Exploring choice set generation approaches for public transport connection choice

Author(s):
Rieser-Schüssler, Nadine; Rieser, Marcel; Montini, Lara; Axhausen, Kay W. (iD)

Exploring choice set generation approaches for public transport connection choice

Nadine Rieser-Schüssler

Marcel Rieser

Lara Montini

Kay W. Axhausen

ETH

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Working paper   1023

Exploring choice set generation approaches for public transport connection choice

Nadine Rieser-Schüssler   Marcel Rieser   Lara Montini
IVT   Senozon AG   IVT
ETH Zürich   Schaffhauserstrasse 331   ETH Zürich
CH-8093 Zürich   CH-8050 Zürich   CH-8093 Zürich
phone: +41-44-633 30 85   phone: +41-44-520 14 62   phone: +41-44-633 30 88
fax: +41-44-633 10 57   fax: +41-44-633 10 57
nadine.rieser@ivt.baug.ethz.ch   rieser@senozon.com   lara.montini@ivt.baug.ethz.ch

Kay W. Axhausen
IVT
ETH Zürich
CH-8093 Zürich
phone: +41-44-633 39 43
fax: +41-44-633 10 57
axhausen@ivt.baug.ethz.ch

August 2014

Abstract

This technical working paper documents the choice set generation algorithms for public transport connection choice in their state of development in August 2014. The choice set generation methods have been developed to generate the non-chosen alternatives for public transport trips that have been observed through GPS traces recorded either by dedicated GPS devices or by smart-phones.

The choice set generation approaches that we have developed build on the idea of generating a routing network that contains the entire schedule and modifying that Dijkstra algorithm so that takes into account multiple start and end stops in the public transport network and the corresponding access and egress segments. The basic version is then combined with other least cost path based choice set generation procedures: First, with a procedure to randomly vary the link cost and cost function parameters to obtain a doubly stochastic choice set generation and, second, with a new approach of selecting potential transfer opportunities in the network enforcing a routing via these selected points.

A first evaluation of the generated choice sets was done with public transport trips found in GPS traces originating from a person-based GPS diary survey with participants living in and around Zurich. It shows that all three approaches generally work. However, it also reveals several areas for improvement that are discussed within in this working paper

Keywords

GPS, public transport, choice set generation, schedule, network

Preferred citation style

Rieser-Schüssler, N., M. Rieser, L. Montini and K.W. Axhausen (2014) Exploring choice set generation approaches for public transport connection choice, *Working paper*, **1023**, Institute for Transport Planning and Systems (IVT), ETH Zurich, Zurich.

1 Introduction

This technical working paper documents the choice set generation algorithms for public transport connection choice in their state of development in August 2014. The choice set generation methods have been developed to generate the non-chosen alternatives for public transport trips that have been observed through GPS traces recorded either by dedicated GPS devices or by smart-phones. The goal of the envisaged models is to gain a better understanding of passengers' evaluation of the different elements of a public transport trip which is crucial for the investigation of new strategies for innovative public transport systems.

Generating choice sets for public transport trips is a complex topic. An appropriate procedure has to account for spatial as well as temporal components, several dimensions of similarities and has to explicitly include the access and egress segments of the trip (Bovy, 2009). Due to these obstacles only few approaches to generate full size timetable-based connection choice sets have been presented in the literature so far. Prevalent are branch & bound based approaches such as the ones presented by Friedrich et al. (2001) and Hoogendoorn-Lanser et al. (2006). A different approach is to generate a public transport routing network that does not only represent spatial but also temporal aspects of the public transport network and use least-cost path based choice set generation procedures on this new network. The choice set generation procedures presented by Nielsen (2000) and Fiorenzo-Catalano (2007) use this idea. However, they do not use the exact timetables but average values such as average frequencies and transfers times which makes it difficult to account for schedule differences in the course of the day.

The choice set generation approaches that we have developed and that are described in the Sections 2, 3 and 4 build on the work by Rieser (2010) who developed a procedure to generate a routing network that contains the entire schedule and a modified Dijkstra algorithm that takes into account multiple start and end stops in the public transport network and the corresponding access and egress segments. As already discussed in Rieser-Schüssler et al. (2012), the advantages of this approach are that the cost function for the router can be specified individually by the analyst and that the basic version described in Section 2 can be combined straightforwardly with other least cost path based choice set generation procedures. We have done this by combining it first with a procedure to randomly vary the link cost and cost function parameters to obtain a doubly stochastic choice set generation as described in Section 3 and second with a new approach of selecting potential transfer opportunities in the network enforcing a routing via these selected points. This approach is discussed in Section 4.

For a first evaluation of these approaches choice sets are generated with each method for public transport trips found in GPS traces originating from a person-based GPS diary survey with

participants living in and around Zurich. The participants were asked to carry a person-based GPS receiver for a week and to fill out three psychometric scales and a household and person questionnaire on-line. Moreover, they confirm and correct the results generated by the automatic post-processing of their GPS tracks in a web-based prompted recall survey. The public transport trips are extracted using automatic mode detection and a procedure called public transport map-matching that is described in Rieser-Schüssler et al. (2012).

The evaluation of the generated choice sets in Section 5 shows that all three approaches generally work. However, it also reveals several areas for improvement that are discussed within the section and summarised in Section 6.

At the point of writing, the code for the choice set generation algorithms is not yet publicly available since it contains code that is owned by the Senozon AG. If you are interested in using it, please contact us under nadine.rieser@alumni.ethz.ch or rieser@senozon.com.

2 The basic approach for the choice set generation

Currently, three approaches for generating public transport connection choice sets have been implemented:

- the basic public transport choice set generation (**Basic PT CSG**),
- the doubly stochastic choice set generation (**DSCSG**) and
- the via point choice set generation (**Via Point CSG**)

All of them are least cost path based approaches and based on the same fundamental idea. The Basic PT CSG has already been presented in Rieser-Schüssler et al. (2012), but since the DSCSG and the Via Point CSG are extensions of the Basic PT CSG, it is shortly summarised in this section. Subsequently, this section discusses the filtering criteria that are used to clean the results of all three algorithms from unrealistic connections.

2.1 The Basic PT CSG

As discussed in Rieser-Schüssler et al. (2012), the underlying idea for our choice set generation approaches is to generate a *routing network* that does not only represent spatial but also temporal aspects of the public transport network and schedule and then use least cost path based choice set generation procedures on this routing network. The routing network is generated with the approach by Rieser (2010). The routing network is generated by first creating a node for each stop in the schedule. Then, the nodes are connected by links that represent the sequence in which the stops are served for each line and course. The travel times on these links are taken directly from the schedule. Third, transfer opportunities to other lines are added by connecting each node with all the nodes within a configurable distance. The travel mode on these links is walk and the travel time is an estimation of the walking times. In addition, other cost components can be assigned to these links such as additional transfer penalties. Subsequently, heuristics are used to reduce the network size: Nodes that are start locations of a public transport line are defined as departure locations only. Thus no transfer links can start there. The reverse applies to nodes that are only arrival locations of a public transport line. In addition, no transfer links will be added between nodes that belong to different courses or directions of the same line at the same stop.

The algorithm used for the least cost path calculation is an adapted version of the least cost path algorithm presented by Rieser (2010). The performance of the original algorithm was improved and it was opened up to give more than just one alternative per origin destination pair. For this, the algorithm starts by identifying all nodes in the routing network that are within acceptable

walk radius around the origin and destination of the trip and stores them as the potential start and end nodes for the routing. Then, the algorithm determines the least cost route for each pair of start and end nodes.

The threshold for the acceptable walking distance is configurable by the analyst along with the parameters of the cost function used by the router. The least cost path algorithm can cope with any kind of additive cost function so it is up to the analyst to define the most suitable one. In this implementation, the cost function is a weighted linear combination of

- in-vehicle travel time
- access and egress walk times
- transfer and waiting times
- flat penalty for doing a transfer

In addition, the least cost path search is constrained by a minimum time for each transfer.

2.2 Filtering

The main goal of the filtering is to remove redundant and unrealistic connections from the choice set. However, as can be seen in the results section, some of these criteria are not ideal and should be reviewed carefully before the filtering is applied to future data sets.

For the removal of redundant connections, connections are removed if there is another connection that

- serves the same stop sequence
- serves the same line sequence
- starts or ends at a "better" stop served by the same line

In the first two cases the connections are compared and the connection with the shorter overall travel time is retained in the choice set and the other one is discarded. The last case can occur because several stops of the same line can be within the acceptable walking distance around the origin or the destination. If this is the case the connection with the longer access or egress walk time is removed. However, while being intuitive this filter turned out to remove the chosen route in a few cases. Therefore, it should be reviewed and adapted before future applications.

For the removal of unrealistic connections, a connection is removed if

- it does not contain public transport stage

- it contains unnecessary line changes
- its overall travel time is 3 times higher than that of the fastest connection

The filtering for unnecessary line changes actually consists of two filters: A filter for line changes at a potential start or end stop around the origin or the destination and a filter for unnecessary line changes during the trip. The filter for unnecessary line changes at a potential start or end stop captures cases where the line change only occurs due to the requirement that the algorithm searches for a connection between each pair of start and end stops and the line used to reach these stops ends at one of these stops. An example occurring in the test data is when the trip origin or destination is ETH Hoenggerberg. Potential start or end stops for these trips are the stops "ETH Hoenggerberg" and "Lerchenhalde". Since routing to both stops is enforced there are cases where the connection would entails getting off the 69 bus at ETH Hoenggerberg and waiting for the next 80 or 37 to get to Lerchenhalde just to walk back to ETH Hoenggerberg. These cases are removed by this filter. The second unnecessary line change filter is designed to find connections that contain a transfer even though the current line serves one of the subsequent transfer stops or one of the potential end stops of the trip. While these filters are helpful in eliminating unrealistic connections in many cases, it still has to be examined if there might be cases where meaningful connections are falsely removed.

The filter based on the overall travel time is the most problematic one. As can be seen in the results section, this filter imposes a constraint that strongly influences the composition of the choice set. Moreover, it biases a parameter that should be determined by the model. Therefore, the current version of this filter is not recommendable and it should definitely be rethought before any future application.

During the filtering there is also a small postprocessing step for connections that contain lines which do not serve the same sequence of stops in both directions. There it can occur that it is reasonable to board the vehicle on the way to the end point of the line and remain in the vehicle at the end point. Due to the way the schedule data is stored, this remaining in the vehicle would constitute a transfer in the unprocessed results while it is not from a passenger's perspective. Therefore, the number of transfers, transfer waiting times and in-vehicle times are adapted to reflect this perception.

3 Doubly stochastic choice set generation (DSCSG)

The doubly stochastic choice set generation (DSCSG) is a straight forward extension of the Basic PT CSG similarly to those presented by Bovy and Fiorenzo-Catalano (2007) and Nielsen (2000). The general idea is the same as for car-based route choice set generation procedures: To obtain a choice set the least cost path search between origin and destination is repeated iteratively and at the beginning of each iteration the costs of the network links and/or the parameters of the cost function are varied. The adaptation of the link costs reflects the uncertainty of the costs actually occurring when travelling through the network while the variation of the cost function parameters represents the taste differences between travellers. Both the network link costs as well as the cost function parameters are drawn from probability distributions.

This idea can also be combined with the Basic PT CSG though it has to be set up slightly differently because the Basic PT CSG already produces a set of alternatives in each step. In the implemented version each least cost paths search iteration results in a set of alternatives and the sets are joined together after all iterations are finished. During the joining duplicate connections are sorted out before the filtering described in the previous section is applied.

The variation of the link cost and cost function parameters is done in two hierarchical loops. In the outer loop, the parameters of the cost function are adapted and in the inner loop the link costs are changed. Following the recommendations by Nielsen (2000), the link costs are drawn from a Gamma distribution and the cost parameters are drawn from a log normal distribution. The standard deviations for both probability distributions have to be set by the analyst, the expected value for the Gamma distribution is set to be the original link cost and the expected value of the log normal distribution the value of the cost parameter set by the analyst. The cost parameters are the same as in the Basic PT CSG. In addition, the analyst can specify separately how many times the cost parameters should be varied and how many times for each cost parameter setting the link costs are drawn.

4 Via point choice set generation (Via Point CSG)

The via point choice set generation (Via Point CSG) is a completely new approach. The underlying assumption is that travellers often first consider which points in the network their route may pass by and only then select specific lines or connections. For example for a trip from ETH Hönggerberg to Zurich HB, a traveller might consider travelling via Bucheggplatz, via Meierhofplatz or via Oerlikon. The goal is to represent this process in the choice set generation approach and thus achieve a more meaningful kind of variation. The drawback, however, is that there is little research available on how these via points are chosen and how they influence the choice set of the travellers. Therefore, the approach is implemented in a modular way that allows the analyst to test different hypotheses and investigate this research question further.

The framework for the choice set generation approach using via points consists of several steps. As in the DSCSG, first several sets of alternatives are generated that are joined in the end taking into account duplicate connections before the final filtering is carried out. Based on this the main steps are:

- Run the basic PT CSG without consideration of via points to find direct trips
- Determine the set of potential via points
- Repeat until the number of iterations specified by the analyst is met:
    - Select a via point
    - Generate a set of routes that connect all combinations of potential start and end points and use the selected via point
    - Remove the via point from the set of potential via points before drawing the next one
- Join the generated sets of alternatives to one set
- Apply filtering as described in Section 2.2

The three crucial steps of this approach are the determination of the set of potential via points, the selection of the via point used in a particular iteration and the routing process itself. Therefore, these three steps are discussed in more detail below. Particularly the steps concerning the selection of vial points are the ones where several assumptions about the traveller's behaviour have to be made. The current version works with simplistic assumptions about consideration distances and relative weighing of the via points based on characteristics of the stops derived from the schedule. The main topic for future research would be to test the behavioural realism of these assumptions and revise the selection processes.

First, however, the concept of a transfer set is introduced. A transfer set is a set of stops between which transfers are possible and that are often perceived by passengers as one stop area.

Prominent examples for this in the Zurich network are Zurich HB, the stops around Central or the stops associated with any of the S-Bahn stations. In our approach transfer sets are used as the basis for the via point selection instead of stops because it makes the routing more efficient (e.g. by not repeatedly routing via one of the stops of Zurich HB or avoiding routing via a train stop for a trip that reasonably only contains bus and tram) and is more in line with how travellers perceive the public transport network.

4.1 Generation of transfer sets

The assignment of stops to a particular transfer set is not done by the algorithm but has to be specified in an input file. The choice set generation package does, however, contain a script to create this assignment semi-automatically. In the current version, it is possible to combine a manual assignment with an automatic assignment using a distance threshold. The manual option was created for transfer sets like Zurich HB where the individual stops might be further away than the distance threshold. Given this input, the script first generates the transfer set for these manual stops. Then, it checks for all remaining stops in the network if there are other stops within a distance threshold of 100 metres. If this is the case, these stops are assigned to one transfer set. It is important to note that for subsequent stops it is sufficient to be within the distance threshold of one of the stops in a transfer set. After the distance based transfer sets are created, transfer sets are created for all remaining stops with each transfer set containing just one stop.

4.2 Determination of the set of potential transfer sets

Currently, a rather simplistic approach is used for the determination of potential transfer sets. For this, the mid-point between origin and destination is calculated and all transfer sets within a circle around this mid-point are considered to be potential transfer sets. The most difficult issue here is to determine the appropriate radius for this circle which has to be specified by the analyst. So far only a few test runs have been conducted during which a radius of 1.2 times the distance between the mid-point and the origin (or destination) looked most promising. However, more systematic testing is needed before any recommendations can be made.

4.3 Selection of a transfer set

The selection of the specific transfer set for an iteration from the set of potential transfer sets is done using a weighted random draw. Therefore, each transfer set is assigned a score that is based on the characteristics of the transfer sets derived from the schedule. A few different formulations for the score calculations have been tested – although not systematically. Currently, the transfer set score $score_{TS}$ accounts for the number of lines per mode and the number of departures of a transfer set:

$$score_{TS} = \beta_{lineVSDep} * score_{lines} + nofDep/nofServH \tag{1}$$

where $\beta_{lineVSDep}$ is a parameter that weighs $score_{lines}$ – the score derived from the characteristics of the lines serving the transfer set – against the ratio between $nofDep$, the number of departures for the whole transfer set, and $nofServH$ the number of hours the transfer set is served by at least one line. $score_{lines}$ is defined as the weighted sum of the number of train lines ($nofTrainLines$), number of tram lines ($nofTramLines$) and the number of bus lines ($nofBusLines$) serving the transfer set. Each line is only counted once even if it serves more than one stop of the transfer set.

$$score_{lines} = \beta_{trainLines}*nofTrainLines+\beta_{tramLines}*nofTramLines+\beta_{busLines}*nofBusLines \tag{2}$$

The weighting parameters $\beta_{trainLines}$, $\beta_{tramLines}$ and $\beta_{busLines}$ as well as the parameter $\beta_{lineVSDep}$ have to be specified by the analyst. The current parameter setting is reported in Section 5.1.2. However, more systematic testing is required. In addition, other formulations should be tested, maybe adding attributes that are not derived from the schedule. Ideally, the selection process should model the perception of travellers. To derive this model, either a new survey or additional analyses of the Swiss Microcensus and the GPS traces should be done.

4.4 Routing via transfer sets

To consider the transfer sets in the routing the router used in the Basic PT CSG had to be adapted. The new routing is done in three steps:

1. Routing from the origin to the transfer set
2. Routing from the transfer set to the destination
3. Joining of the routes

Similar to the Basic PT CSG routing, the routing from the origin to the transfer set starts with determining the potential start stops around the origin. Then, the least cost path between each of these stops and the stops of the transfer set is calculated. Thereby, the router stops when it reaches at least one of the stops of the transfer set and determines that no better connection can be found to any remaining stop of the transfer set. The result is one route alternative for each potential start stop.

In the second step, for each of these routes the subsequent routes to each of the potential end stops around the destination are calculated. This is done taking into account the arrival time of the origin-based alternative at the specific stop of the transfer set as well as the transfer opportunities between the stops in the transfer set and the walking times required to realise these transfer opportunities. The result is then one route for each origin-based connection part and each potential end stop.

In the third step, the origin-based connection parts are joined with their destination-based counterparts. During the joining process it is checked whether a transfer is necessary at the transfer sets. Accordingly, the legs are either joined – in the case of no transfer – or the required waiting times and transfer walk legs are added.

5 Preliminary results

No systematic testing of the algorithms has been done so far. However, for STRC 2014 a few runs have been executed to test the general functionality of the algorithm and detect further areas for improvement. For these tests a preliminary set of algorithm parameters has been determined, which can serve as a start point for future experiments. These parameters are reported in the next subsection along with a short overview of the data used for the tests. Then, a few analyses of the resulting choice sets are discussed in the second subsection.

5.1 Input data and runs

5.1.1 Data

The data used for these tests is the tram and bus sub-sample from the test data used in Rieser-Schüssler et al. (2012). By using the pretest data as well as the data from the main study participants, in total 193 cleaned and double-checked public transport trips made by participants living in and around Zurich were available. There was no need to use the GPS raw data for this study, instead the results of the map-matching described in Rieser-Schüssler et al. (2012) were used directly. If available in the GPS data the trips include the walk access and egress stages. Then the routing is started from the real origin of the trip and not the first public transport stop.

5.1.2 Parameters used

The configuration of the **Basic PT CSG** requires only the parameters for the router as summarised in Table 1. These parameters were taken from the more in-depth parameter testing in Rieser-Schüssler et al. (2012) and remain the same for all choice set generation approaches.

The parameters for the DSCSG are presented in Table 2. Here the analyst has to specify how many times the parameters and link costs are varied and which standard deviations for the distributions are assumed. Note, that due to the hierarchical loop structure the combination of 5 iterations with parameter draws and 5 iterations with link cost draws result in 25 iterations overall because for each of the drawn parameter settings the link costs are drawn five times.

The parameter settings for the Via Point CSG are presented in Table 3. Besides the number of transfer sets or via points used, the analyst has to specify the parameters for the selection of the

Table 1: Routing parameters

Parameter	Value
Maximum walk radius [m]	600
Additional transfer time [sec]	60
Marginal utility of waiting [util/s]	-0.00167
Marginal utility of travel time walk [util/s]	-0.005
Marginal utility of travel time PT [util/s]	-0.00167
Utility of line switching [util]	-1

Table 2: Parameters for the doubly stochastic choice set generation

Parameter	Value
Number of parameter draws	5
Number link cost draws per parameter setting	5
Standard deviation Gamma distribution	1
Standard deviation Log Normal distribution	1

transfer sets. This is the biggest area for testing that remains. The current parameters can be seen as really preliminary.

Table 3: Parameters for the via point choice set generation

Parameter	Value
Number of transfer set draws	5
Radius factor transfer set search	1.2
Transfer set score contribution train lines ($\beta_{trainLines}$)	5
Transfer set score contribution tram lines ($\beta_{tramLines}$)	2
Transfer set score contribution bus lines ($\beta_{busLines}$)	0.5
Transfer set score lines vs departures ($\beta_{lineVSDep}$)	10

5.2 Analyses of the resulting choice sets

The main purpose of the analyses presented in the remainder of this section is to provide a starting point for further improvements of the algorithms. The goal is to show where the algorithms already provide reasonable results and where more work is needed.

The first of these analyses is an overview about the **Number of alternatives** generated by each algorithm depending on the beeline distance between origin and destination that is presented in Figure 1. It can be seen that the Basic CSG and the DSCSG result in very similar choice set sizes while the Via Points CSG often generates more alternatives, particularly for longer distances. This was to be expected since one of the goals of the Via Point CSG is to generate more (meaningful) variation among the generated connections.

Figure 1: Number of alternatives
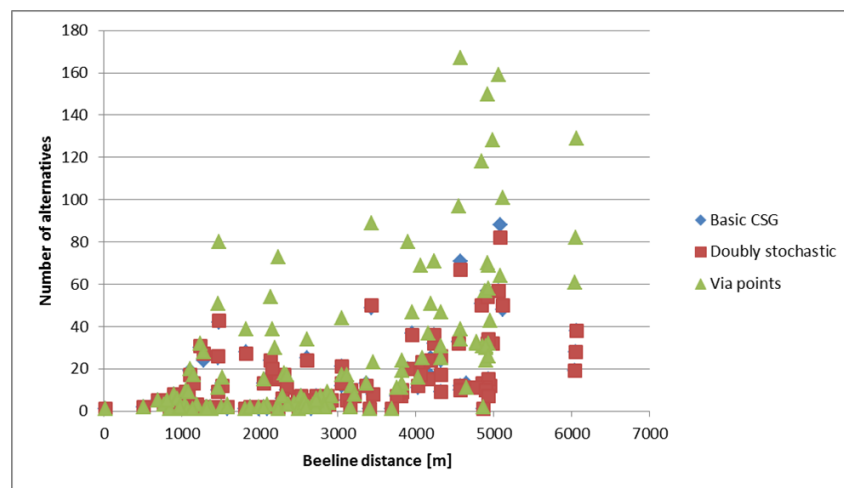
Table 4 shows how often the three algorithms were able to **reconstruct the chosen route**. All three algorithms perform similarly and not very satisfactory. Therefore, one of the next analyses should focus on the trips where the chosen route was not reconstructed to check if there are any systematic issues that could be resolved by an improvement of the algorithm. Moreover, it should be examined if the filtering criteria might also cause some problems here. One indicator for that is that the DSCSG and the Via Point CSG perform even worse than the Basic PT CSG even though the Basic PT CSG is first executed in both algorithms.

Table 4: Reconstruction of the chosen route

Algorithm	Chosen route reconstructed
Basic PT CSG	107 (77%)
DSCG	106 (76%)
Via Point CSG	106 (76%)

One potential issue with regards to the Via Point CSG is that it would result in unrealistically high **numbers of transfers**. However, as can be seen in Figure 2, where the minimum, maximum and average number of transfers is depicted against the beeline distance between origin and

destination, this effect is not so big that it is cause for concern. There are a few connection where the maximum number of transfers is higher than three or where the average number of transfers is slightly higher than for the Basic CSG or the DSCSG, particularly for longer distances. Overall, however, the number of transfer patterns appear to be rather similar.
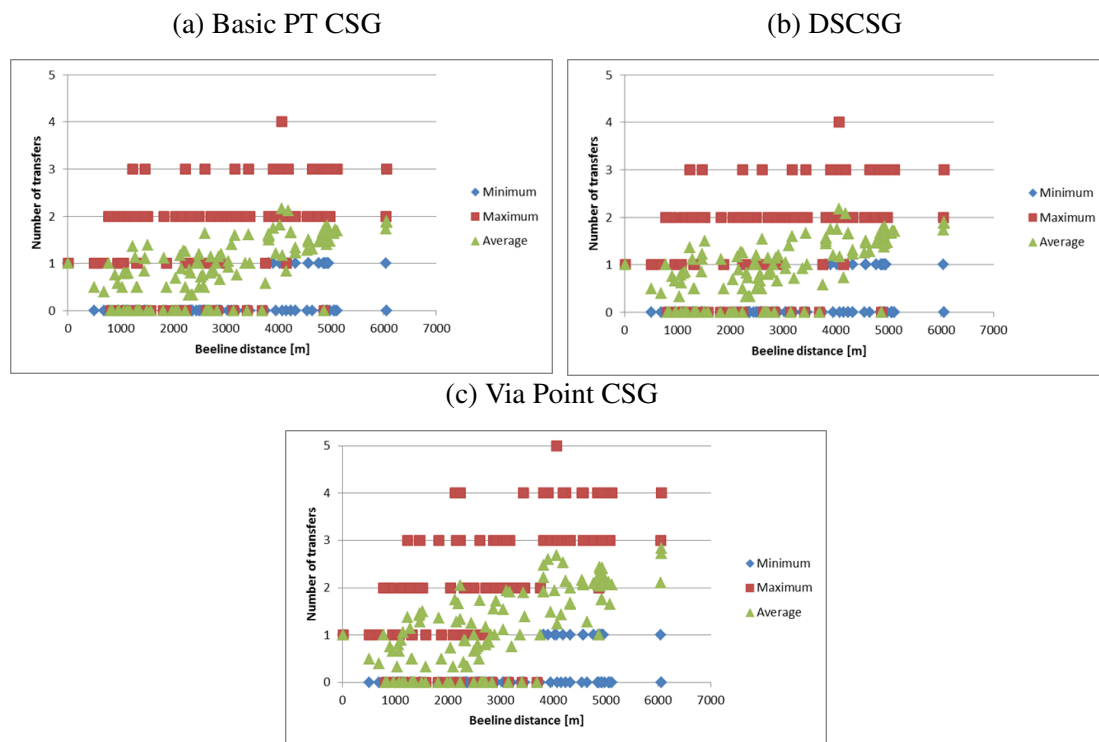
Figure 2: Number of transfers

(a) Basic PT CSG

(b) DSCSG

(c) Via Point CSG

Figure 3 on the other hand reveals a very substantial issue that needs to be addressed. The figure shows the minimum, maximum and average total travel time of the connections in each choice set. Notably the pattern of the maximum travel times shows an effect that is not desirable for the subsequent modelling process and this is the strong influence of the travel time based filter. It can be clearly seen that this filter affects nearly all choice sets and therefore imposes a bias on any choice models that try to estimate the effect of the total travel time on the choice of the travellers. Therefore, it needs to be reconfigured or removed from the set of filtering criteria. Apart from that Figure 3 does not show any systematic issues with any of the choice set generation approaches. The travel time patterns are rather similar even though the Via Points CSG can lead to slightly higher travel times which was expected due to the slightly higher number of transfers.

The final analysis looks at the **Spatial distribution of the connections** generated by the different choice set generation procedures. Figure 4 and Figure 5 show this for two different trips and for all three choice set generation approaches. In both figures, there is no big difference between

Figure 3: Total travel time

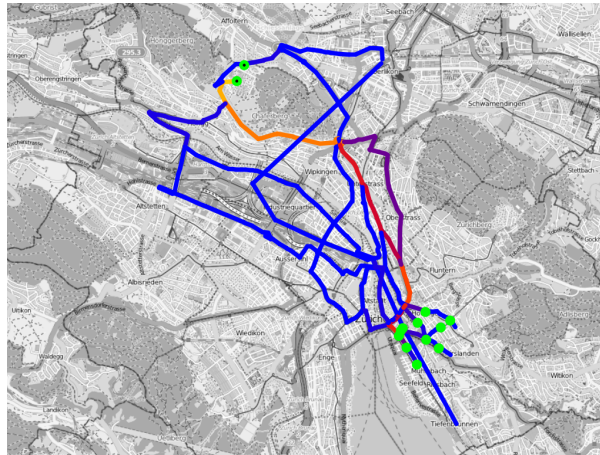(a) Basic PT CSG

(b) DSCSG

(c) Via Point CSG
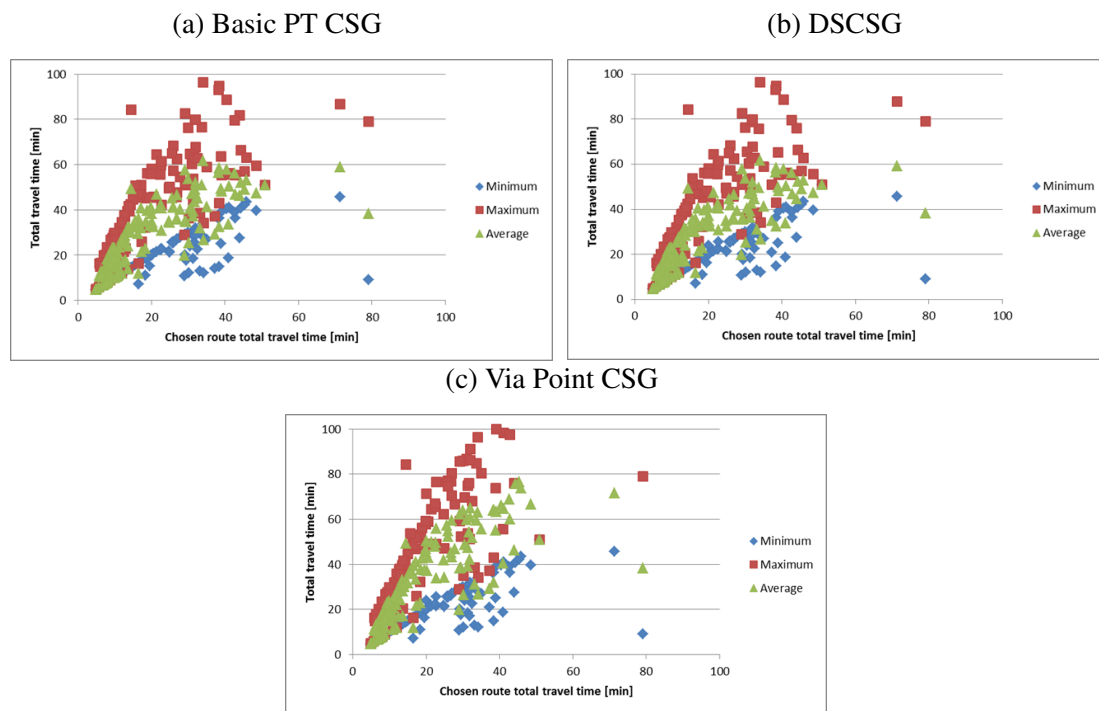
spatial distribution of the connections found by the Basic PT CSG and those found by the DSCSG. This was similar for all other trip examples we looked at. This seems to indicate that the DSCSG does not provide any benefit in terms of more variation compared to the Basic PT CSG, maybe because the variation of the Basic PT CSG is already rather high. Further analysis is necessary to examine this. The analysis should include testing other parameter settings for the DSCSG, in particular, higher standard deviations for the link cost and parameter distributions.

When comparing the spatial distribution of the Basic PT CSG and the Via Point CSG the differences are much more pronounced. How strong they are does, however, depend on the individual trip as can be seen when comparing Figure 4 and Figure 5. As expected the Via Point CSG is suitable to generate more variation but more research and testing is necessary to determine how much variation and what type of variation best represents travellers' different perceptions.

Figure 4: Spatial distribution of the connections - Example trip 1

(a) Basic PT CSG
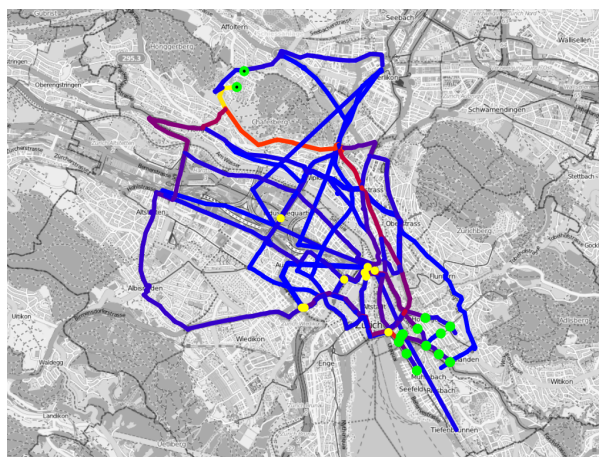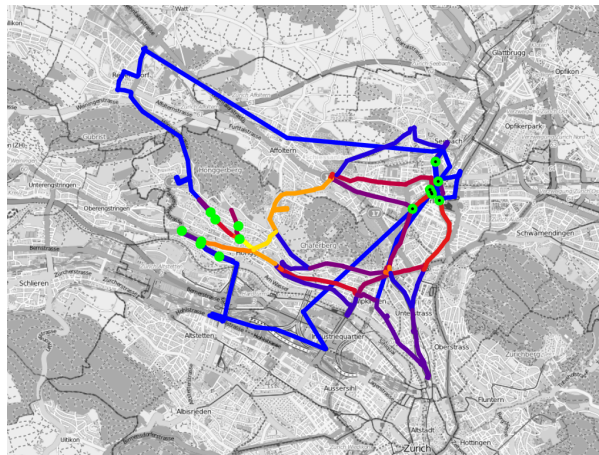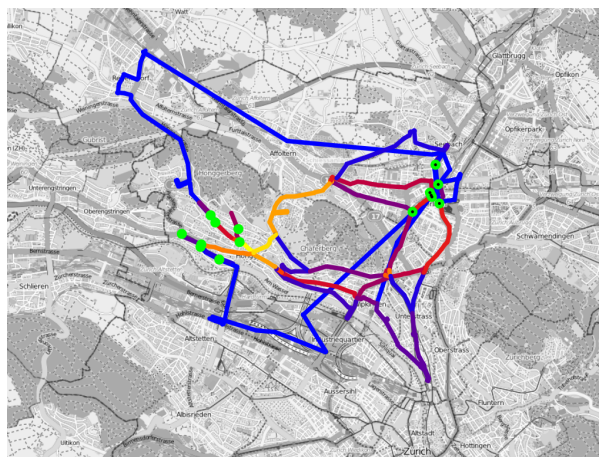

(b) DSCSG
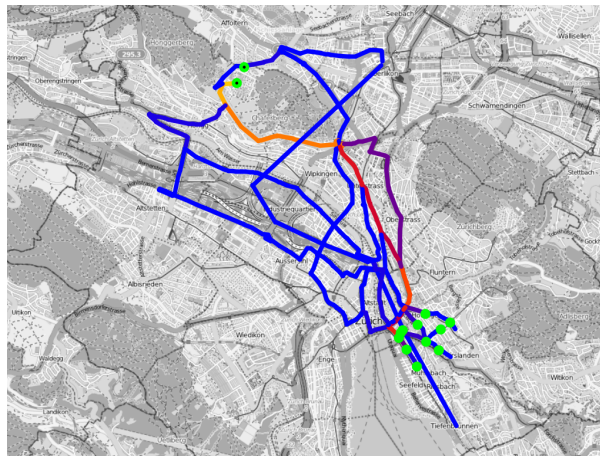
(c) Via Point CSG

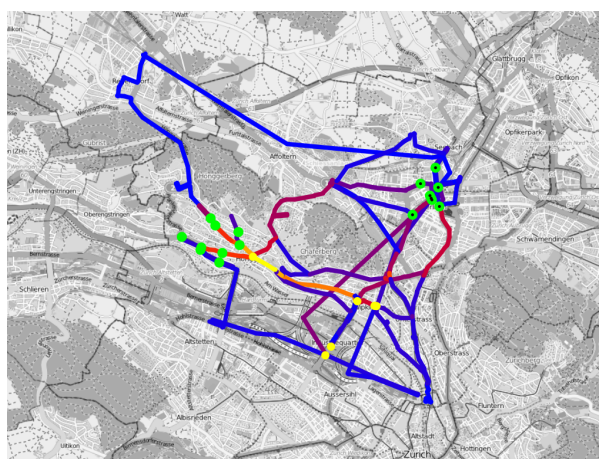Figure 5: Spatial distribution of the connections - Example trip 2

(a) Basic PT CSG

(b) DSCSG

(c) Via Point CSG

6 Conclusion and outlook

This working paper documents the choice set generation algorithms for public transport connection choice in their state of development in August 2014. As the different tests in Section 5

show, the algorithms show a lot of potential but more analysis of the resulting alternative sets is necessary as is more testing of different parameter settings.

In the preliminary tests the Basic PT CSG and the DSCSG produce very similar results. However, no systematic testing of the influence of the parameters settings of the DSCSG – particularly of the standard deviations – have been done so there is still potential for improvement.

The connections generated by the Via Point CSG show more variation, however, the algorithm also produces a larger number of connections that are judged unrealistic by a human observer. Therefore, more examination of the algorithm is required with a focus on the selection process of the via points or transfer sets. The current version uses characteristics of the public transport schedule to guide the selection but is not based on any user behaviour model.

Moreover, the filtering methods currently implemented are not ideal and some of them come with the danger of biasing the subsequent modelling results. Therefore, they should be revisited and adapted accordingly.

The next steps should be:

- Explore why the chosen connections could not be reproduced for certain trips
- Revising the filtering criteria
- Derive a more behaviourally meaningful selection process for the via points
- Systematic parameter testing
- Choice modelling

7 Acknowledgements

8 References

Bovy, P. H. L. (2009) On modelling route choice sets in transportation networks: A synthesis, *Transport Reviews*, **29** (1) 43–68.

Bovy, P. H. L. and S. Fiorenzo-Catalano (2007) Stochastic route choice set generation: Behavioral and probabilistic foundations, *Transportmetrica*, **3** (3) 173–189.

Fiorenzo-Catalano, S. (2007) Choice set generation in multi-modal transportation networks, Ph.D. Thesis, Technical University Delft, Delft.

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

Hoogendoorn-Lanser, S., R. van Nes and P. H. L. Bovy (2006) A rule-based approach to multi-modal choice set generation, paper presented at the *11th International Conference on Travel Behaviour Research (IATBR)*, Kyoto, August 2006.

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

Rieser, M. (2010) Adding transit to an agent-based transportation simulation, Ph.D. Thesis, Technical University Berlin, Berlin.

Rieser-Schüssler, N., L. Montini, M. Rieser and K. W. Axhausen (2012) Preparations for estimating the influence of attitudes on public transport connection choice, paper presented at the *13th International Conference on Travel Behaviour Research (IATBR)*, Toronto, July 2012.