threshold. This is the minimal cost of a transit trip in Switzerland. For estimation, MNL models with linear utility functions were used. NL or other formulations were not tested, since the goal of these models is to provide a basis for comparison with the recursive logit model, which collapses to a MNL formulation on a route level.

3.2 Recursive logit models

This chapter describes the procedures followed to run a multimodal recursive logit model estimation (Figure 7). There are two main steps for the estimation of a recursive logit model, namely processing the observations to be matched on a network and reading out the sequence of chosen edge ID's for a given path. At the same time, the network has to be read out as an edge list containing edge variables. An edge list in the context of this thesis therefore always corresponds to the entire network for each mode. Given that survey data instead of GPS data is used in this study, the routes are not directly observed and as such have to be routed on a network. The routes are split into stages in the Microcensus data, which allows for a higher level of detail. The stages are separately routed and joined ex-post to form each path.

A multimodal router making use of the Conveyal R5 router (https://github.com/conveyal/r5) for transit observations and a modified Graphhopper router (https://github.com/graphhopper/graphhopper) for all other modes was implemented in Java. The output of this router was two files. One, from Graphhopper, contains a sequence of edge ID's for each observation. The non-modularity of the R5 Java code did not make such an output for transit stages possible. Chapter 3.2.2 shows how this issue was treated.

Figure 8 Scheme of the methodology



3.2.1 Street network

As mentioned, the network input for the recursive logit model is a directed edge list containing an edge ID, as well as the to and from nodes and variables for each edge. Such attributes are subject to the availability of information and can vary largely depending on the provider of the original street network. While the use of the Google Directions API proved infeasible since it is not possible to read out an edge list of the underlying Google map, open source routers such as Graphhopper and R5 allow this, since they can both run locally. While both are feasible, memory and estimation issues called for using the Graphhopper router for all non-transit stages.

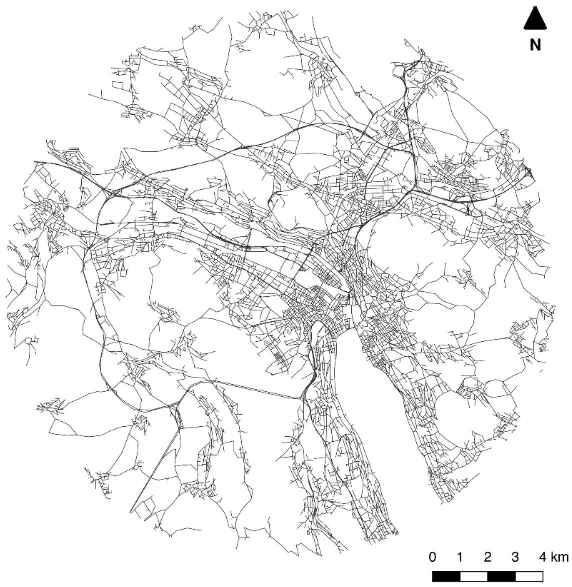
One main issue when estimating recursive logit models is memory usage and computation time. This is partly an issue that arises from using MATLAB for estimations, which is a memory-hungry application. Besides memory issues, the estimation time increases dramatically with larger networks given that for each edge choice in each observed path, utilities are calculated for all edges in the network. The issue of computation time is not as dramatic when estimating models on a supercomputing cluster, but another, more serious issue calls for a reduction of the network size. A too large network can cause the identity matrix **I** (see chapter 2.5) to be singular, and thus not invertible. Therefore, it is good practice to reduce the size of the network down to the smallest amount of edges possible. For this reason, topological instead of navigation networks should be used for recursive logit estimations.

The OSM maps used in the graphs of Graphhopper and R5 are navigation maps. The underlying graph of an OSM map is a sequence of straight edges, which become exceedingly small in curves. When running a recursive logit estimation, each of these small edges must be explicitly represented, yet that does not add any explanatory power to the model. For a route choice model, we are only interested in a topological representation of the network and the characteristics of its edges. To achieve a smaller amount of edges, with the same explanatory power as the original network, one should reduce the navigation network to a topological one. In the MATSim framework (Horni et al., 2016) such a simplification method is already implemented for the generation of the network. A pseudocode for this simplification algorithm used in the generation of the MATSim network is presented below.

```
for (i in edges)
    if (#edges following i ==1)
        j=follow up edge to i
        newedge=i
        fromnode(newedge) = fromnode(previousedge)
        tonode(newedge) = tonode(j)
        length(newedge)=length(i)+length(j )
```

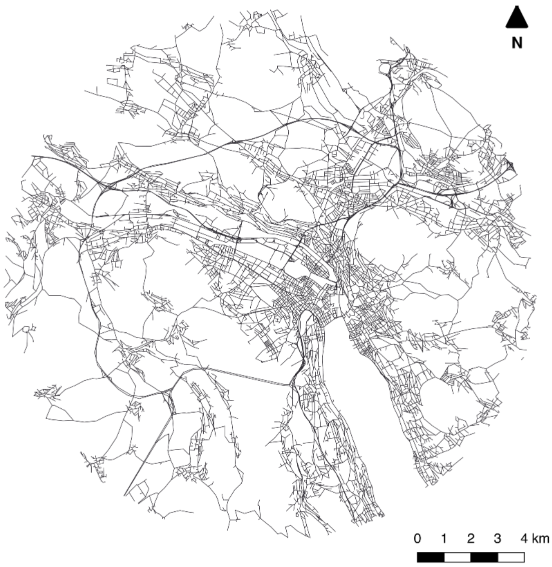
Molloy (2017) implemented methods in a standalone Graphhopper router to work with the MATSim network. This solution was then adapted to print a sequence of edge ID's for each routed observation as required by the recursive logit model. For the study area, the MATSim simplified network contains 30'372 edges and 13'829 nodes (Figure 8). The network simplification algorithm did not completely reduce the network to its topological features, but only simplified the network until the aggregated edges resulted in lengths of 50m. While a significant improvement is achieved through this code, it does not simplify complex intersections, such as roundabouts. The street network above was used to route all car stages. For bike and walk stages, a subset of this network was used, containing all roads except motorways (Figure 9). To draw a comparison, Figures 10 and 11 show the OSM car and bike/walk networks, respectively, used in the R5 router. These networks contain 154'342 and 210'171 edges, respectively.

Figure 9 Car network



Source: IVT 2015 Baseline Scenario (Bösch et al., 2016)

Figure 10 Bike/Walk network

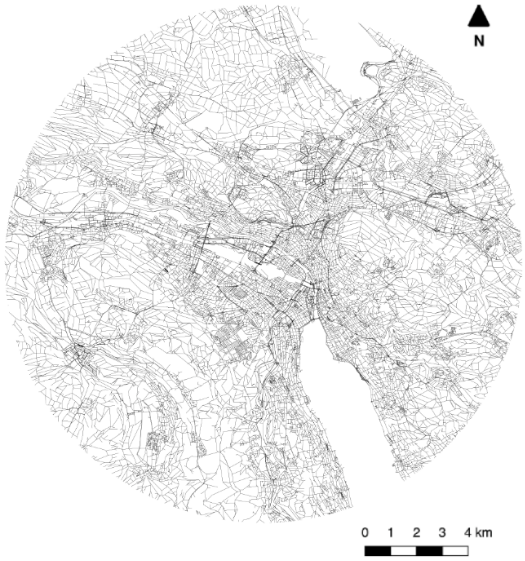


Source: IVT 2015 Baseline Scenario (Bösch et al., 2016)

Figure 11 OSM car network

Source: Open Street Maps (2018)

Figure 12 OSM bike/walk network



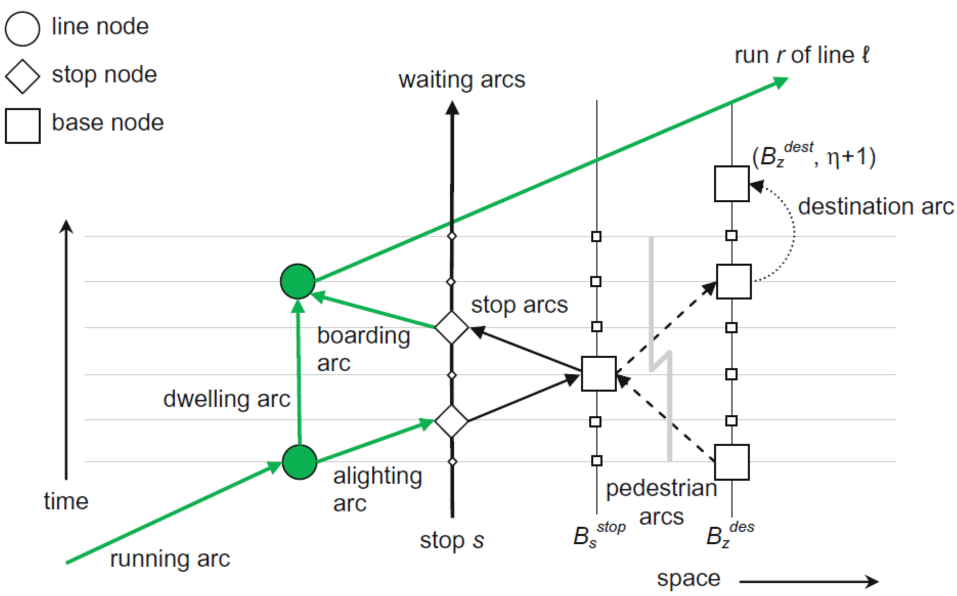
Source: Open Street Maps (2018)

3.2.2 Transit network

The transit router employed for routing transit stages was the Conveyal R5 router as it was the only transit router tested, which allowed a transit network to be read out. The R5 routing graph is easily created by joining an OSM network in a pbf format, together with a Generalized Transit Feed Specification (GTFS) file covering the same geographical area. The OSM data as well as the GTFS data used for the R5 router are open source. The first one is provided by Geofabrik (download.geofabrik.de), while the second one is available on the Swiss open data platform (opendata.swiss). The Swiss GTFS data created and published by the Swiss Federal Railways (SBB), which maintains databases on transit schedules for the entire country. The GTFS dataset for Switzerland contains the nation-wide timetable for the calendar year 2017, consisting of 5’505 transit routes. For the study area, there are 449 unique transit routes.

The R5 graph created with OSM and GTFS data is a diachronic supernetwork. Sheffi (1985) defines a supernetwork as a multi-modal network containing all the possible connections between nodes in each network for each mode. The diachronic term extends this concept to a supernetwork which is discrete in time. Diachronic graphs (Figure 12) provide a suitable topological description of a diachronic supernetwork. It consists of the road network, the pedestrian network, the transit network, as well as waiting arcs, transfer arcs and intermodal arcs connecting the stop nodes to the pedestrian network. While such a graph is internally created by the R5 router, the non-modularity of its Java code does not allow one to read out all of its elements.

Figure 13 Detail of the topology of a diachronic graph



Source: Gentile et al. (2016)

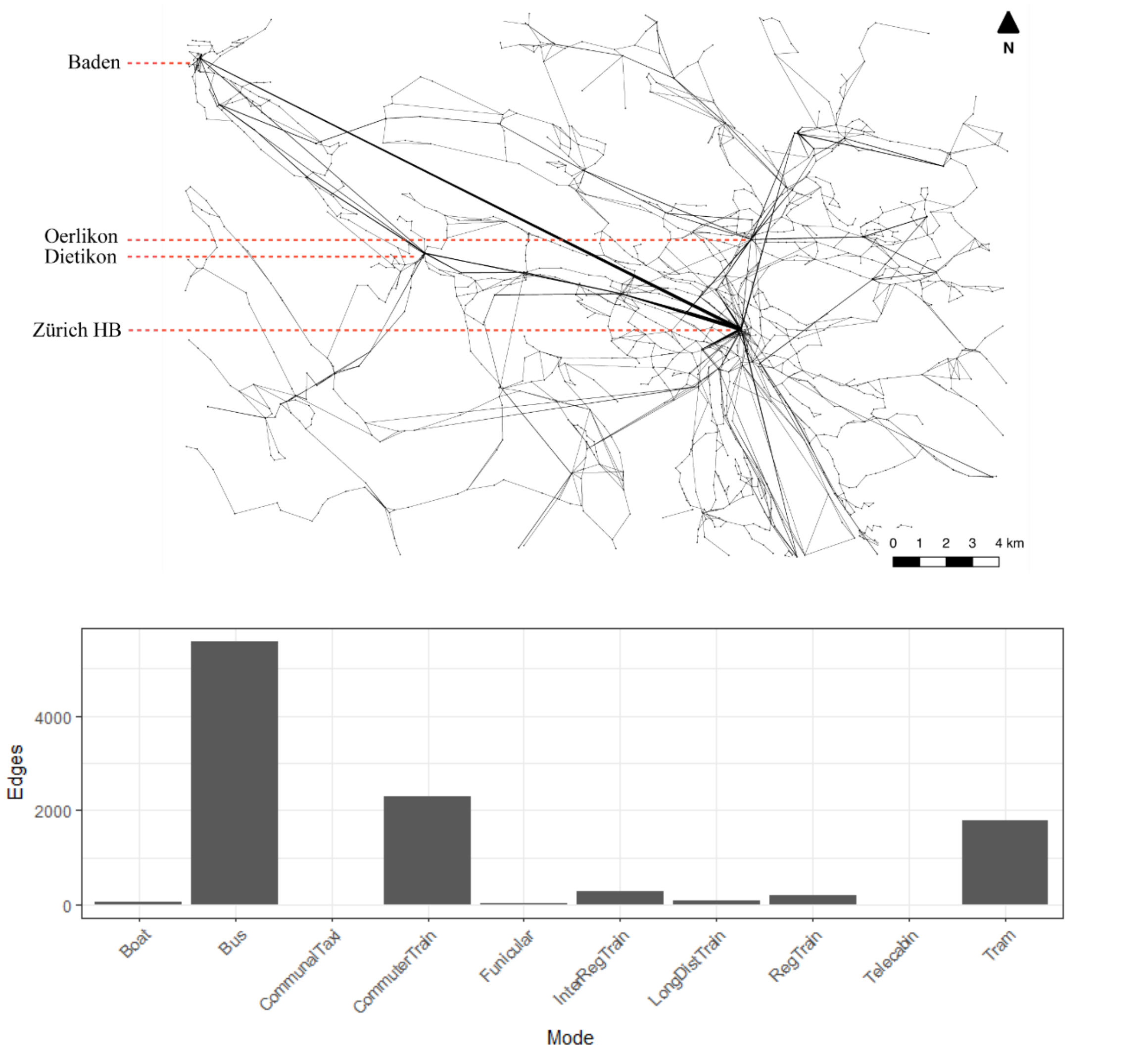
The transit part (only the time-dependent running arcs in Figure 12) of a diachronic graph representing the supernetwork for the study area consists of over 2 million edges. Estimating a recursive logit model on such a network would be unfeasible. Furthermore, a great part of the information in such a network is irrelevant, since travelers will usually disregard all other transit runs except those minutes away from their desired departure time. Zimmerman et al. (2017b) approached this issue by reducing the network to all transit runs which are available up to an hour after the departure time for each observation. This considerably reduces the computational effort needed for estimating the model. The approach adopted in this project is different, though. Here, a static instead of a dynamic transit graph is constructed. Two facts speak for representing reality in such a coarser manner. The first one is the lower level of detail in the observations from the Microzenus compared to the available GPS tracking data used by Zimmerman et al.

(2017b). The paths available in the Microzensus dataset are collected by a phone interview, which is subject to significantly more errors than a GPS tracker. Furthermore, the departure times are reported in 15-minute slots. For a dense transit network, as the one in Zurich (with several transit lines running in headways under 10 minutes), it becomes impossible to uniquely identify the transit service used by each traveler. The aforementioned computational effort is another obstacle. While Zimmerman's et al. (2017b) work was restricted to a transit network, the inclusion of a walk, car and bike network in this project calls for a simplification of the supernetwork from a dynamic to a static one in a bid to avoid not only long computational times, but also the singularity issue in the identity matrix. The construction of the static transit network from the diachronic GTFS data is explained below.

The GTFS data format contains a dozen of text files, some of which are not always included in feeds (such as one containing headways for different lines). The relevant files for the router and for this work are the routes, trips, and stoptimes files. Each trip belongs to a route and each trip has a sequence of stops and departure times for each stop. R5 combines this information into what Conveyal calls a pattern. In their own words, a pattern consists of *"all the trips on the same route that have the same sequence of stops, with the same pickup/drop-off options."* (Conveyal R5, 2017). A pattern thus contains several courses of the same line. Courses are called trip schedules in R5. To create an edge list of the transit graph in R5, one has to first iterate through all the patterns in the network within the study area and then through all the schedules within a pattern. In the study area, there are 7'947 patterns and 23'750 schedules. The average number of schedules, or courses (runs of a transit line), per route is, therefore, 53 courses per day. Unfortunately, a route-based edge-list cannot be read out directly from R5 but has to be constructed manually. There are no route edges with uniquely identifiable edge ID's or stop ID's with coordinates. The non-modularity of the R5 package also blocks any attempt to make such a construct in Java. Therefore, the network has to be read out with another approach. Luckily, R5 has an object named geometry, which is internally represented by line string representing the connections between all stops in a pattern. This object also stores coordinates of the stops.

Therefore, while there are no stop ID's, one can uniquely identify a stop by its coordinates and the pattern (and thus route) it belongs to. From the R5 router, a table is read out containing all the schedules of all the trips in the study area, namely the diachronic supernetwork with 2 million edges. This level of detail is needed in order to match the observations to each route later on. A table is then read out containing the following information: route identifier; mode of the route; pattern identifier; direction; trip identifier; number of the edge in the pattern; from stop; to stop; travel time. This results in a table identifying 2'044'129 edges for each unique schedule in each pattern. To obtain the final transit edge list, the diachronic edges from the transit pattern were aggregated by the route ID, direction and stops. The stops were not ID's, but a numbered stop sequence for each schedule. The resulting amount of edges from this aggregation is 10'298 edges representing the final transit network (Figure 13).

Figure 14 Transit network and edge count per mode



The modes available in the network for the study area are tram, commuter rail, bus, regional trains, interregional trains, long distance trains, funiculars, telecabin, boat and communal taxis. Telecabins, funiculars and communal taxis have few edges (Figure 13). For estimation, telecabins and funiculars were aggregated to steep modes. Long distance, regional and interregional trains were also aggregated to a train mode. This aggregation is based on the fact that these modes possess similar service characteristics within the study area.

The travel times for each edge are gathered by comparing the arrival time from the ‘to stop’ in an edge to the departure time from the ‘from stop’ in the same edge. Almost 10% of the edges read out from the schedules in R5 had a travel time of zero though (Figure 14). This issue arises from the fact that the GTFS data is based on the published transit schedule which has the minute as the smallest time unit and the travel times between pairs of stops is sometimes less than a minute. Some edges, on the other hand, had unrealistically long travel time. For instance, the travel time between Bahnhof Stadelhofen and Stettbach was 45 minutes, whereas in reality this trip does not take longer than 5 minutes. To overcome these errors, travel times were updated by using edge pair coordinates as inputs for the router and then reading out the transit travel time between them. To ensure that the mode of each link was used, the routing request also included the edge mode. Furthermore, link travel times below 1 minute were all set to 1 minute, independently of length, to ensure that no zero travel times were present in the network (Figure 15).

Figure 15 Edge travel times read out from R5

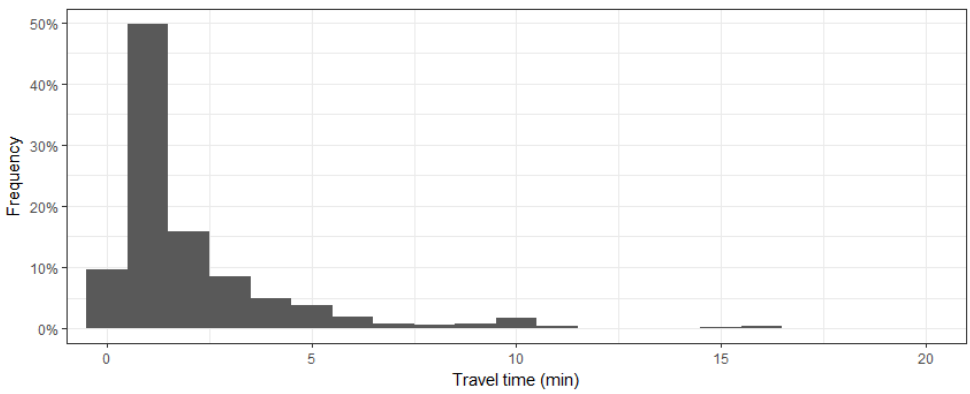
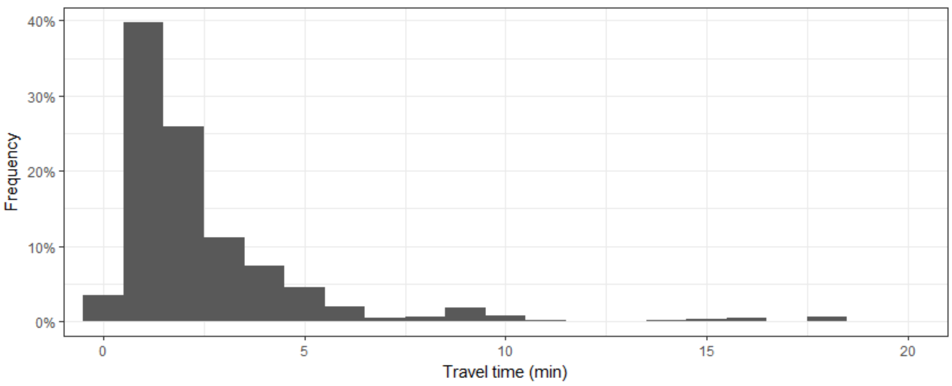


Figure 16 Edge travel times after processing



Another variable included in the edge list is the headway for each link. The GTFS specification for Switzerland does in fact have the optional headways file, which contains headways for routes. However, the data in this file is not to be taken into account given that 59.4% of the trips had a headway of 60 seconds. A first attempt to construct headways was done by constructing an algorithm in the router in Java. The algorithm compares all the schedules available for the same pattern and prints the minimum non-zero difference between two patterns. The edge headways obtained this way are very unsatisfactory and unrealistic (Figure 16). The reason for this result is the fact that a pattern has different schedules for different days. A bus which is scheduled to arrive at 15:34 at a certain station during weekdays, might arrive at 15:35 on weekends. Therefore, comparing schedules of all weekdays is not a feasible method for retrieving headway information.

A second attempt was made using the same method used for retrieving more realistic travel times. As a proxy for the entire day, headways were calculated for the morning peak (the time window for departures in the routing request between 7:30-8:30 was used). While the headways during the actual departure time of a traveler represent the true headway for the observed path such a level of detail is not possible in a static network. The results obtained with this method are much more satisfactory (Figure 17). The success rate for obtaining headways using this method was of 100%. The assumption that headways are stable across the day is therefore accepted, since most of the headways do not change significantly during the day (Figure 18). Headways for each mode are reported in Appendix 1.

The transit network constructed here is a time-dependent static snapshot of the real diachronic transit network. The specification of the time of the day used for the routing requests made to obtain travel times and headways thus represents the state of the network at a specific time-window. This allows the static network to represent the network state at different times of the day. Recursive logit models that aggregate observations from different times of day, e.g. peak, off-peak and night time windows in the network representing the respective time of day can thus be estimated, providing a higher level of fidelity of the variables to the experienced ones by the traveler.

Figure 17 Relative frequencies of headways read out from R5

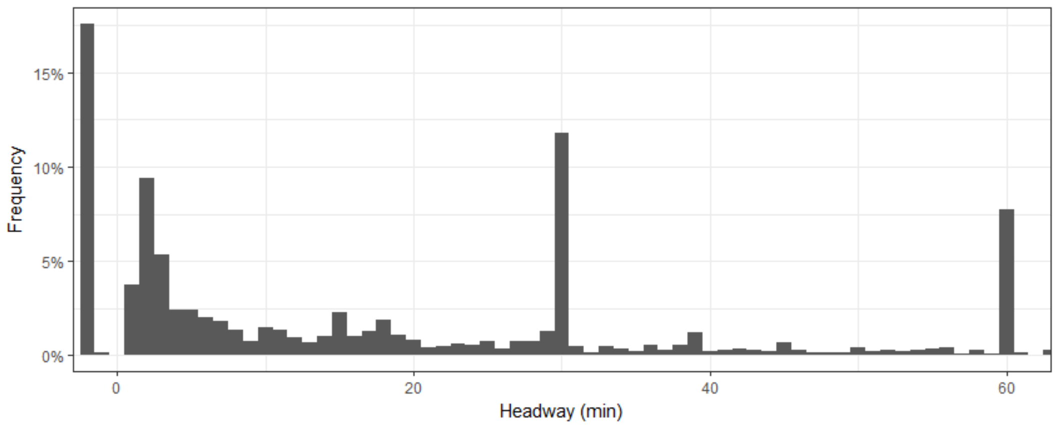


Figure 18 Relative frequencies of headways after processing

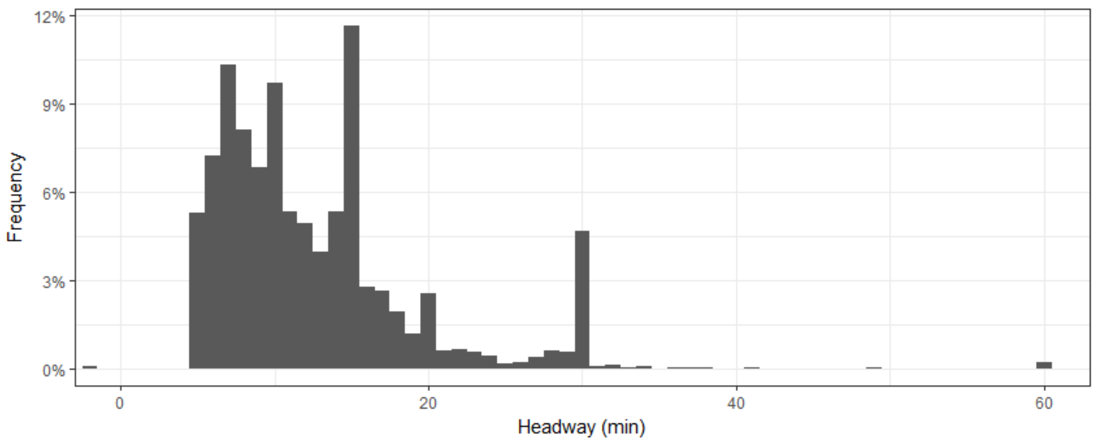
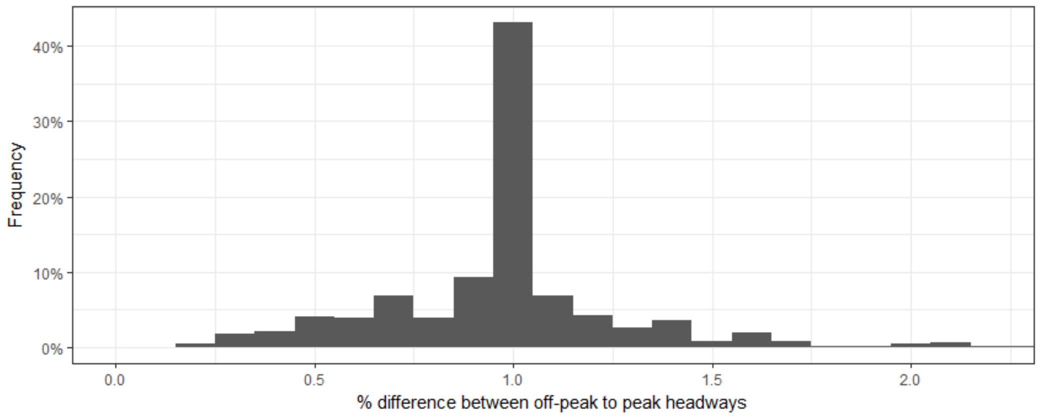


Figure 19 Ratio of headways in peak hours and off-peak hours



The transit network described so far is still incomplete though, since transfer edges are not available to connect separate stages. Virtual connection edges were created to connect all stops within a 150m radius from each stop. Since the actual connection time for each possible transfer cannot be calculated, a minimal transfer time was used. This minimal time is calculated as 1.2 times the crowfly distance divided by an average walking speed of 1.2 m/s plus a 3-minute waiting time. Round edges also connected each node with itself. 9’556 transfer edges were created this way, resulting in a transport network with a total of 19’854 edges. Transfer times are actually composed of transfer waiting times plus transfer walking times (Hoongedoorn-Lanser, 2005). The transfer times represented here only take the latter into account, since the latter can only be represented by a diachronic graph. Transfer station related attributes such as number of available transfers and quality of the transfer facility itself could not be represented but they also play an important role in the transfer decision making (Hoongedoorn-Lanser, 2005).

Matching of observations to the static network

While the static transit network simplifies the dynamic network for estimation purposes, the routed paths are still in a dynamic form. In order to match the observations to the static network, each edge of the 2’044’129 long dynamic edge list was matched to an edge in the 10’298 long static edge list. For each routed transit stage, the pattern, schedule board and alight stops were read out. With this information, it was possible to match the boarding and alighting stops uniquely to a schedule in the dynamic transit edge list and therefore, to the static network. The transfer links in the case of paths with multiple stages were included by finding an edge connecting the last stop in the first stage, to the first stop in the second stage.

Static supernetwork

Having a street network and a transit network in place, both have to be connected. The transit network was connected to the street network by adding virtual connection edges from each node in the street network, to each node within a 150m radius in the transit network. 19'302 edges were created this way. The travel times for these edges are null since they only provide a virtual connection between both networks and do not correspond to actual link choices . It is important to note that the algorithm has to iterate through the street network first, since it has significantly more nodes than the transit one, thus creating more connection possibilities from each transit stop to the street network. This redundancy is needed, since the two routers operate on two independent networks and the more connections are available between both networks, the higher the probability of being able to effectively connect paths with stages from both networks. After this step is performed, a supernetwork is created with 97'501 edges and 14'593 nodes (Figure 19). The matching of different stages belonging to the same path in both networks is done analogous to the method for finding transfer links for separate transit stages.

Figure 20 The multimodal supernetwork

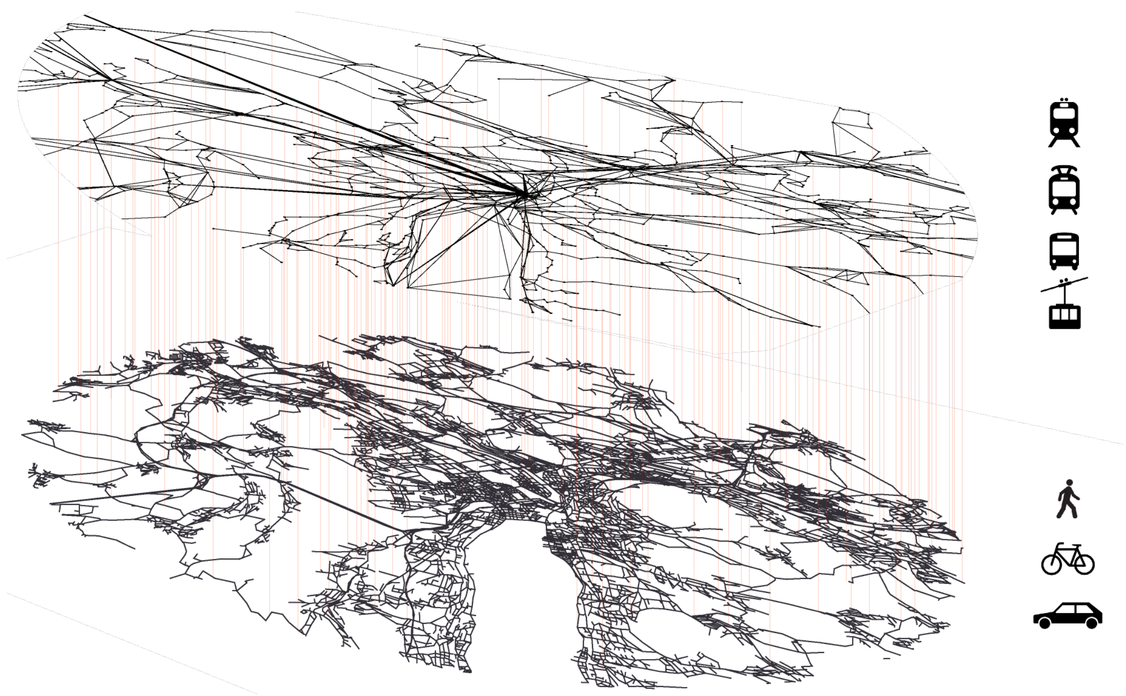
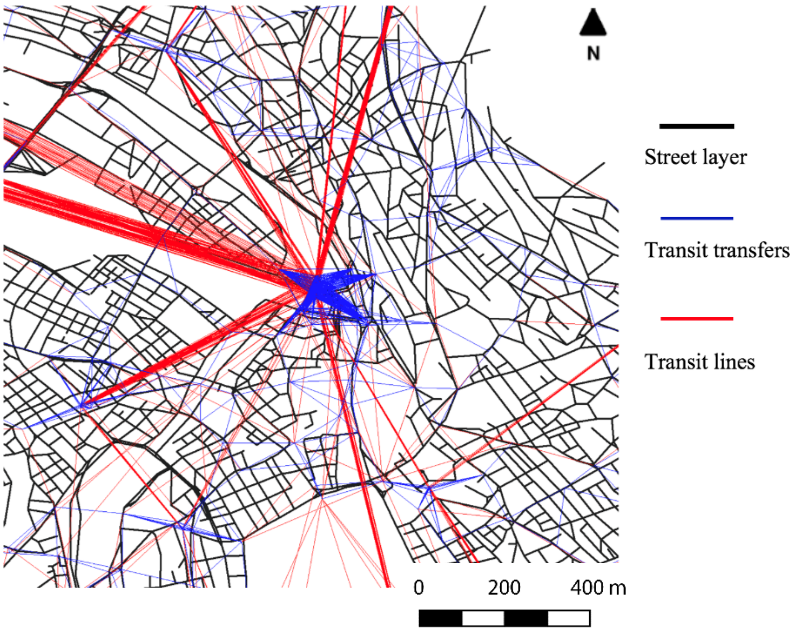


Figure 21 Detail of the multimodal supernetwork around Zurich main station



3.3 Microcensus data

The data employed for this project stems from the Swiss mobility Microzensus 2015 (BFS/ARE, 2017). This survey is conducted by the Federal Statistical Office (BFS) every five years since 1974 and aims at understanding the mobility behavior of the Swiss population. The data is collected through a computer assisted telephone interview in which individuals are asked about sociodemographic characteristics, the ownership of mobility tools (car, transit tickets), attitudes towards mobility and transport policies, irregular mobility behavior (trips with over-night stays, daytrips) as well as their everyday mobility behavior. The sampled population for the 2015 study consists of 57’090 randomly selected individuals from the entire country, which represents a response rate of 53.0% (BFS, 2018). In Zurich, 4’272 individuals participated in the interview. The path observations are collected at a stage level. For car trips over 5km of length, respondents are also asked about points passed in the route, so that verification points can be established for these car stages. The reference days for the everyday mobility behavior spanned from 19.01.2015 until 18.02.2016. Each respondent was contacted within 2 days after a personal reference day set to him/her and asked in detail about the trips conducted during the reference day.

Table 3 presents summary statistics for the data used in the estimation of the models. This data corresponds to all observations available in the Microcensus, minus round trips (with the same start and end points), since these cannot be routed. These accounted for 7.1% of all trip observations and consisted mostly of leisure walk and bike trips. The path based modal split for the selected area corresponds to the one official published by the city of Zurich (Stadt Zürich, 2017). The most probable reason for this difference is the fact that the bounding area for the modal split for the city are trips strictly within the city boundaries, whereas political boundaries are not taken into account for this project. The predominance of transit is very high in the study area, especially when the mode share is measured by the trip length.

Besides the raw input data from the Microcensus, Table 3 also shows the number of paths that could be used for estimation after routing the paths and processing them for joining the separate stages of the transit trips. For bike and walk, a rather small number of paths couldn't be routed. This was mainly because the start and/or end points for the paths were not close enough to a node in the sparse Graphhopper network. Graphhopper was then unable to assign a closest point with accuracy and route the observations. The number here is negligible though.

For transit trips, only 54.2% of trips could be joined together. The issue here were mostly not due to the routers but the joining of observations. For most of the cases, one of the possible connections between the transit network and the street network could not be found. While in areas with a high density of links (such as the city center of Zurich) this was not an issue, this problem is more severe in the fringes of the network. Since the transit network is a rectangular map extract, and the street network a circular one, many stations at the fringes of the network did not have street edges close by. Transit trips making use of these stations could therefore not be used. The tolerance for the maximum distance between a transit stop and the nearest street network node (150m) also played a role here, since for some separate stages, end stop of one stage and the start node of another stage were further away than this distance. For this reason, future work should focus on using the same router for all observations rather than virtually joining observations ex-post.

Table 4 presents a detailed insight into the stages of the transit trips, which are truly multimodal trips in the model. Most of the transit follow the pattern: access walk – main transit mode – egress walk. While these trips are strictly speaking multimodal, since different modes are used in the trip chain, it could be argued that they are not multimodal because access and egress walk modes always must be used for transit trips. Hoongedoorn-Lanser (2005) reports that most multimodal trips (excluding access and egress modes) occur in intercity travelling in the Netherlands. When looking at the stage mode share for transit trips in the entire Microcensus data, the total share of Bike in the stages rises from 0.7% to 1.5%, and the aggregated car share (as driver and as passenger) rises from 0.8% to 2.3%, indicating that also at a country average the share of multimodal trips, with transit as the main mode is still small.

Table 3 Summary statistic for observations

	Car	Transit	Bike	Walk	All modes
Paths	1846	2192	486	2355	6879
Mode share	26.8%	31.9%	7.1%	34.2%	100.0%
Paths (after routing)	1846	1188	462	2310	5806
Mode share (after routing)	31.8%	20.5%	8.0%	39.8%	100%
Average trip length (km)	5.57	7.60	3.00	0.83	4.38
Trip Length mode share	13.9%	72.2%	5.4%	8.5%	100.0%
Average path travel time (min)	16	37	14	11	21
Travel time mode share	22.1%	55.8%	4.3%	17.8%	100.0%
Average travel speed (m/s)	5.79	4.40	3.62	1.52	2.86
Stages	2938	7759	519	2419	13635
Average stages per trip	1.03	3.54	1.07	1.03	1.87
Average waypoints per path	3.71	-	-	-	-
Main trip purposes					
Work	30.3%	31.7%	32.1%	17.2%	24.3%
Education	20.3%	12.0%	11.7%	13.5%	10.1%
Shopping	36.9%	14.9%	16.5%	25.5%	19.8%
Leisure	50.3%	30.7%	32.7%	32.7%	32.0%

Table 4 Stages by mode in the transit trips

Mode	Stage count	Stage percentage
Walk	2192	54.9%
Bike	28	0.7%
Motorcycle (driver)	2	0.1%
Car (driver)	13	0.3%
Car (passenger)	18	0.5%
Train	343	8.6%
Bus (Postauto)	16	0.4%
Bus	637	16.0%
Tram	737	18.5%
Boat	1	0.0%
Funicular/Telecabin	2	0.1%

Other	2	0.1%

4 Results

4.1 Mode choice

First, mode choice models were estimated for the study area as well as for the entire Microcensus dataset. While more disaggregated results can be seen in the Appendix 2, only the results for the entire data and Zurich are presented in Table 5. The results presented in this table use the same observations which were used for estimating the recursive logit model for transit, complemented with observations for bike, walk and car. The router used for all the mode choice models was the Google Directions API for the sake of consistency. The utility functions used for estimating the model are the linear functions presented below. Simple linear functions were used, since the first goal of these models is to allow for comparisons with results from the recursive logit model, which collapses to a simple MNL.

$$V_{car} = \beta_{tt_{car}} TravelTime_{car} + \beta_{cost} TravelCost_{car} \tag{31}$$

$$V_{walk} = ASC_{walk} + \beta_{tt_{walk}} TravelTime_{walk} \tag{32}$$

$$V_{bike} = ASC_{bike} + \beta_{tt_{bike}} TravelTime_{bike} \tag{33}$$

$$\begin{aligned} V_{pt} = \; & ASC_{pt} + \beta_{ivt}\, InVehicleTime_{pt} + \beta_{at}\, AccessTime_{pt} + \beta_{et}\, EgressTime_{pt} \\ & + \beta_{tt}\, TransferTime_{pt} + \beta_{cost}\, TravelCost_{pt} \end{aligned} \tag{34}$$

The cost variable is a linear function of the path length as explained in chapter 3.1. While the results for the entire dataset are satisfactory and the values of travel time savings are similar to the ones found by Schmutz (2015), the same cannot be said for the results of the Zurich model. Schmutz (2015) found values of travel time savings of 2 CHF/h for transit and 16 CHF/h for car in a model with utility functions analogous to the ones employed here, but with Mackie interaction terms for the cost parameters. The simplest model estimated for the combined RP and SP data for the Microcensus (ARE, 2017) resulted in a value of travel time savings of 21 CHF/h for car and 14 CHF/h for transit.

Table 5 MNL model results for Zurich

	All		Zurich	
ASC Bike	-0.944	***	-0.798	***
ASC Transit	-1.534	***	-6.156	***
ASC Walk	1.867	***	2.339	***
Travel time car (min)	-0.095	***	-0.167	***
Cost (CHF)	-0.086	***	1.226	***
Travel time walk (min)	-0.14	***	-0.137	***
Travel time bike (min)	-0.128	***	-0.117	***
Transit in-vehicle time (min)	-0.04	***	-0.036	***
Transit access time (min)	-0.063	***	-0.062	***
Total transit time (min)	-0.033	***	-	
Transit egress time (min)	-0.054	***	-0.043	***
Observations:	124306		5564	
Final log likelihood	-69751		-4809	
Rho2:	0.43		0.33	
AICc:	139525		9637	
VoT car (CHF/h)	66.0		-8.2	
VoT transit (CHF/h)	27.8		-0.03	

Confidence levels: ***>99%, **>95%, *>90%

While the VoT for the entire data does seem more realistic than the ones estimated by Schmutz (2017) if compared to the officially published values by ARE (2017), the results for Zurich do not correspond to assumptions or expectations. The most striking result is a positive value for the cost parameter, which leads to negative travel time saving values (-8.2 CHF/h for car and -0.03 CHF/h for transit). When looking at the other models estimated (Appendix 2) one can see that positive cost parameters are estimated for models which restrict the observations. Boyce & Williams (2016) show that this issue is not new. Model results which were inconsistent with assumptions on travel behavior are one of the reasons for the development and wide acceptance of SP methods in the first place. The reason for that, is the very nature of RP data. It does not usually contain enough trade-offs between parameters and sometimes not enough variance within the variables used as an input. The model estimated for all the data does indeed contain more variation of choices and more variance in its parameters which is a possible reason for its better parameter estimates. Another important factor is the choice environment itself. While in an SP environment, respondents are directly observing the trade-offs between alternatives, and make decisions based on this trade-off, RP data reflects behavior which is often enough not based on any conscious choice, but reflect everyday behavior often based on simple rules of thumb.

4.2 Route choice

4.2.1 Car only

The first model to be estimated using the recursive logit model is a route choice model containing only car observations. Utility functions will not explicitly be shown here, because they are the same for all alternatives (i.e. links). The utility functions are simple linear functions of the parameters presented in the model results. Table 6 shows the correlations of the variables available for modelling. Cost and travel times are both a function of the length of the link. The freespeeds are on the other hand, not a function of the length. Still, the correlation of the cost and travel time variable is 0.86 and the estimation of a model containing both variables was not possible.

The estimation of this model is in first hand useful for validating the Microcensus data as well as the use of the recursive logit model for this type of data. Two other questions were also examined here. The first being, whether the inclusion of waypoints for car observations significantly changes the model inputs and therefore the results and second, whether parameter estimates are affected by an enlargement of the network. The same observations were used in the estimation of the different models. Figure 21 shows the original network with the observed paths. Figure 22, on the other hand, shows the network comprising an area with double the diameter of the first one. The results of the three estimated models are presented in Table 7.

Table 6 Variable correlations for the car model

	Capacity	Freespeed	Nr of Lanes	Travel Time	Cost	Type	Length
Capacity	1.00	0.62	0.90	-0.14	0.07	-0.14	0.07
Freespeed	-	1.00	0.42	-0.20	0.19	0.26	0.19
Nr of Lanes	-	-	1.00	-0.12	0.02	-0.08	0.02
Travel Time	-	-	-	1.00	0.86	-0.06	0.86
Cost	-	-	-	-	1.00	0.08	1.00
Type	-	-	-	-	-	1.00	0.08
Length	-	-	-	-	-	-	1.00

Figure 22 Observed car paths in the network

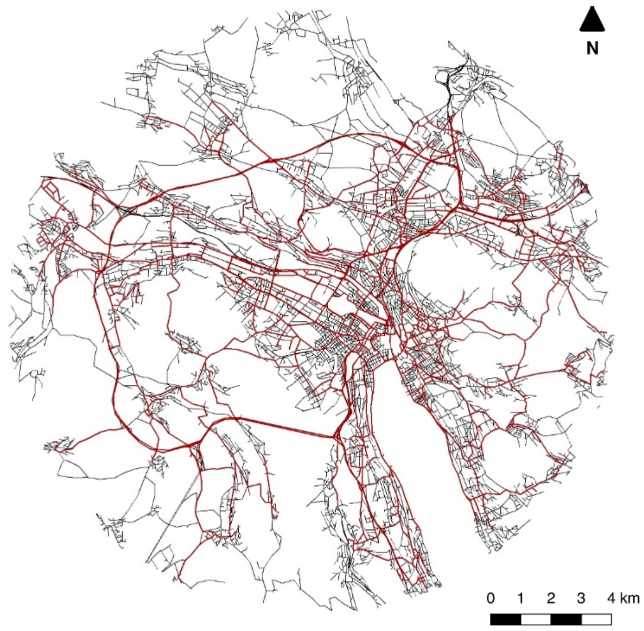


Figure 23 Observed car paths in the enlarged network

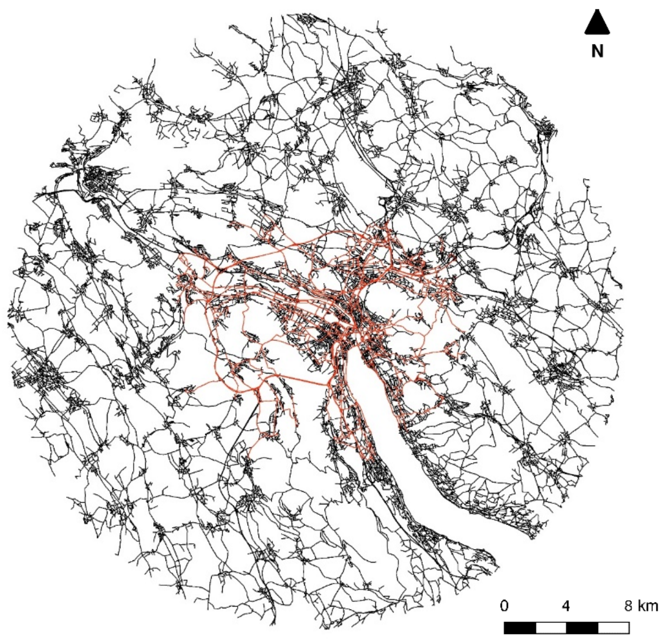


Table 7 Estimated car route choice model parameters

	No waypoints		With waypoints		With waypoints, larger network	
	β	**t-test**	**β**	**t-test**	**β**	**t-test**
Freeflow speed (km/h)	0.93	29.05	0.49	18.99	0.48	19.71
Cost (CHF)	-12.49	-17.25	-5.47	-10.99	-5.59	-11.1
No of Lanes	0.22	17.3	0.14	20.05	0.13	18.5
Edge constant	-2.61	-33.42	-1.74	-30.48	-1.69	-32.02
Log likelihood	-5003		-11872		-11909	
Observations	749		749		749	

By comparing the model results of the model with and without waypoints, it becomes evident, how important precise information on the choice of a route is. The model only with start and end points has a cost parameter which is more than double the value of the first one. The higher t-test value of the cost parameter in this model, compared to the models for the routes with waypoints represents the lower variation in route cost, which is expected, since they are always the fastest ones resulting from the routing algorithm. The higher determinism in chosen paths is also reflected through the considerably lower log-likelihood of the model with waypoints, since the probability of choosing the observed route becomes higher. While data quality for the observed route significantly changes the model results, the detail of the network does not make a difference. The use of a network double the size of the first one did not significantly change parameter estimates, confirming that results from the recursive logit model do not depend on the size of the study area.

4.2.2 Multimodal transit

The first multimodal model to be estimated used transit stage observations, that is, without access and egress stages and only the transit network consisting of 19’609 edges. First, aggregated models without a differentiation of travel modes was estimated (Table 8). A simple model consisting only of a link constant and travel time was estimated and compared to an identical model specification for car observations. A second model adds a headway parameter and a link dummy for transfer links. Further on, disaggregated transit models, which estimate separate parameters for different transit modes were estimated (Table 9). The mode-specific travel time and headway parameters are calculated as a multiplication of the variable with the link dummy of the respective mode.

Table 8 Aggregated transit models

	Car	Transit 1	Transit 2
Link Constant	-0.94***	-2.43***	-1.91***
Travel Time (min)	-7.66***	-1.05***	-1.22***
Headway (min)			-0.03**
Transfer dummy			-2.06***
Log likelihood	-13744	-25491	-24182
Observations	749	1769	1769

Confidence levels: ***>99%, **>95%, *>90%

Table 9 Disaggregated transit models

	Transit 3	Transit 4	Transit 5	Transit 6
Link constant	-2.81***			-4.04***
Travel time (min)			-1.09***	
Transfer time (min)	-1.27***	-1.03***		-1.02***
Train travel time (min)	-2.22***	-2.37***		-2.35***
Commuter train travel time (min)	-1.83***	-2.01***		-2.47***
Bus travel time (min)	-0.95***	-1.18***		-1.48***
Tram travel time (min)	-0.7***	-1.01***		-0.31***
Boat travel time (min)	-0.62***	-0.74***		
Funicular travel time (min)	-0.98***	-1.16***		
Headway (min)		-0.03**		
Train headway (min)			-0.35***	0.1***
Commuter train headway (min)			-0.01	-0.02
Bus headway (min)			0.09***	-0.04**
Tram headway (min)			0.1***	-0.03***
Transfer dummy		-1.81***		0.01
Train dummy				-0.74
Commuter train dummy				3.11***
Tram dummy				0.8***
Bus dummy				2.5***
Log likelihood	-24057	-23669	-24394	-24059
Observations	1769	1769	1769	1769

Confidence levels: ***>99%, **>95%, *>90%

After estimating transit only stages, transit models containing access and egress stages by walk were estimated using the supernetwork. The number of observations here is smaller than in the model before because of the overlap between both networks. While the transit network is a rectangular map extract, the street network is a circular extract, making it impossible to connect the transit network to the street network at its fringes. While the models presented above are already expensive to estimate, with estimation times reaching 20h, the ones below take even longer. The model Transit 8 took 29h to estimate and the model Transit 9, which has only 1 parameter more took 39h to estimate. The CPU used for estimations is an Intel i7-6700 with 3.40GHz. This happens since for each added variable, a matrix with the new parameter estimates is calculated for each interaction. Table 10 presents results for models with access and egress stages and Figures 23 – 24 present a comparison between ratios of the estimated parameters.

Table 10 Disaggregated transit models with access and egress

	Transit 7	Transit 8	Transit 9
Link constant	-10.60***	-10.70***	-10.64***
Transit Travel time (min)	-1.34***		
Transfer time (min)	-1.34***	-1.23***	
Train travel time (min)		-2.40***	-2.56***
Commuter train travel time (min)		-2.60***	-2.75***
Bus travel time (min)		-1.91***	-1.88***
Tram travel time (min)		-0.52***	-0.55***
Headway (min)			-0.08***
Transfer dummy	6.75***	6.90***	6.85***
Transit dummy	8.50***		
Train dummy		6.71***	7.67***
Commuter train dummy		9.19***	10.05***
Tram dummy		7.49***	8.04***
Bus dummy		8.69***	9.18***
Access/egress dummy	9.68***	9.80***	9.75***
Access/egress travel time (min)	-1.01***	-1.10***	-1.11***
Log likelihood	-18158	-23712	-17249
Observations	1197	1197	1197

Confidence levels: ***>99%, **>95%, *>90%

Figure 24 Travel time parameter ratios

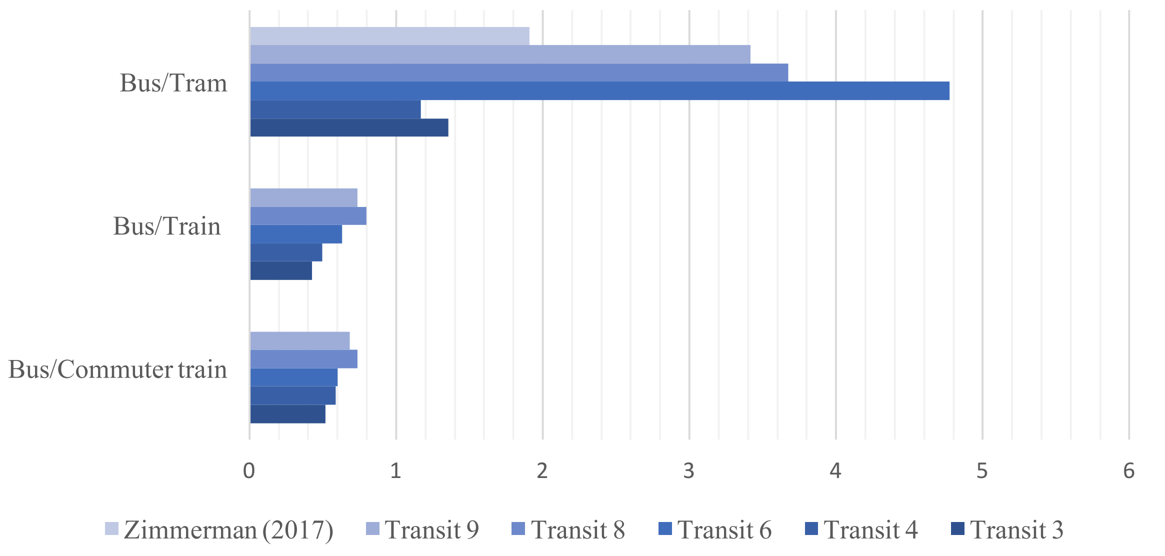


Figure 25 Travel time/ headway parameter ratios

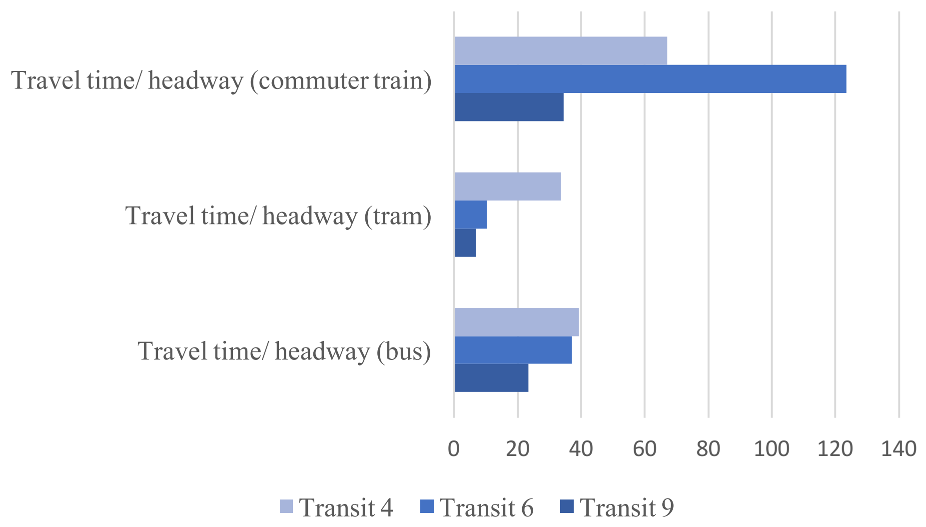


Figure 26 Travel time/ transfer time parameter ratios

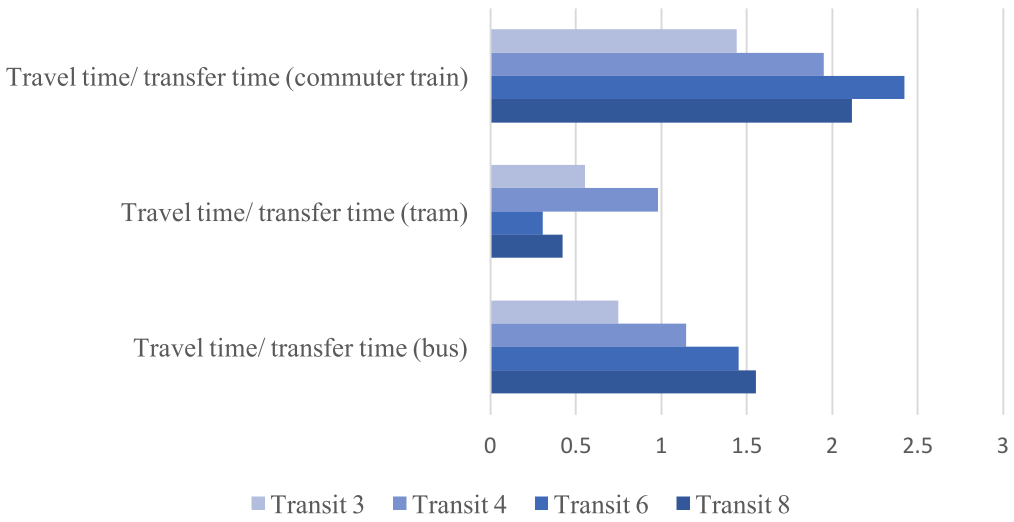
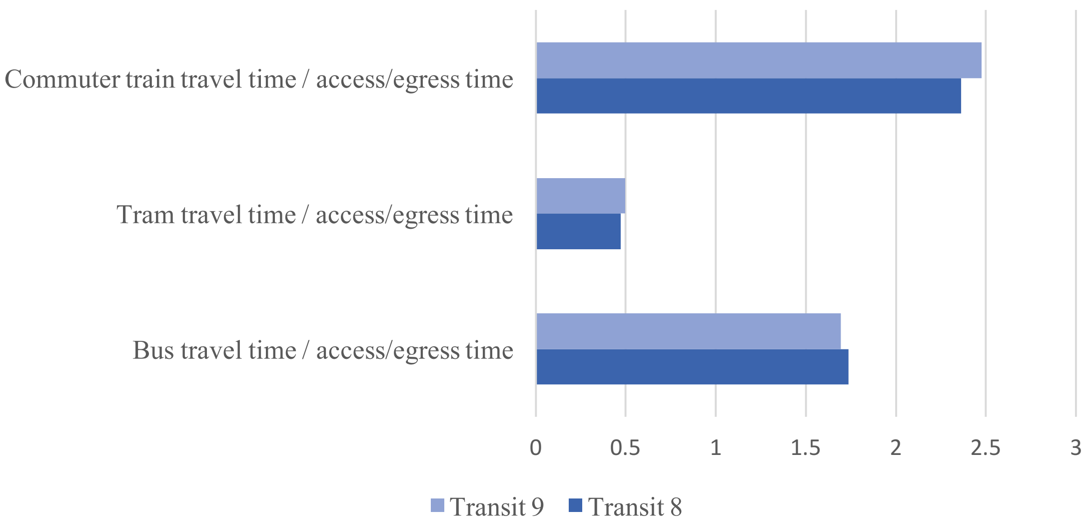


Figure 27 Travel time/ access/egress time parameter ratios



The results above are satisfactory as far as parameter estimates are significant and parameter signs are in tune with expectations. The exception to this rule are the estimated headway parameters in the transit models 5 and 6. It is not expected that travelers value an increase in headways positively. This and the low t-test values indicate that these variables should not be considered in further model specifications. A possible explanation for these results is that individuals do not consider headways for their route decision making. This probably occurs because the average headways are rather short in Zurich (Figure 15) and because people trust and orient themselves on the transit schedule. The very existence of a choice when travelling by transit is not always a given. Often enough, there is only one service between two points in the transit network available and the transit traveler will make use of this alternative regardless of its attributes, since he or she do not have a choice once transit is chosen as a mode.

Figures 23-26 provide a comparison between the different models through parameter ratios. Surprisingly, travelling by tram is not valued as high as travelling by bus (Figure 22). This result is in tune with model estimates by Zimmerman et al. (2017b) who uses GPS traces to estimate a transit recursive logit model for Zurich. The ratios for transit models 3 and 4 are close to 1. This corroborates findings from Scherer (2011) who finds that the Swiss population has a similar image of buses and trams. On the other hand, this author asserts that only frequent transit travelers judge trams more positively than buses. Models 3 and 4 as well as the other ones are consistent in estimating higher travel time parameters for buses than for trams though. The differences in operational speeds of these modes provides some explanation. Trams in Zurich run at average operational speeds of 15.4 km/h (VBZ, 2014) which is the lowest operational speed for a tram network in the country. Trolley buses and diesel buses have, respectively average operational speeds of 18.7 km/h and 19.7 km/h in Zurich's transit network (VBZ, 2014).

The higher bus/tram travel time ratio in models 6, 8 and 9 has other reasons. The comparison between these models and transit models 3 and 4 show that adding mode-specific link dummies in the model more than doubles the bus/tram travel time ratio. The meaning of this result is possibly related to the fact that tram trips are shorter than bus trips, i.e. less tram links are travelled through in tram trips than in bus trips, leading to a lower value of the tram dummy (model transit 6, 8 and 9) as well as biasing the travel time parameter for trams downwards. The comparison of bus/train and bus/commuter train travel time ratios show that these remain relatively stable across model estimates when compared to the bus/tram ratio. As expected, travel times with the two heavy rail modes are valued higher than with the bus. But the same issue with dummies arises when comparing the two heavy rail modes. The train/commuter train travel time parameter ratio (not shown in Figure 23) is 1.2 in the models transit 3 and 4. For the model transit 8 and the same ratio assumes the value 0.75, therefore inversing the preferences. The reason for this is related to longer paths, calculated by the amount of links per path, for commuter train trips than for train trips. There are just 4 edges for trains in the network (Zürich HB-Oerlikon, Zürich HB- Altstetten, Oerlikon-Airport, Zürich HB-Airport) and a train trip in the network consists, at most, of 3 edges.

The same effects can be observed when comparing transit travel time with the headway parameter (Figure 24). For the model transit 4, the ratio uses the headway parameter for all observations, while for the two others it makes use of the mode specific headway and travel time parameters. Here again a similar value is obtained for bus and tram for the model without link dummies, but very divergent results are obtained when comparing ratios of models 6 and 9, which do include link dummies. An interesting result though is that for commuter trains, headways play a less important role than for trams and buses. This is expected since commuter trains are mostly used because of their faster speeds and higher comfort.

Figure 25 compares travel time parameters to transfer time parameters. Commuter train is the only mode where for all evaluated models, the travel time has a larger value than transfer time. For bus, the ratio is consistently larger than for trams and lower than for commuter rail. The results for the transit model 3 for bus and trams are corroborated in the literature. Iseki et al. (2006) provide a literature review on the valuation of transit transfer time valuations compared to in-vehicle times. Most studies do indeed find a higher valuation for transfers than for in-vehicle times. As Iseki et al. (2006) note, the valuation of a transfer is not only dependent on the transfer time itself, but sometimes even more on the transfer facility, and the mode interchange performed. Here, there was no differentiation between the type of modal interchange and this differentiation would probably give a more detailed insight into transfer valuations. The higher transfer value for commuter trains is not expected. A model without a transfer dummy should be estimated to assess the real valuation differences between transfer and travel times to remove the bias from the low amount of chosen transfer links in the path. As a matter of comparison, Schmutz (2015) finds that transfer times have parameters ranging from 29 to 144 times the value of in-vehicle time parameters.

The last parameter ratios presented are between travel times and access/ egress times for the models estimated on the supernetwork. These remain stable between the two compared models. Both models are very similar, the only difference being the headway parameter in the transit model 9, which model 8 does not. Here the ratios for bus and commuter train are also inconsistent with expectations. Access and egress times are expected to be more onerous than in-vehicle travel times. This is corroborated by the results of the estimated MNL mode choice models. Schmutz (2015) also finds such a relation in his MNL models, although with a higher ratio. While the access/ in vehicle time ratio found here is of 1.7, Schmutz (2015) found one of 2.5. Again, the unrealistic results from the RL models here are possibly caused by the inclusion of mode-specific specific dummies. Bovy (2003) argues that it is expected to observe higher time parameter estimates for access than for egress times. This is found in the MNL estimated here, which shows that access times are perceived as 44% more onerous than egress times. Other factors not considered here are environmental (the walking facilities around a station for example) and personal attributes (such as the preference for walking). These play a large role in explaining the valuation of access and egress times (Krygsman et al., 2004).

While model estimates are satisfactory, the experience obtained here clearly shows that the inclusion of mode specific link dummies biases the model results and therefore the parameter estimates of link attributes. The trade-offs between travel times for different modes are still captured by the models, though. Especially the higher valuation of travel times by heavy rail modes is in tune with expectations. Andersen et al. (2017) show that the literature is unanimous in reporting higher travel time parameters for rail based modes. The higher stability of operation, comfort and shorter travel times, given the exclusive right-of-way of heavy rail based modes, increase their valuation compared to modes which interact with street traffic. Andersen et al. (2017) also find that including headways increases the explanatory power of multimodal transit models. This could not be verified here since no attention was given to parameters of model fit to the data. Nonetheless, the results presented here suggest that the headway variable should be included in the model, without interacting with link dummies, if at all. This assures a higher variance of the headway variable across all the different modes. As already discussed, there are enough reasons to believe that headways do not play a major role for mode choice in Zurich.

Unfortunately, only a few model specifications could be tested, given the extremely long computational time required for the recursive logit model. This is also the first time that such a model is estimated with such a high number of observations. As presented in chapter 2.5, the recursive logit modelling framework also allows for the estimation of models with a link size attribute, which corrects choice probabilities based on overlapping links in different observations, as well as a nested recursive logit model, which allows to account for similarities among alternatives. It would be very interesting to relax the IID assumption of the classic recursive logit model, to evaluate different nesting structures. The nesting of bus and tram links can lead to results which are more in tune with common expectations regarding the travel time ratios of these modes.

5 Conclusion

This thesis presented the first implementation of a multimodal recursive logit route choice model. Furthermore, a methodology was presented to construct a static transit network for model estimation based on GTFS data and the Conveyal R5 router. This network was connected to a street network as well. Since the R5 router and Graphhopper are open-source, and the code used for routing and constructing the network is made available upon request, this methodology can be reproduced for estimating route choice models virtually anywhere. Still there are important improvements which can be made, regarding the construction of the supernetwork.

As reported, link parameters (namely travel times and headways) cannot be read out from the network with the presented code since they produce minmum travel times and average headways across all courses for each route, that is for an entire year. This issue is easy to overcome though by routing a virtual path between each pair of nodes in the network and then assigning these for each link pair. This gives rise to a possible inconsistency, since the travel times in all links are related to the time of the day declared in the virtual routing to obtain link parameters and the routed path observations from the Microcensus are not necessarily at the same time. While travel times are not expected to change significantly between each pair of stations, it was shown that headways are rather stable across the day as well.

The street network employed here was the MATSim network for Zurich. This was mainly due to time reasons, since in the scope of a master thesis it was not possible to develop, test and deploy the pseudo algorithm presented in chapter 3.2.2 for post-processing the street network from R5. As reported, the joining of separate stages from multimodal paths was not free of problems. It would be simpler and more consistent to route all the observations in the same routing engine also using a detailed street network and to simplify the network after the observations were routed. While the R5 router does construct a supernetwork from an OSM map and GTFS feed, it does not allow for route requests to contain multiple stages and to assign different modes for each stage. Therefore, stages would have to be routed separately in R5 and then virtually joined as well. It is interesting to examine the capabilities of Open Trip Planner (2018) with respect to the desired functionalities.

While the estimated mode choice models give results for the entire Microcensus data similar to those found by Schmutz (2015), the subdivision of data by trip length, purpose and area highlight the challenges of working with RP data and the care that must be taken when interpreting the results of such models. Nonetheless, travel time parameter estimates from these models do reflect expected behavior.

The first recursive logit estimations are for unimodal car trips. The estimated models show that results from the recursive logit model are not dependent on the size of the network. A second and more important insight from these estimations concerns the importance of having direct observations for estimating route choice models. This is specially a concern when working with RP data stemming from surveys as in this project. It can be concluded that sparser verification points for a route will lead to parameter estimates further away from the real ones, leading to wrong conclusions on trade-offs and welfare measures. This is the reason, why unimodal bike and walk route choice models were not estimated here. For trips with these modes, the only path data points available in the Microcensus are start and end coordinates.

This issue is minimized when estimating route choice on transit networks though, since the route taken between two points in the network is predetermined by the available transit services. The only reported data needed here are entry and exit points in the transit network, transfer stops (if any) and the chosen mode, since geographical entry and exit stations can overlap with different modes (as many tram and bus lines do overlap in several edges in Zurich).

The recursive logit transit model results are also satisfactory and show that trade-offs between different modes, usually captured by mode choice models can be estimated with this modelling framework. The results also provide the lesson that mode specific link dummies bias the parameter estimates. Nonetheless, even for the model without mode specific link dummies, some parameter ratios differ from common expectations. Another interesting finding is the higher valuation of travel time by bus than by tram. A possible reason for this result is a correlation between tram and bus. To evaluate this, a nested recursive logit model would have to be estimated though.

The models estimated here only used link travel times and headways as explanatory parameters for traveler's utility. These two variables alone are not the only variables playing a role in multimodal decision-making though. One the one hand, a recursive logit framework will not be able to include all the possible decision making variables. Mode choice decision-making for example is not a trip by trip decision as modelled here but a long term one, with different decisions regarding budget, mobility tools ownership and location choice being some of them. On the other hand, some influential attributes are not link additive and cannot be modelled into the utility function. Such attributes are link dummies for separate modes as learned here or attributes evaluating transfer facilities for example. Hoongedoorn-Lanser (2005) reports that waiting times and access/egress walking times as being the most important ones for multimodal transit trips. While the second parameter is captured by the multimodal model presented here, the first one can only be captured in a diachronic supernetwork.

As suggested by Fiorenzo-Catalano (2007) personal attributes play a more important role than network attributes for multimodal route choice. Not much literature has been dedicated to evaluating this claim, but the recursive logit model does allow for the inclusion of decision-maker specific attributes in its model. It would also be interesting to test whether further link additive attributes such as the reliability of a transit service, average delay and the accessibility of a link on the network play a role for multimodal route decision making.

The models estimated in this project show that the methodology employed does generate significant and consistent parameter estimates, representing trade-offs between different modes. These are the first models of this kind using the recursive logit framework. Many more model specifications, with more variables and with personal attributes as well as link size and nested recursive logit formulations should be tested to evaluate the quality of the recursive logit model for multimodal route choice analysis. Still, the results presented here show that a static transit network combined with a street network can provide a good basis for estimating recursive logit models. This thesis shows that the recursive logit model can simplify the route choice modelling process of multimodal trips as well as provide valuable insights for policy makers and transport planners.

6 References

Anderson, M. K., O. A. Nielsen and C. G. Prato (2017) Multimodal route choice models of public transport passengers in the Greater Copenhagen Area, *European Journal of Transportation and Logistics,* **6**, 221-245.

ARE – Bundesamt für Raumentwicklung (2017) *Analyse der SP-Befragung 2015 zur Verkehrsmodus- und Routenwahl,* Bern.

Axhausen, K.W. (2017) *Struktur und Idee der random utility model – Familie, Messung und Modellierung*, ETH Zürich, Zürich, März 2017.

Batley, R. (2008) On ordinal utility, cardinal utility and random utility, *Theory and Decision,* **64**, 37-63.

Beck, M. J., S. Hess, M. Ojeda Cabral and I. Dubernet (2017) Valuing travel time savings: A case of short-term or long term choices? *Transportation Research Part E,* **100**, 133-143.

Bekhor, S., M. E. Ben-Akiva and Ramming, M. S. (2006a) Evaluation of choice set generation algorithms for route choice models, *Annals of Operations Research,* **144** (1) 235–247.

Bekhor, S., T. Toledo and J. N. Prashker (2006b) Implementation issues of route choice models in path-based algorithms*, Proceedings of the 11th International Conference on Travel Behavior Research*, Kyoto, Japan.

Bellman, R. E (1957) *Dynamic Programming*. Princeton University Press, Princeton, New Jersey.

Ben-Akiva M. E., M. Scott Ramming and S. Bekhor (2004) Route Choice Models, in M. Schreckenberg and R. Seltern (eds.) *Human Behavior and Traffic Networks,* 23-45, Springer, Heidelberg.

Ben-Akiva, M.E., M. J. Bergman, A. J. Daly and R. Ramaswamy (1984) Modeling inter- urban route choice behaviour, in J. Volmuller and R. Hamerslag (eds.), *Proceedings of the 9th International Symposium on Transportation and Traffic Theory*, 299-330, VNU Science Press, Utrecht.

Ben-Akiva, M. E. and S. Lerman (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge.

Ben-Akiva, M., D. McFadden, K. Train, J. Walker, C. Bhat, M. Bierlaire, D. Bolduc, A. Börsch-Supan, D. Brownstone and D. S. Bunch (2002) Hybrid choice models: Progress and challenges, *Marketing Letters,* **13** (3) 163–175.

Bierlaire, M. and E. Frejinger (2005) Route choice models with subpath components, Paper Presentation *5th Swiss Transport Research Conference,* Ascona, March 2005.

Biswas, D. (2009) The effects of option framing on consumer choices: Making decisions in rational versus experiential processing modes, *Journal of Consumer Behaviour,* **8**, 284-299.

Bliemer, M. C. J. and P.H.L. Bovy (2008) Impact of route choice set on route choice probabilities, *Transportation Research Record,* **2076**, 10-19.

Bovy, P.H.L., R. Uges, and S. Hoongedoorn-Lanser (2003) Modelling route choice behaviour in multimodal transport networks, Paper presentation *10th International Conference on Travel Behaviour Resarch,* Lucerne, August 2003.

Bovy P. H. L. (2009) On Modelling Route Choice Sets in Transportation Networks: A Synthesis, *Transportation Reviews,* **29** (1) 43-68.

Boyce, D. and H. Williams (2016) *Forecasting Urban Travel: Past, present and future*, Edward Elgar Publishing, Cambridge.

Bundesamt für Statistik (2018) *Microrecensement mobilité et transports 2015: Rapport méthodologique: pland'échantillonage, taux de response et ponderation*, Neuchâtel.

Bundesamt für Statistik and Bundesamt für Raumentwicklung (2017) *Verkehrsverhalten der Bevölkerung - Ergebnisse des Mikrozensus Mobilität und Verkehr 2015*, Neuchâtel and Bern.

Bösch, P., K. Müller and F. Ciari (2016) The IVT 2015 Baseline Scenario, Paper presentation *16th Swiss Transport Research Conference,* Ascona, May 2016.

Cascetta, E. (2009). *Transportation Systems Analysis: Models and Applications*, 2nd edition, Springer, Berlin.

Cascetta, E., A. Nuzzolo, F. Russo and A. Vitetta (1996) A modified logit route choice model overcoming path overlapping problems: specification and some calibration results for interurban networks, in J.B. Lesort (ed.), *Proceedings of the Thirteenth International Symposium on Transportation and Traffic Theory*, 697-711, Pergamon, Lyon.

Cascetta, E. and A. Papola (2001) Random utility models with implicit availability perception of choice travel for the simulation of travel demand, *Transportation Research Part C,* **9** (4) 249-263.

Conveyal R5 (2017) Open source R5 router code, Website, https://github.com/conveyal/r5.

Daganzo, C.F. and Y. Sheffi (1977) On stochastic models of traffic assignment, *Transportation Science,* **11**, 253-274.

De la Barra, T., B. Perez and J. Anez (1993) Multidimensional path search and assignment, *Proceedings of the 21st PTRC Summer Annual Meeting,* Manchester, September 1993.

Dijkstra, E. W. (1959) A note on two problems in connection with graphs, *Numerical mathematics,* **1**, 269-271.

Dugge, B. (2006) *Ein simultanes Erzeugungs-, Verteilungs-, Aufteilungs- und Routenwahlmodell*, Dissertation, Technical University Dresden.

Fiebig, D. G., M. P. Keane, J. Louviere and N. Wasi (2010) The Generalized Multinomial LogitModel: Accounting for Scale and Coefficient Heterogeneity, *Marketing Science,* **9** (3) 393-421.

Fiorenzo-Catalano, M. S. (2007) *Choice Set Generation in Multi-Modal Transportation Networks,* Dissertation, Delft University of Technology.

Flötteröd, G. and M. Bierlaire (2013) Metropolis-Hastings sampling of paths, *Transportation Research Part B,* **48***,* 53-66.

Fosgerau, M., K. Hjorth, and S. V. Lyk-Jensen (2007) *The danish value of time study*, Danmarks Transport Forskning Report.

Fosgerau, M., E. Frejinger and A. Karlstrom (2013) An edge based network route choice model with unrestricted choice set, *Transportation Research Part B: Methodological,* **56**, 70–80.

Frejinger, E., M. Bierlaire, and M. Ben-Akiva (2009) Sampling of alternatives for route choice modelling, *Transportation Research Part B: Methodological,* **43** (10) 984–994.

Frejinger, E. and M. Bierlaire (2007) Capturing correlation with subnetworks in route choice models, *Transportation Research Part B,* **41** (3) 363-378.

Friedrich M., I. Hofsass and S. Wekeck (2001) Timetable-based transit assignment using branch and bound*, Transportation Research Record,* **1752***,* 100-107.

Gao, S. and I. Chabini (2006) Optimal routing policy problems in stochastic time-dependent networks*, Transportation Research Part B: Methodological*, **40** (2) 93–122.

Gentile, G., Y. Hamdouch and M. Friedrich (2016) Schedule-Based Assignment on Transit Space-Time Networks, in G. Gentile and K. Noekel (eds.), *Modelling Transport Passenger Flows in the Era of Intelligent Transport Systems*, 334-352, Springer, Berlin.

Henley, D. H., I. P. Levin, J. Louvière and R. J. Meyer (1981) Changes in perceived travel cost and time for the work trip during a period of increasing gasoline costs, *Transportation,* **10***,* 23-34.

Hensher, D. A. (1994) Stated preference analysis of travel choices: the state of practice, *Transportation,* **21***,* 107-133.

Hess, S., A. Erath and K. W. Axhausen (2007) Reducing bias in value of time estimates by joint estimation on multiple datasets: A case study in Switzerland, *Proceedings of the European Transport Conference 2007,* Leiden, October 2007.

Hoongedoorn-Lanser, S. (2005) *Modelling Travel Behaviour in Multi-modal Networks,* Dissertation, Delft University of Technology.

Horni, A., K. Nagel and K. W. Axhausen (eds.) *2016 The Multi-Agent Transport Simulation MATSim,* Ubiquity Press, London.

Iseki, H., B. D. Taylor and M. Miller (2006) The effects of out-of-vehicle time on travel behaviour: Implications for transit transfer, Deliverable 1 for the project *Tool Development to Evaluate the Performance of Intermodal Connectivity (EPIC) to Improve Public Transportation*, University of California, Los Angeles.

Ivehammar, P. and J. Holmgren (2015) The relation between perceived and actual private travel costs – A key question for efficient modal split*, Proceedings of the European Transport Conference 2015*, Frankfurt, September 2015.

Jansen, G.R.M. and D. N. Den Adel (1985) *Routekeuze van automobilisten: een onderzoek naar kwalitatieve keuzfactoren,* Delft University Press, Delft.

Kazagli, E., M. Bierlaire and G. Flötteröd (2016) Revisiting the route choice problem: A modelling framework based on mental representations, *Journal of Choice Modelling,* **19***,* 1-23.

Keane, M. (2006) The generalized logit model: Preliminary ideas on a research program, Paper presentation *Motorola CenSoC Hong-Kong meeting,* Hong Kong, October 2006.

Krygsman, S., M. Djist and T. Arentze (2004) Multimodal public transport: an analysis of travel time elements and the interconnectivity ratio, *Transport Policy,* **11**, 265-275.

Lew, A. and H. Mauch (2007) *Dynamic Programming: A computational tool,* Springer, Heidelberg.

Louviere, J. J. D. A. Hensher and J. D. Swait (2000) *Stated choice methods: Analysis and applications,* Cambridge University Press, Cambridge.

Mai, T., M. Fosgerau, and E. Frejinger (2015a) A nested recursive logit model for route choice analysis, *Transportation Research Part B: Methodological,* **75***,* 100–112.

Mai, T., F. Bastin, E. Frejinger (2015b) A decomposition method for estimating recursive logit based route choice models*, CIRRELT Working Paper*, **2015-66**, Montréal.

Marzano, V. and A. Papola (2004) A edge-based path-multilevel logit model for route choice which allows implicit path enumeration*, Proceedings of the European Transport Conference 2004*, Strasbourg, October 2004.

McFadden, D. D. (1975) *Urban Travel Demand: A Behavioral Analysis*, North Holland Publishing.

Montini, L., C. Antoniou and K. W. Axhausen (2017) Route and mode choice using GPS data, Paper presentation *96th Annual Meeting of the Transportation Research Board*, Washington D.C., January 2017.

Molloy, J., C. Tchervenkov and K. W. Axhausen (2017) Estimating the externalities of a sustainable mobility platform using GPS traces, Paper submitted to *mobil.TUM Conference 2018,* Munich, June 2018.

Newman, P. and J. Kenworthy (1999) *Sustainability and Cities: Overcoming Automobile Dependence*, Island Press, Washington D. C.

Nielsen, O. A. (2000) A stochastic transit assignment model considering differences in passenger's utility functions, *Transportation Research Part B: Methodological,* **34** (5) 377-402.

Papinski, D., D. M. Scott, and S. T. Doherty (2009) Exploring the route choice decision-making process: A comparison of planned and observed routes obtained using person-based GPS, *Transportation Research Part F: Traffic Psychology and Behaviour*, **12** (4) 347–358.

Park, D. and L. R. Rilett (1997) Identifying multiple and reasonable paths in transportation networks: a heuristic approach, *Transportation Research Record,* **1607***,* 31-37.

Prato, C.G. and S. Bekhor (2006) Applying branch & bound technique to route choice set generation, *Transportation Research Record,* **1985***,* 19-28.

Prato, C.G. and S. Bekhor (2007) Modeling route choice behavior: how relevant is the choice set composition?, *Transportation Research Record,* **2041**, 64-73.

Prato, C. G. (2009) Route choice modeling: Past, present and future research directions, *Journal of Choice Modelling*, **2** (1) 65–100.

Prashker, J.N. and S. Bekhor (1998) Investigation of stochastic network loading procedures, *Transportation Research Record,* **1645**, 94-102.

Rabin, M. (1998) *Psychology and Economics*, Journal of Economic Literature, **36** (1) 11-46.

Ramming, S. (2002) *Network knowledge and route choice*, Dissertation, Massachusetts Institute of Technology.

Raveau, S., O. Larsen, K. Jansson (2016) Mode and Route Choice, in G. Gentile and K. Noekel (eds.) *Modelling Transport Passenger Flows in the Era of Intelligent Transport Systems*, 215-231, Springer, Berlin.

Richardson, A. (1982) Search models and choice set generation, *Transportation Research Part A: Policy and Practice,* **16** (5-6) 403-419.

Rieser-Schüssler, N., M. Balmer and K.W. Axhausen (2013) Route choice sets for very high-resolution data, *Transportmetrica A: Transport Science,* **9** (9) 825-845.

Rieser-Schüssler, N., M. Rieser, L. Montini and K.W. Axhausen (2014) Exploring choice set generation approaches for public transport connection choices, *Arbeitsberichte Verkehrs- und Raumplanung,* **1023***,* IVT, ETH Zurich, Zurich.

Scherer, M. (2011) The image of bus and tram: first results, Paper presentation *11th Swiss Transport Research Conference*, Ascona, March 2011.

Schmutz, S. (2015) *Effect of analytical units and aggregation rules on mode choice model,* Master thesis, Institute for Transportation Planning and Systems, ETH Zurich.

Schüssler, N. and K.W. Axhausen (2009) Processing GPS raw data without additional information*, Transportation Research Record,* **2105**, 28-36.

Sheffi, Y. (1985) *Urban transportation network - Equilibrium analysis with mathematical programming methods,* Prentice Hall, New York.

Soto, R., B. Crawford, R. Olivares, R. Herrera, F. Johnson, and F. Paredes (2015) Enumeration Strategies for Solving Constraint Satisfaction Problems: A Performance Evaluation, in R. Silhavy, R. Senkerik, Z. K. Oplatkova, Z. Prokopova, and P. Silhavy (eds.), *Artificial Intelligence Perspectives and Applications: Proceedings of the 4th Computer Science On-line Conference 2015 (CSOC2015), Vol 1: Artificial Intelligence Perspectives and Applications*, 169–179, Springer, Berlin.

Stadt Zürich (2017) *Stadtverkehr 2025: Bericht 2016,* Tiefbauamt, Stadt Zürich.

Stern, E. and D. Leiser (1988) Levels of spatial knowledge and urban travel modeling, *Geographical Analysis,* **20** (2) 140-155.

Thurstone, L. (1927) A law of comparative judgement, *Psychological Review,* **34***,* 273-286.

Train, K. (2009) *Discrete Choice Methods with Simulation*, 2nd edition, Cambridge, New York.

Tversky, A. (1972) Elimination by aspects: A theory of choice, *Psychological Review,* **201**, 269-283.

Open Street Maps (2017) Website, https://www.openstreetmap.org.

Open Trip Planner (2018) An open-source multimodal trip planning tool, Website, http://www.opentripplanner.org.

VBZ – Verkehrsbetriebe Zürich (2014) *Bericht zur Servicequalität und Nachfrage auf dem Netz der Verkehrsbetriebe Zürich,* Zürich.

Vrtic, M. (2003) *Simultanes Routen- und Verkehrsmittelwahlmodell,* Dissertation, Technical University Dresden.

Vrtic, M., N. Schuessler, A. Erath and K. W. Axhausen (2010) The impacts of road pricing on route and mode choice behaviour, *Journal of Choice Modelling*, **3** (1) 109–126.

VöV (2017) *T-600 Allgemeiner Personentarif,* Verband öffentlicher Verkehr, Bern, 2017.

Web of Science (2018) Metadata search on mode and route choice publications, Website, http://apps.webofknowledge.com.

Williams, H. C. W. L. (1977) On the formation of travel demand models and economic valuation measures of user benefit, *Environment and Planning A,* **9** (23) 285-344.

Wittinik, L. T. (2011) *Choice Modelling: An overview of theory and development in individual choice behaviour modelling,* BMI Paper, Vrije Unversiteit Amsterdam, Amsterdam.

Zantema J., D. H. Almersforst, M.C.J. Bliemer, P.H.L. Bovy and D. Busche (2007) Route set generation within a dynamic modeling framework: a comparison of methods and GPS route choice data*, Proceedings of the European Transport Conference 2007,* Leiden, October 2007.

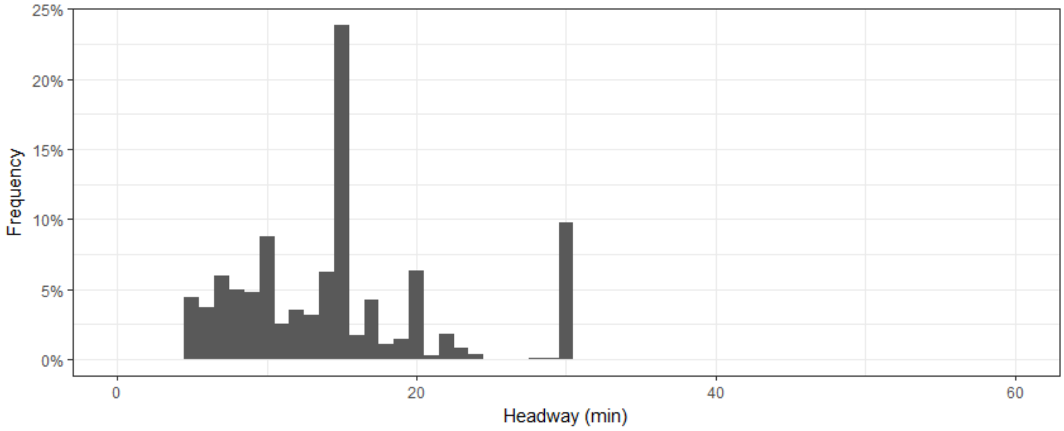
Zimmerman, M., T. Mai and E. Frejinger (2017a) Bike route choice modelling using GPS data without choice sets of paths, *Transportation Research Part C: Emerging Technologies,* **75**, 183-196.

Zimmerman, M., K. W. Axhausen and E. Frejinger (2017b) Multi-modal route choice modeling in a dynamic schedule-based transit network*,* Submitted to IATBR, Santa Barbara, July 2018.

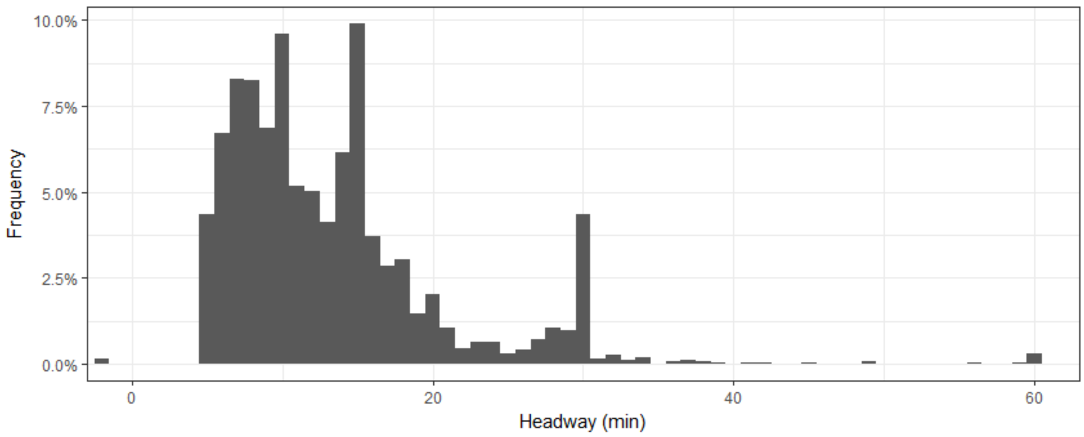
Appendix

A 1 Headways by mode

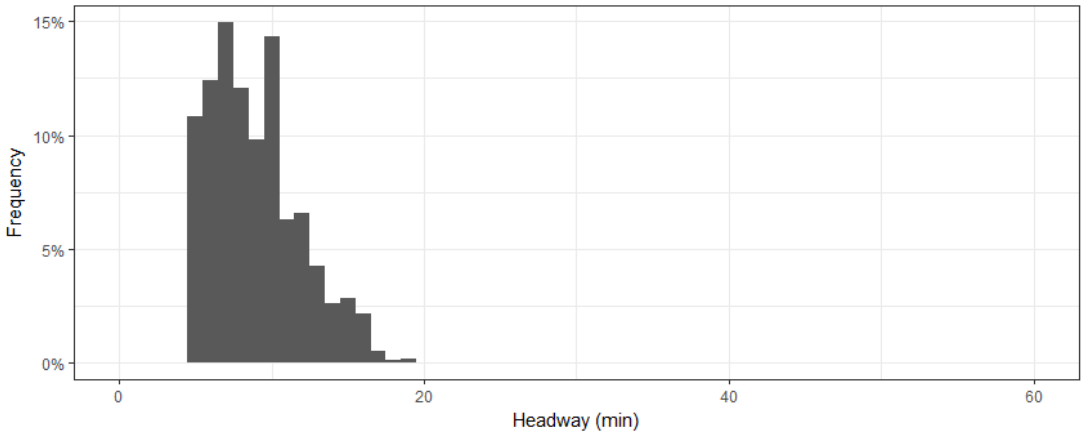
Commuter rail (S-Bahn) Headways (mean=14.6 minutes).



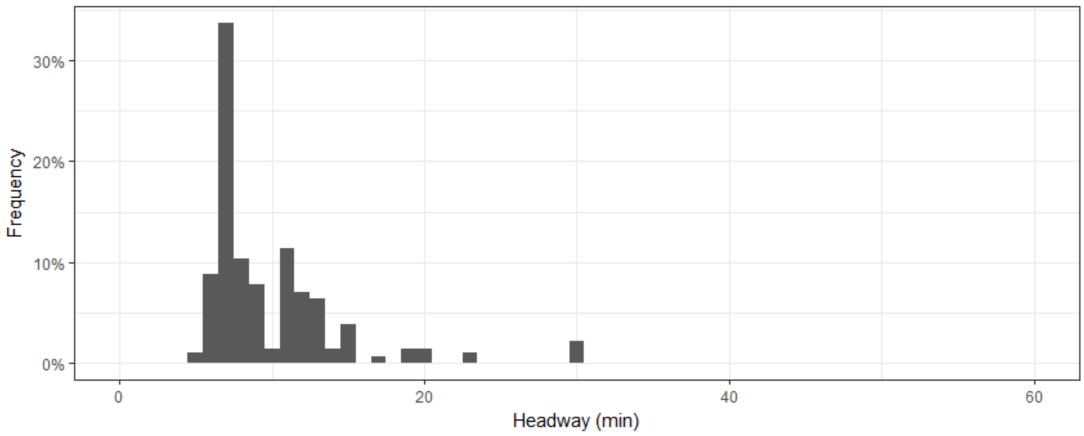
Bus Headways (mean=13.4 minutes).



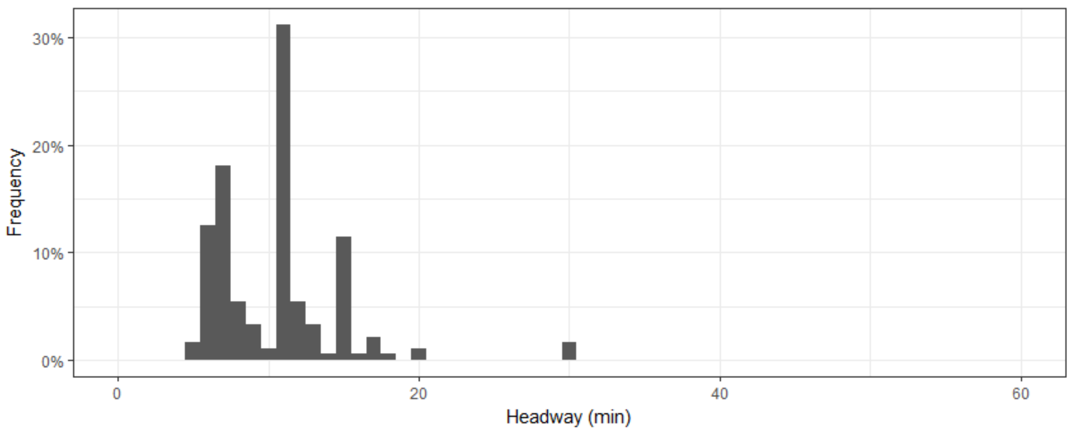
Tram Headways (mean=8.9 minutes).



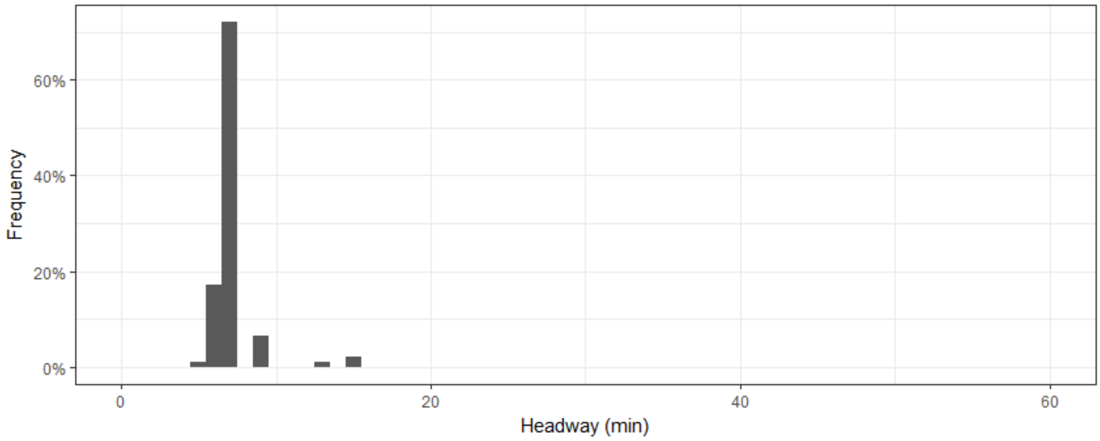
Interregional Train Headways (mean=9.9 minutes).



Regional Train Headways (mean=10.5 minutes).



Long distance train Headways (mean headway=7.1 minutes).



A 2 Mode choice models

	All	Non-leisure	Leisure	Short dist.	Medium dist.	Long dist.	Urban travel	Zurich
ASC Bike	-0.944 ***	-0.696 ***	-0.872 ***	-0.914 ***	-3.873 ***		-0.315 ***	-0.7986 ***
ASC Transit	-1.534 ***	-1.64 ***	-1.346 ***	-2.162 ***	-2.123 ***	-0.034 ***	-1.756 ***	-6.1561 ***
ASC Walk	1.867 ***	2.294 ***	1.686 ***	1.827 ***	-6.298 ***		2.608 ***	2.3396 ***
Travel time car (min)	-0.095 ***	-0.119 ***	-0.074 ***	-0.151 ***	-0.131 ***	-0.058 ***	-0.181 ***	-0.1673 ***
Cost (CHF)	-0.086 ***	-0.138 ***	-0.047 ***	0.09 ***	-0.004	0.023 ***	0.194 ***	1.2267 ***
Travel time walk (min)	-0.14 ***	-0.184 ***	-0.107 ***	-0.151 ***	-0.018 ***		-0.186 ***	-0.1376 ***
Travel time bike (min)	-0.128 ***	-0.175 ***	-0.103 ***	-0.161 ***	-0.055 ***		-0.221 ***	-0.1174 ***
Transit in-vehicle time (mir	-0.04 ***	-0.039 ***	-0.039 ***	-0.029 ***	-0.064 ***	-0.055 ***	0.001	-0.0363 ***
Transit access time (min)	-0.063 ***	-0.062 ***	-0.069 ***	-0.086 ***	-0.059 ***	-0.043 ***	-0.114 ***	-0.0625 ***
Total transit time (min)	-0.033 ***	-0.051 ***	-0.018 ***	-0.084 ***	-0.045 ***	-0.012 ***	-0.121 ***	
Transit egress time (min)	-0.054 ***	-0.053 ***	-0.058 ***	-0.076 ***	-0.048 ***	-0.03 ***	-0.097 ***	-0.0434 ***
Observations:	124306	66311	47788	96908	21992	5405	26655	5564
Final log likelihood	-69751	-36276	-28094	-55604	-9291	-2656	-17476	-4809
Rho2:	0.43	0.43	0.43	0.43	0.56	0.29	0.35	0.33
AICc:	139525	72573	56210	111230	18604	5326	34975	9637
VoT car (CHF/h)	66	52	94.5	-100.6	1829.3	-148.2	-56	-8.2
VoT transit (CHF/h)	27.8	16.9	50.1	-19.2	886	-142.1	0.2	-0.03

Confidence levels: ***>99%, **>95%, *>90%

Short distance: <20km

Medium distance: >20, <40 km

Long distance: >40km