Veröffentlichung / Publication

Route Choice Set Generation in Road Networks

Autoren / Authors:

Juliane Pillat
Eileen Mandir
Markus Friedrich

Lehrstuhl für Verkehrsplanung und Verkehrsleittechnik
Email: vorname.nachname@isv.uni-stuttgart.de

Veröffentlicht in / Published in:

Pillat, J., Mandir, E., Friedrich, M. (2010): Route Choice Set Generation in Road Networks, Proceedings of the 5th International Symposium Networks for Mobility, Stuttgart.

Universität Stuttgart
Institut für Straßen- und Verkehrswesen
Lehrstuhl für Verkehrsplanung und Verkehrsleittechnik
www.isv/uni-stuttgart.de/vuv/

Route Choice Set Generation in Road Networks

Juliane Pillat, Eileen Mandir, Markus Friedrich

Universität Stuttgart
Department for Transport Planning and Traffic Engineering
Pfaffenwaldring 7
70569 Stuttgart, Germany
firstname.lastname@isv.uni-stuttgart.de

Abstract: The paper reports on a method which generates choice sets for commuters in the Munich metropolitan area. The method utilizes GPS trajectories and interview data of 300 participants over an 8 week survey to combine chosen, known and generated routes to choice sets for route choice modelling. The main idea of the method is to use revealed preference routes as well as stated preference routes to calculated accepted detour factors, which are then used as boundary conditions for choice set generation using path enumeration. Based on a spatial choice set, the method generates time-dependent choice sets by attributing all routes with actual travel times at the time travelled.

Key Words: Choice set generation, branch-and-bound, route choice behavior, GPS trajectories, stated preference

1 Proposed Method

Forecasting and modelling the effect of traffic control systems on driver's route choice has been in the centre of research interest since some years. Along with the estimation of route attributes and their impact parameters on driver's utility as well as identification of suitable choice models, a major task is developing methods for choice set generation to capture so called unchosen alternatives. Current heuristic procedures focus on algorithmic performance as well as number of routes in choice sets. However, including observed route choice behaviour and driver's network knowledge in choice set generation has rarely been studied and is the contribution of this work.

The paper will describe the method of choice set generation with path enumeration based on Stated Preference (SP) and GPS data in three chapters:

- Identification of routes from survey data
- Route Generation
- Choice Set Composition

First, SP data is geo-coded and used to generate known activity locations as well as known routes for trips from home to work. Second, GPS data is processed to generate trips from recorded trajectories in order to identify revealed routes as well as additional visited activity locations. In order to generate unchosen and not stated alternative routes in the choice set a NAVTEQ road network of the study area is simplified to a strategical road class level and computational manageable size. On this network a spatial choice set is generated. The choice set generation is done by path enumeration under empirically – from SP data – derived criteria for accepted detour factors of routes in the choice set. All three route sets are then combined to a choice set for an OD pair. Hereafter, the spatial choice set composition is supplemented by dynamic route attributes, such as current travel time, according to the departure time of each particular trip.

Finally, the paper gives a summary of the method, shows statistical results for choice set size and important attributes in driver's route choice. The paper closes with an outlook on future work on choice model estimation.

2 Identification of routes from survey data

2.1 Survey design

The Project *wiki* focuses on route choice behaviour in the major road network. The greater Munich area has a very dense network of motorways, highway and arterials north of the city area and thus is an optimal test bed for observing route choice. Therefore, 300 commuters working in the northern part of Munich city area and living north of the city in the greater metropolitan area were recruited.

In part 1 of the survey, the participants had to fill out a questionnaire to state information about socio-demographic characteristics, car ownership etc. Further, the participants were asked for their commonly frequented activity locations including purpose of activity (leisure, shopping) and the exact address, if possible.

In part 2 of the survey, all participants were equipped with a smart phone and GPS sensor. The GPS sensor calculated a position every second which's latitudinal and longitudinal coordinates, date/time and speed were transmitted to a server via a phone network every five minutes.

In part 3 of the survey, a personal interview, every participant was requested to state his known routes from home to work. The routes were clicked link by link on a digital map.

2.2 Geocoding of activity locations and routes from SP data

After evaluating and validating the questionnaire the activity locations home, working place and usual other activities were geocoded and matched on a high resolution NAVTEQ road network. Activities without a stated purpose were referred to a nearby point of interest to deduce a trip purpose. Points of interests were also utilized to derive the correct activity location if the address was incomplete or missing in the questionnaire.

Furthermore, for every stated route it had to be ensured that origin and destination matched stated home and working place address taken from the questionnaire.

2.3 Identification of activity locations and routes from GPS data

Sending data through the mobile phone network can result in data loss. This happens if the smart phone is switched off at the destination before the current data package is sent or if GSM network connection fails in between a trip. Moreover, GPS files stored on the server include data recorded between switching on and switching off the smart phone. This may include several trips.

The trips are map matched by assigning GPS logs to links and adding route elements to get complete trajectories without gaps. One problem in map matching GPS data on a digital map is that mostly at the trip ends the location cannot be found because minor roads are likely to be missing in the network model (SCHÜSSLER and AXHAUSEN [1]). To avoid these problems the highly resolute NAVTEQ network was used.

Further processing is needed to advance from link trajectories to actual trips. In this context three criteria for trip identifying are deduced:

- Calculation of current speed and smoothed speed → current speed criterion
- Identification of space and time gaps → gap criterion
- Calculation of current detour factor → detour criterion

The first criterion deals with the varying accuracy with which the GPS data is recorded. This accuracy is related to the number of available satellites and ranges between 5 and 10 meters (SCHÜSSLER and AXHAUSEN [1]). Calculating the current speed based on the coordinates results in "jumping" speed values. Hence, a smoothed speed value is calculated using GPS coordinates and timestamps two seconds before and ahead for the current GPS log.

Now time periods with a smoothed speed lower than a certain threshold are determined. If the time period is longer than five minutes, the trip is divided in two parts. Hereby, a detection of trip ends within a continuous GPS track is achieved. The threshold values for the smoothed speed and length of the according time period have to be strict enough to avoid that congested traffic situations or stops due to signal-controlled intersections are not wrongly identified as stops at activity locations and thus trip ends.

The second criterion deals with remaining time and space gaps. For each GPS log the distance x_{gap} to the subsequent GPS log and the difference of the respective time stamps t_{gap} are calculated. If t_{gap} is lower than the threshold t_{acc} the space gap is connected. Otherwise, the GPS track is split. The accepted time t_{acc} is determined as follows:

$$t_{acc} = \begin{cases} t_{crit.low}, & \text{if } t_{gap} \le t_{crit.low} \\ x_{gap} \cdot 1.6, & \text{if } t_{gap} \le t_{crit.high} \\ t_{crit.high}, & \text{if } t_{gap} > t_{crit.high} \end{cases}$$

t_{acc} [s]	accepted time for activity detection
x_{gap} [m]	space gap between two GPS logs
$t_{crit,low}$ [s]	threshold for minimum activity duration
$t_{crit,high}$ [s]	threshold for maximum accepted time gap

The formula implies that for $t_{gap} < t_{crit,low}$ = 300s the space gap is always connected. For $t_{gap} > t_{crit,high}$ = 1200s the trip is always split, because completing a track with gap larger than 20 minutes is not reasonable. Between these two thresholds t_{acc} depends on the space gap x_{gap}. The larger the space gap the larger is t_{acc}. The parameter 1.6 in the formula can be interpreted as threshold for a minimum accepted speed in meters per second.

The last criterion focuses on the current detour factor in relation to the smoothed speed. If the calculated detour factor exceeds a certain value while the current speed value is below a threshold an activity is detected. The combination of critical values for speed and detour factor allows to determine several intermediate activities which are shorter than five minutes, typically pick up and drop off. Figure 1 shows an exemplary trajectory with the time-dependent course of speed and detour factor. The position of the detected activity is marked red.

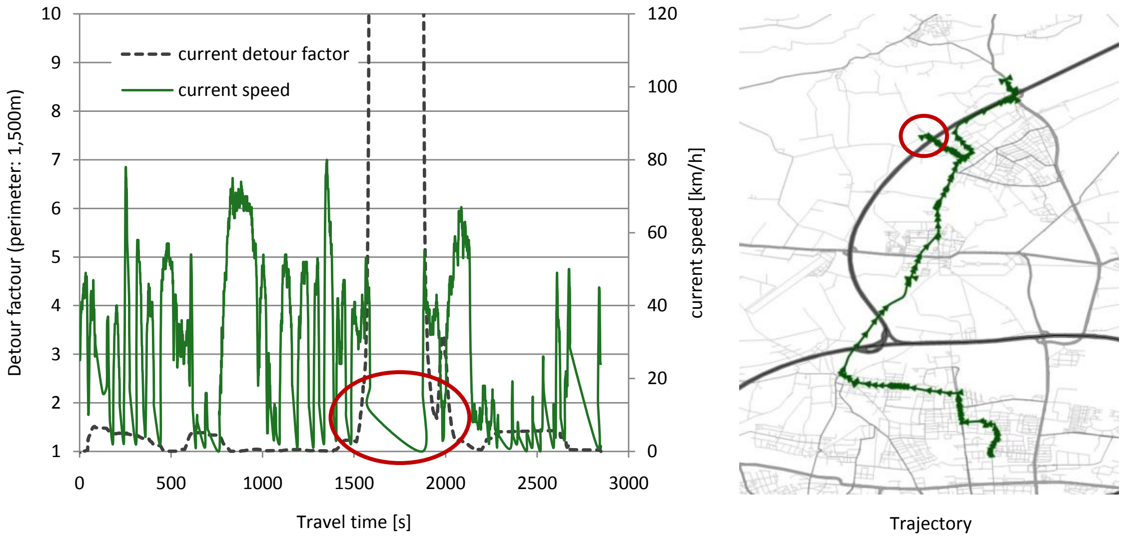

Figure 1: Trajectory with time-dependent course of current speed and detour factor

As a result all routes between two main activities which are significantly influenced by an intermediate activity are split. Table 1 gives an overview of collected data volume and detected trips.

Data Volume		
	Total over 300 participants	Per participant
Total time of detection in hours	8,850	29.5
Number of detected GPS tracks	20,000	66
Number of identified trips	24,000	80
Number of trips between identified activity locations	18,300	61

Table 1: Data volume

Knowing the activity purpose at trip end and beginning is crucial for modeling route choice behavior. Therefore, every stated activity in a perimeter of 2,500 meters is allocated to the trip ends. If a trip end is related to more than one known activity a distinct activity match using arriving and departure time and duration between last and next trip is done. For example, if activities “work” and “leisure” are assigned to a trip ending at 7:05 AM with an activity duration of 8 hours (the time until the next trip begins) the activity “work” is matched to the trip end. Each trip end which was not related to any known activity was clustered in groups with other unspecified trip ends within a perimeter of 2,500 meters. Because of this additional activity locations which were not stated could be derived. Figure 2 shows the chosen routes for a total of 25 identified trips for an example from home to work.

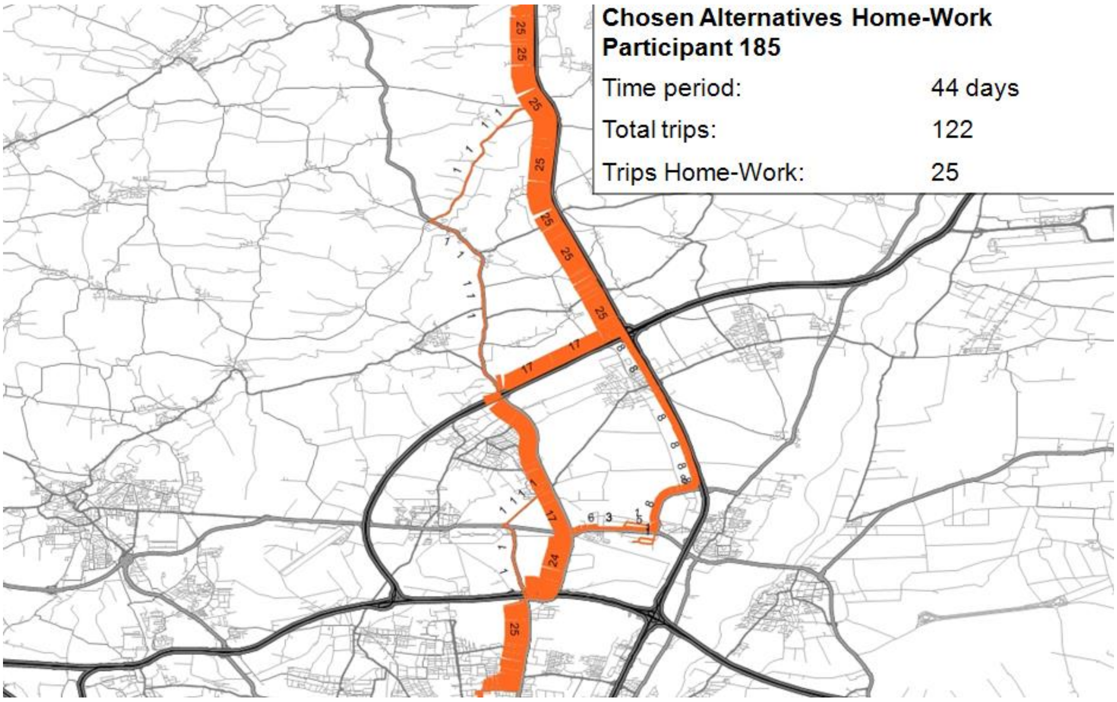

Figure 2: Chosen alternative routes from home to work for one participant

3 Route Generation

Heuristic choice set generators calculate k shortest paths by varying route impedance criteria. The number of routes and resulting impedance detour factors depend on the number of different criteria and iterations. Although, routes exceeding a maximum allowed detour factor can be excluded from the choice set subsequently, it is not possible to consider detour factors on parts of the route rather than the entire route. Choice set generators, which consider detour factors only after the actual generation process, have a major deficit. Maximum allowed detour factors on route level are often chosen to be constant or decreasing with increasing route travel time. Shortcomings of these approaches are that a constant detour factor will result in biased choice sets with more routes for shorter distance OD pairs. A decreasing detour factor solves this problem, yet is unable to exclude long routes with minor and implausible detours along the way from the choice set. Therefore, including criteria for maximum detour factors on route parts within the generation process is desirable.

To include detour factors within the choice set generation process a path enumeration algorithm is needed, where certain branches of the route tree are deleted due to a too high detour factor. Complete enumeration in its classical meaning aims at identifying every possible route from origin to destination by using a directed graph of the road network to build a route tree where the origin is the tree source and the tree branches are the routes in the network. By including branch cutting criteria as the allowed detour factors, this method of choice set generation allows to control the generated routes (branch-and-bound). The resulting routes are highly dependent on the allowed detour factors of the tree branches compared to the corresponding shortest path. For every new node that is added as a branch element to the route tree, the branch cutting criteria are checked. SCHLAICH [2] and [3] developed a method in with three branch cutting criteria are checked for each current node added to a branch end:

- No cycles in route (branch is cut if current node is already element of this branch)
- Maximum detour factor for entire branch (branch is cut if detour factor from origin to current node is larger than a defined maximum difference to shortest path)
- Maximum detour factor for rear part of branch (branch is cut if detour factor from the current node to a node 15 minutes upstream is larger than a defined maximum difference to shortest path)

The more complex the network structure and the more divers the trips for which choice sets need to be generated (short distance trips, long distance trips, inner city trips, motorway commute), it is recommendable to redesign the third branch cutting criteria for the rear part of a branch to match the task, by including more rules for various lengths of rear route sections. For the case study area of Munich, which spans the southern half of Bavaria and includes network parts with less density in rural areas as well as high network density in the Munich city centre, four maximum detour factors were defined and lead to the following function for the allowed impedance from origin (respectively an upstream node in the branch) to current node at the end of the branch:

$$\text{Allowed Impedance} = \min\left[\left\{\begin{matrix} 1.0 \cdot t_{\min}, \text{ if } t_{\min} < 250 \text{ sec} \\ 2.0 \cdot t_{\min}, \text{ if } t_{\min} < 450 \text{ sec} \\ 1.6 \cdot t_{\min}, \text{ if } t_{\min} < 1000 \text{ sec} \\ 1.4 \cdot t_{\min}, \text{ if } t_{\min} \geq 1000 \text{ sec} \end{matrix}\right\}, t_{\min} + 1800 \text{ sec}\right]$$

with t_{min}: travel time of shortest path between upstream node and branch end

As the generated routes depend highly on the defined maximum detour factors, these parameters need to be chosen carefully matching the network topology of the studied area. In this case study the route choice model focuses on the major road network north of the city area. The choice set generator is fitted to produce choice sets which's routes capture the main network loops in this area. From the SP-interview known routes from home to work (via this part of the road network) are available. By comparing known routes which use the same network loops, minimum and maximum travel time within a loop and thus accepted detour factors for each participant of the survey can be determined. Figure 3 shows the detour factor for the analyzed network loops as data points plotted over the minimum travel time in the loop t_{min} against the detour factor within that loop t_{max} / t_{min}. From these empirically derived detour factors the impedance function for route generation is derived. The grey line in Figure 3 shows the t_{min}-sections and corresponding allowed impedances and is chosen to cover most of the survey detour factors except for single extreme values. All survey detour factors below t_{min} = 250 seconds are ignored because route choice for very short distance trips is not in the primary focus of this analysis and allowing detour factors as large as 2.1, as indicated by the survey data points, results in unreasonable large choice set sizes.

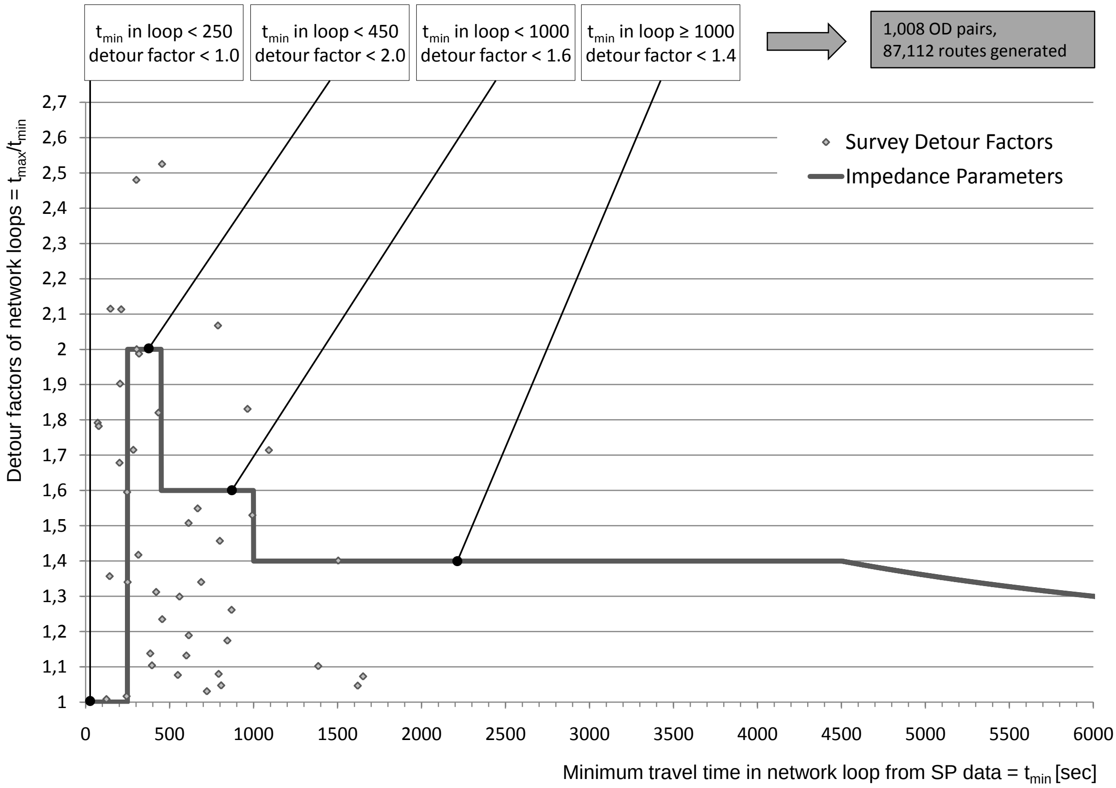

Figure 3: Observed detour factors in network loops and allowed impedance

Figure 4 illustrates a route tree generated by this method for a small example. For each possible route in the network from origin to destination there is a branch. With complete enumeration there would be 8 possible routes (branches), excluding routes with cycles. Every branch with node elements shown with a dashed frame is not part of the generated route tree, because it is subject to one of the branch cutting criteria (cycle in branch or detour factor exceeded). For each current node (the branch end at the current route tree generation step) the impedance of the rear part of the branch is compared to t_{min} of the shortest path between the respective node pair. In Figure 4 this comparison is highlighted, for example, in green for the route section from node 1 to node 2. The shortest path between the two nodes is the direct connection from 1→2 and has a travel time of 10 minutes. The middle branch of the route tree includes the path from 1→3→2 which has a travel time of 12 min. Thus the impedance is lower than the maximum allowed impedance for this $_{tmin}$-section and the branch stays in the route tree. Two examples of branches that are excluded from the route tree because their impedance exceeds the allowed impedance are displays with dashed nodes in red and blue.

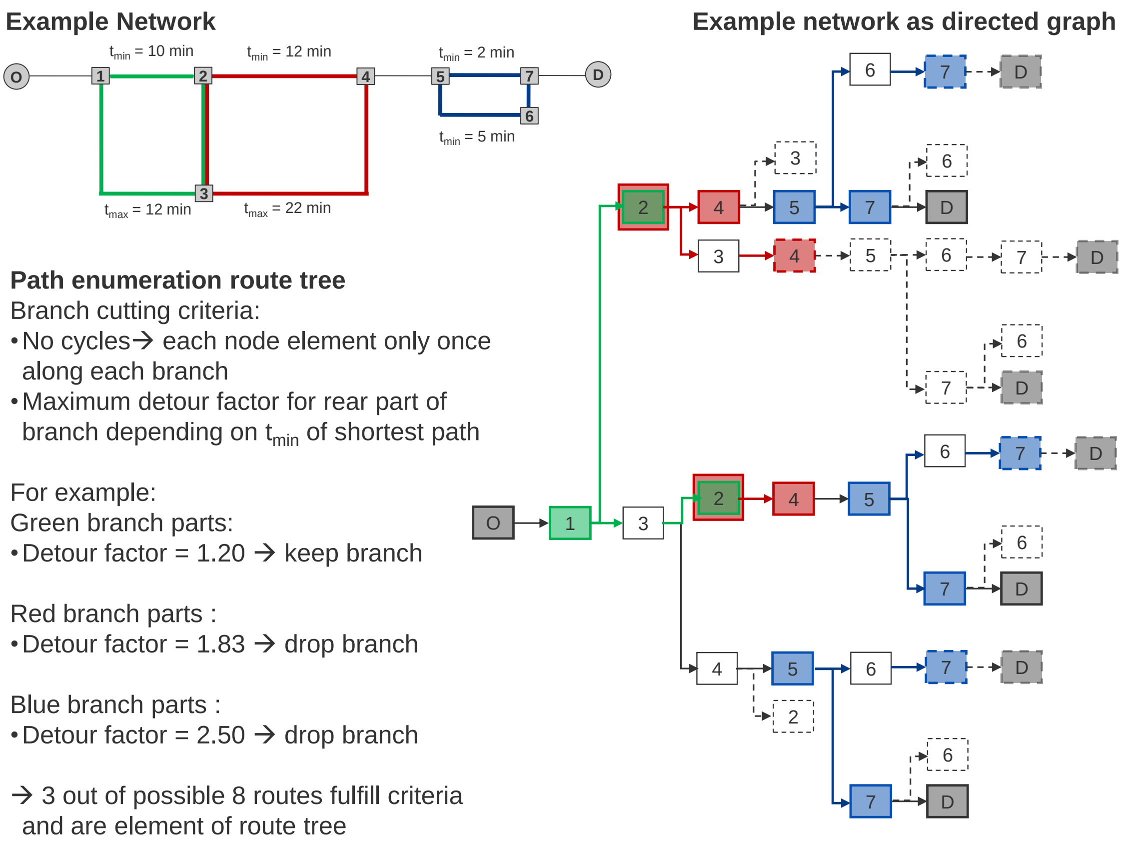

Figure 4: Route tree for seven node example network

4 Choice Set Composition

For each surveyed OD pair revealed, known and generated routes were fused to a spatial choice set. To ensure each included route is unique and to reduce the choice set to a computable size a commonality factor C is applied (CASCETTA [4]). Before a route is added to the choice set the commonality factor to already included routes is calculated. Only routes with a maximum commonality factor of 0.9 (see SCHÜSSLER [5]) are put in the choice set additionally. First, all revealed routes resulting from observed trips are added. Secondly, stated routes and last generated routes are added. That means, e.g. a generated route which is similar to a chosen route is not included. Figure 5 illustrates the spatial choice set for one OD pair.

The estimation of the route choice model is based on evaluating each choice situation of the participants. For this OD pair a spatial choice set with static attributes (length, detour factor etc.) exists. Knowing the starting time of the trip, time-dependent attributes of the routes such as current travel time, historical travel time, relevant TMC-messages are set with data from the traffic archive. Finally, for every observed trip a choice set including time-dependent and time-independent attributes is composed.

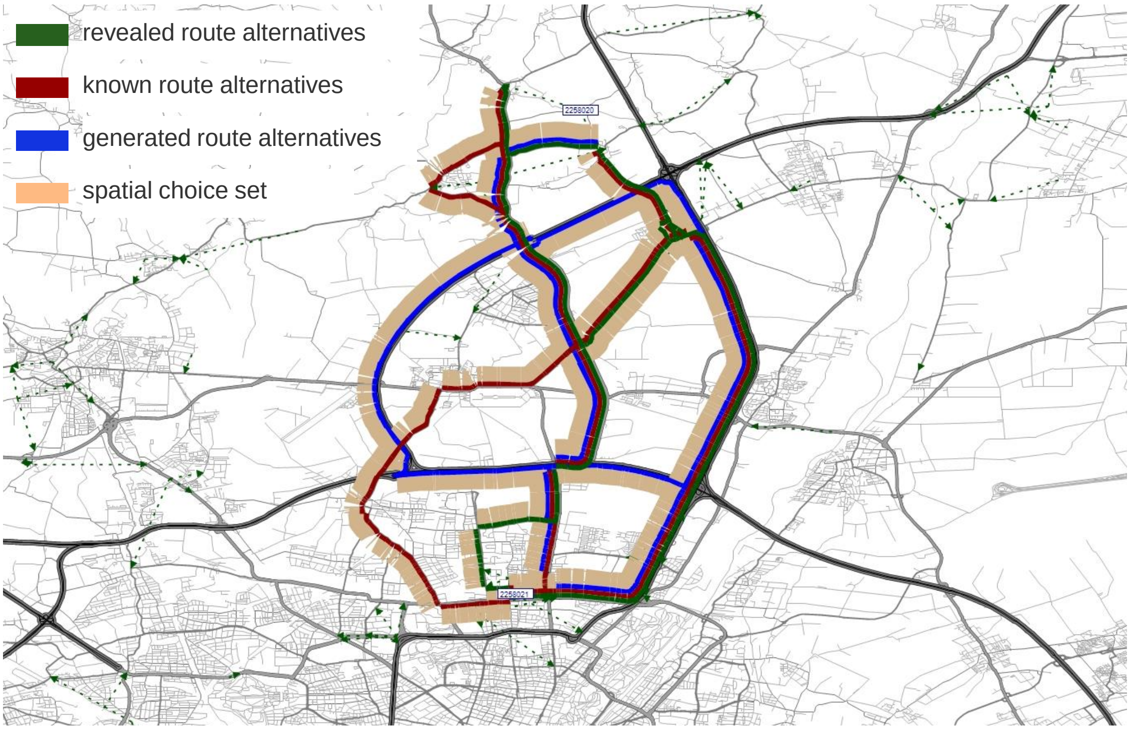

Figure 5: Spatial choice set for one OD pair

5 Conclusions and Future Work

Table 1 analyzed three routes types: chosen (if the route was chosen at least once for a trip), known (if the route was stated in the interview but never chosen) and generated (if the route was never chosen nor stated in the interview). Furthermore, the size of the developed choice sets is displayed separately for all OD pairs as well only the home to work OD pair.

Table 2 shows the importance of different route attributes for driver's route choice by analyzing how often the currently fastest, historically fastest or shortest route was chosen over a total of 4,216 trips. The percentages indicate that the current travel time plays a minor role in driver's route choice compared to the historical travel time or the route length and that a major part of choices are towards routes with other, more complex benefits. This emphasizes the need for analyzing the impact of traffic information as well as other route attributes.

Route type	**Total number for all participants on all OD pairs**	**Total number for all participants for OD pair home to work**	**Mean number of routes per OD pair**	**Maximum number of routes per OD pair**
chosen	1,755	426	1.74	12
known	119	119	1.67	4
generated	17,028	4,908	16.89	253
sum	18,902	5,453	20.30	269

Table 2: Number of routes in choice sets and coverage of chosen and known routes by generated choice sets (known routes are only given for trips from home to work)

Route type	Morning peak 6 – 9 am Chosen routes	Off-peak 6 pm – 9 am, 9 am – 3 pm Chosen routes	Evening peak 3 – 6 pm Chosen routes	Total 24 hours Chosen routes
currently fastest	4%	4%	5%	4%
historically fastest	28%	30%	29%	30%
shortest	18%	20%	16%	19%
other	50%	46%	50%	47%
sum	100%	100%	100%	100%
absolute numbers	497 trips	3,158 trips	561 trips	4,216 trips

TABLE 3: Number of routes types chosen over survey period in % (values based on ground population of 4,216 trips on 1,008 OD pairs over 8 weeks of survey)

These preliminary results are based on trips from 100 of the total 300 participants of the survey. The underlying traffic archive data is still subject to plausibility checks, which will be completed briefly. Future work involves choice model estimation as well as model identification, especially focusing on the effect of traffic information in driver's route choice. The proposed route choice model aims at identifying device acceptance, driver compliance as well as potentials of adaptive traffic management for reduction of network congestion.

Acknowledgments

This research is part of the project *wiki*, which is funded by the German Federal Ministry of Economy and Technology (Bundesministerium für Wirtschaft und Technologie). The GPS map-matching and dynamic traffic data was provided by PTV AG in Karlsruhe, Germany, based partly on data from the Highway Administration of South Bavaria, Germany.

References

[1] Schüssler, N. and K.W. Axhausen: Processing Raw Data from Global Positioning Systems Without Additional Information, *Transportation Research Record*, *28-36,* 2009.
[2] Schlaich, J.: Nutzung von Mobilfunkdaten für die Analyse der Routenwahl, *PhD Dissertation at the Department of Traffic Engineering and Transportation Planning, University of Stuttgart Stuttgart*, Germany, 2009.
[3] Schlaich, J., Otterstätter, T. and M. Friedrich: Generating Trajectories from Mobile Phone Data, *TRB 89th Annual Meeting Compendium of Papers DVD-ROM*, Transportation Research Board of the National Academies, Washington, D.C., USA, 2010.
[4] Cascetta, E.: Transportation Systems Engineering: Theory and Methods, *Kluver Academic Publishers,* Dordrecht, Netherlands, 2001.
[5] Schüssler, N., Balmer, M. and K.W. Axhausen: Route Choice Sets for Very High-Resolution Data, paper presented at the 89th Annual Meeting of the Transportation Research Board, Washington, D.C., 2010.