ETH zürich

ETH Library

Identifying chosen public transport connections from GPS observations

Conference Paper

Author(s):
Rieser-Schüssler, Nadine; Axhausen, Kay W.

Publication date:
2013-01

Permanent link:
https://doi.org/10.3929/ethz-b-000073415

Rights / license:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection.
For more information, please consult the Terms of use.

Identifying chosen public transport connections from GPS observations

Date of submission: 2012-07-20

Nadine Rieser-Schüssler
IVT, ETH Zürich, CH-8093 Zürich
phone: +41-44-633 30 85
fax: +41-44-633 10 57
nadine.rieser@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

Words: 6771 words + 3 figures = 7521 word equivalents

ABSTRACT

Transport planners around the world are currently searching for innovative strategies for customer-friendly and efficient public transport systems. An important element in this process is the understanding of the passengers' valuation of different elements of public transport trips. A challenge associated with this is the observation of the actual passenger behaviour in all its complexity. One way to address this challenge is to use person-based GPS devices for the observation of the public transport connections chosen by passengers. GPS-based studies have become increasingly popular in the last two decades and their advantages for observing and modelling car and bicycle route choice have been shown by many studies. However, for public transport connection choice the processing routines to extract the chosen connections and their relevant attributes from the GPS traces have so far been missing.

This paper reports on a first implementation of such a procedure called *"public transport map-matching"*. The basic idea is to first employ a car map-matching procedure for each stage of a public transport trip to determine the route within the public transport network. Then, this route is used to find the most likely public transport line and the respective boarding and alighting stops. Finally, the stages of the public transport trip are joined together including the access and egress stages by walk or bicycle. The procedure is tested using the data from an ongoing GPS study in Zurich – an area in Switzerland with a very dense public transport network.

INTRODUCTION AND RELATED WORK

A major issue in modern transport planning is how to accommodate future travel demand while resources are getting scarcer at the same time. One solution is to increase the attractiveness of public transport. Therefore, innovative strategies for customer-friendly and efficient public transport systems are required. An important element in this process is to better understand the passengers' behaviour and their valuation of different elements of public transport trips. This task is particularly challenging since not only the behaviour of passengers is complex but also the choice situations they are confronted with. This is particularly true in dense urban networks where passengers can choose between a large variety of public transport options.

One of the difficulties associated with the complexity of the choice task is observing the actual passenger behaviour and more specifically, the chosen public transport connections with all their attributes, e.g. lines, departure times, access and egress modes and times, transfer points, etc. Traditionally, questionnaire based travel diaries are used *(*e.g. *1, 2, 3)*. However, they require the respondents to remember a lot of details about their trips which many respondents are not able to recall. Another – increasingly popular – approach is to use smart-card interactions *(*e.g. *4, 5, 6, 7)* to deduce the public transport connections chosen by passengers. But often the destination stop of the trip has to be imputed because the system does not require alighting smart-card interactions and access and egress stages are inherently missing in these data sets.

An alternative to these approaches is observing the connection choice using person-based GPS devices. Since the beginning of this century, an increasing number of person-based GPS studies have been conducted world-wide *(*e.g. *8, 9, 10, 11, 12)*. The GPS device is carried by the participants for several days and records in high frequency the participants' location. Post-processing procedures *(*e.g. *13, 14, 15, 16)* are then used to reconstruct travel diaries including activities and trips, modes, departure and arrival times etc. Moreover, several studies have shown the advantages of GPS records to observe car *(*e.g. *17, 18, 19, 20, 21)* or bicycle *(*e.g. *22, 23)* route choice because the chosen routes are observed at a level of accuracy and detail that is hard to reach with recollection-based surveys.

The identification of the routes chosen by car drivers or bicyclists is usually done in a processing step called map-matching. Over the years, a number of different procedures *(*e.g. *24, 25, 26, 14, 27)* have been developed. However, processing routines that extract public transport specific information such as the chosen lines, boarding, alighting and transfer stops have – at least to our knowledge – so far been missing. What is needed is a procedure that works analogously to car map-matching and identifies the chosen public transport connection including access to and egress from the public transport system.

This paper reports on a first implementation of such a procedure called *"public transport map-matching"*. The basic idea is to first employ the same map-matching procedure *(14)* as for car or bicycle trips for each stage of a public transport trip to determine the route the participant travelled along within the public transport network. Then, the result of this step is used to find the public transport line the person most likely chose and the respective boarding and alighting stops. Finally, the stages of the public transport trip are joined together including the access and egress stages by walk or bicycle.

The GPS traces used to test the procedure originate from a person-based GPS diary survey that is still ongoing with participants living in and around Zurich – an area in Switzerland with a very dense public transport network. The participants are asked to carry a person-based GPS receiver for a week and to confirm and correct the results generated by the automatic post-processing of their GPS tracks in a web-based prompted recall survey. More details about

this data and the public transport network information used can be found in the next section. In the subsequent section, the methodology of the public transport map-matching is elaborated. Then, the performance of the procedure in terms of how well it was able to extract the chosen connections is presented before the paper closes with some final remarks and an outlook on future work.

SURVEY DESIGN AND PUBLIC TRANSPORT NETWORK DATA

Two sources of data are necessary to identify the public transport connection actually chosen by the study participants: The GPS traces of their public transport trips including associated access and egress stages and information about the public transport network, including the lines, their routes and preferably also their schedules. This section shortly discusses the origins of the GPS traces and the available public transport network data.

GPS observations

The survey conducted to obtain the GPS traces consists of three components. The first component is a basic questionnaire for socio-economic characteristics such as age, gender, income, mobility tool (e.g. car or public transport season ticket) ownership and employment status. The second component comprises three psychometric scales measuring the attitudes towards risk, environment and variety in the daily schedule. The third component is a one-week diary collected by tracing the participants via GPS, automatically generating a travel diary from these traces and asking the participants to confirm and correct their diaries using a prompted recall interface on the website. In the following, the GPS component is briefly discussed. For a more detailed description of GPS survey the reader is referred to Rieser-Schüssler *et al. (28)*.

For the data collection of the GPS-based travel diary component, the respondents are handed out a MobiTest GSL by MGE Data *(29)* along with a charger. They are asked to carry the device with them for a week whenever they leave the house and to charge it during the night. While the device is charging the GPS data is transmitted via GSM to our server and automatically processed. The results are fed into a database and displayed in the interactive prompted-recall diary.

The data collected by the GPS device comprises GPS traces and accelerometer measurements. The GPS points are collected at 1 Hz frequency and consist of the three-dimensional position, the timestamp, the horizontal and vertical accuracy of the position and the number of satellites in view. The accelerometer provides three-dimensional acceleration information in 10 Hz intervals.

The goal of the prompted recall survey is to check and correct the processing results so that the resulting diaries are trustworthy and ready for further analysis. The participants can add, delete, confirm or correct their trip stages and activities that are depicted on digital maps and described in textual form. Initially, the participants were supposed to do this on their own the day after the actual travel day and return the GPS device by priority mail service. However, it turned out that a lot of respondents had difficulties to do this properly even after being carefully trained at the beginning. Therefore, the survey protocol was changed and, instead of receiving a training at the beginning, the interviewers visit the participants after the data collection is completed, collect the GPS device and assist the participants in the prompted-recall survey. In addition, the diaries of the participants that had already completed the survey before these changes are manually checked and corrected by the IVT team.

The focus of the prompted recall survey was put on the correct identification of stages,

activities, modes and trip purposes. The participants were not queried about details of their stages such as the public transport lines used or the name of their transfer stops. This was left out because it would have increased the already rather high participant burden even further. Instead, the details of the public transport trips where added manually by the research team taking into account several sources of information such as the network, departure times and schedule information as well as the recommendations of online trip planning tools.

At the time of writing, the main survey is still ongoing. The aim is to survey 300 participants until the end of the year. Participants have to be older than 18 and live within a radius of 25 km around Zurich. The radius was chosen to just include the cities of Winterthur and Zug, which are the regional centres closest to Zurich. The data set used in this paper contains the travel diaries of the survey participants that have already been cleaned and double-checked for consistency. It contains 2507 stages conducted by 58 participants on 435 observation days. The majority of these stages (1073/42.8%) were done by car, but 129 (5.15%) bus, 111 (4.43%) tram and 175 (6.98%) train stages are available for the public transport map-matching. In addition, the 668 (26.65%) walk and 332 (13.34%) bike stages are used to construct the whole public transport trips including the walk access and egress stage.

Moreover, data from the pretest conducted with members and friends of the institute is used. The data set contains 339 stages by 7 individuals on 65 observation days. 79 of these stages were done by bus, 23 by tram and 33 by train. In addition, 117 walk and 15 bike stages are available for the trip construction.

Available public transport infrastructure data

The public transport network data originates from the official transport model for the Canton Zurich *(30)* that was built based on the HAFAS-timetable for Switzerland. The network covers all major public transport lines, including ships and cable cars. It includes all major train connections to and from Zurich and all local public transport lines within Canton Zurich. The links in this network represents the actual geographic location of the link which is essential for this procedure to work. The HAFAS-timetable represents the status in the year 2005. We have updated the network, since a few lines have been extended and two new tram lines have been added since then.

The original VISUM *(31)* network was converted into a MATSim *(32)* format. In this process, the network was automatically cleaned with regards to obvious errors such as zero length links. Links that were not used by any public transport line were removed. As a result, three components jointly describe the public transport infrastructure: the network on which the public transport lines operate, the public transport stops and the public transport lines.

The network contains the roads on which buses and trams run, the rail tracks of the train lines. Each network link is described by several criteria, such as its length, free-flow speed and which modes are allowed to use it. The geographical reference is provided by the nodes and their coordinates. The public transport stops are also geographically referenced by their coordinates. However, for each stop only one coordinate is given even if the stop is actually distributed over several stop locations, e.g. for different directions or lines. The description of the lines requires the most extensive data structure. Each line is described by its name, its route(s) in terms of stops and in terms of network links. Each route has at least one time profile and the departure times are assigned to the time profiles forming the courses of the line. By this, variations in routes and scheduled departure times during the course of the day are accounted for.

In this application, two variations of this network and schedule data were used: One containing all public transport modes and one containing only the rail tracks and the train schedule.

THE PUBLIC TRANSPORT MAP-MATCHING PROCEDURE

The public transport connection choice identification is done in a module of the open source GPS data analysis framework POSDAP *(33)* called *public transport map-matching*. Before applying the public transport map-matching the GPS data needs to be cleaned, stored in a suitable database along with the accelerometer measurements and subdivided into stages and stop points where the participants performed an activity or a transfer. This can be done using the respective procedures in POSDAP, that are described in Schüssler and Axhausen *(34)* and Rieser-Schüssler *et al. (28)*, any other procedure or prompted recall information given by the participants.

In terms of network data, the public transport map-matching essentially requires a public transport network containing all links and nodes used by public transport lines, the location of all public transport stops and a description of the public transport line routes both as a sequence of stops and as a sequence of links such as the data described in the last section. In addition, schedule information such as departure times and travel times can be used but are not mandatory.

The public transport map-matching is conducted in fives steps:

- Car map-matching to the public transport network
- Identification of line candidates
- Selecting the chosen line
- Determination of boarding and alighting stops
- Joining stages to trips

In the car map-matching, the route the participant travelled along within the public transport network is determined. The second step, this route is compared to the routes of the public transport lines to identify potentially used public transport lines. If schedule information is available, the chosen line identification can take into account departure and travel times, otherwise the chosen line is selected based on the distance between the closest stops and random draws. The fourth step determines the most probable boarding and alighting stops by finding the stops along the line route that are closest to the start and end point of the stage. In the last step, the stages are joined to trips taking into account distance and time to the surrounding stages. Each of the five steps is explained in more detail below.

Car map-matching to the public transport network

The identification of the routes the participants travelled along within the public transport network is done using the map-matching that was developed for car and bike trips in Schüssler *(14)*. This approach employs Multiple Hypothesis Technique by developing as set of possible paths through the network and only choosing the most likely path when the entire GPS trace has been processed. To keep this procedure computationally efficient not all possible paths can be considered. Instead only the best candidates are kept in a consideration set. The size of the consideration set has to be determined by the analyst weighing the accuracy of the algorithm against the performance. For the rather sparse public transport network limiting the size of the consideration set to 20 alternatives was sufficient not to encounter any accuracy problems.

The final selection of the most likely path is based on the score of each candidate path which is calculated as the sum of the distances between the path and all GPS points. The result of the map-matching is a sequence of links as shown in Figure 1 that can then be compared to the routes of the public transport lines.

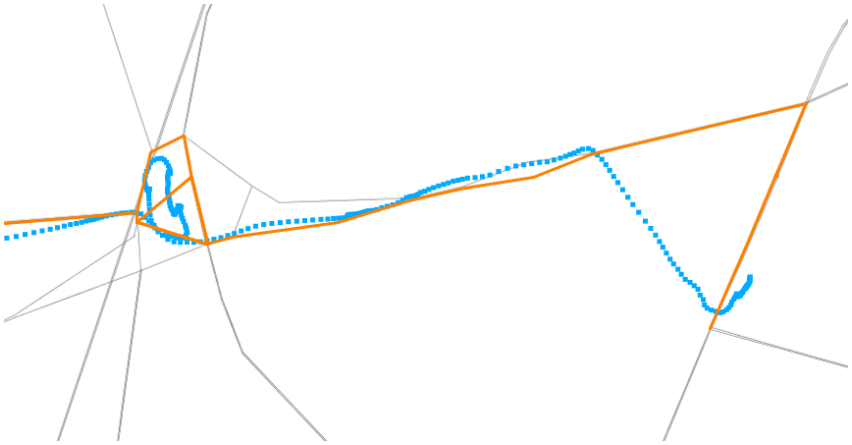


FIGURE 1 Links identified by map-matching GPS traces to the public transport network

Since the public transport network contains considerably less links than the car or bike network, the map-matching is faster and even more reliable in finding the right route in this application. A major challenge, however, is dealing with inaccurate network codings such as the one seen on the right side of Figure 1. The blue GPS points follow the route the bus actually takes. But there is no link in the network because the line is coded along the orange marked links. In the network that is used in this study, many errors like this have been found. We have corrected the most significant ones. But since the resources to find all of them are prohibitive, the subsequent identification of line candidates accounts for the possibility of this kind of errors.

A change compared to the car map-matching described in Schüssler *(14)* is the treatment of long recording gaps (i.e. no GPS point for more than 120 s or 1000 m) within a stage. Similar to the car version, for each long recording gap, the stage is subdivided into two segments containing the GPS points before and after the gap, respectively. But different from the car map-matching the route is not closed using a shortest path search. Instead, the line identification is performed separately for each segment and the subsequent line selection takes care of joining the segments back together as described in the corresponding section.

Since GPS reception in trains is usually bad resulting in traces consisting mainly of recording gaps, the gap segmentation was not applied to train stages. However, this was only feasible because the map-matching of the train stages employed the very sparse rail network. Only in a very sparse network, the map-matching is – at least in most cases – able to follow the route during a recording gap as long as enough GPS points before and after the gap are available.

Identification of line candidates

Due to the network coding issues discussed above, a two-step process was implemented for the identification of line candidates for each segment. In the first step, the algorithm does an exact link-by-link comparison between the route determined by the car map-matching and the link routes of all lines. The result can either be one unique line or several ambiguous lines .In case more than one line is found, all of them are stored in a set of line candidates for later processing.

If no valid line route can be identified this way, the second step is executed. This time, the matching between the map-matched links and the links of the line route does not have to be exact. Instead, certain deviations are tolerated. Since these deviations can occur at the beginning,

in the middle or at the end of the stage the length of all deviating links is summed up. If this sum does not exceed a certain tolerance level, the line is added to the set of line candidates. The tolerance level is currently defined as 15% of the length of the map-matched route.

There are two reasons for using a two-step process and not using the deviation tolerance approach for all stages. First, an exact match should have priority over an approximate solution. Second, the exact matching is faster to compute because it can discard unsuitable routes quicker by not keeping them in memory until the tolerance level for the deviation is exceeded. Especially in dense public transport networks this can make a noticeable difference in computation time.

Selecting the chosen line

The first step in selecting the line that was most likely used, is to check whether the car map-matching subdivided the stage into segments due to long recording gaps. If that is the case, these segments are joined before any further selection procedure is applied. For each recording gap, the joining procedure compares all line candidates of the segments before and after the gap. If no common line candidate was detected, the stage is filtered because it needs further checking and might not be a public transport stage. If there is at least one common line, the segments are joined together and the common lines make up the set of line candidates for the entire stage.

Subsequently, the set of line candidates for the entire stage is evaluated. If it contains only one line, this line is set as the chosen line for the stage and the line selection procedure ends. If the set of line candidates contains more than one line, the actual line selection executed. For this, currently three different approaches are available that can also be executed in sequence.

The first approach is based on the distance between the first and last GPS points of the stage and the respective closest public transport stops. First, the public transport stop closest to the first GPS point of the stage out of all the stops served by the lines in question is determined. Only the lines that serve this stop are retained in the set of line candidates. Then, the set of line candidates is further reduced by repeating this procedure for the last GPS point of the stage. This procedure is particularly useful in cases such as the one depicted in Figure 2 where the grey and the pink line share the same link route but serve different stops. The grey route represents an express train that serves only the grey stop in the top left corner whereas the pink line, representing a local train, also serves the other two stops. The blue GPS points clearly show that the participant boarded the train at the pink stop in the lower right corner, therefore, he can only have used the pink line.

The basic idea of the second approach is to compare the actual departure time of the stage with the scheduled departure times of the line candidates. For each line the stop closest the departure times of the lines at the stop closest to the first GPS point are compared to the actual departure time in the GPS observations. Then, the line(s) with the minimum time difference is/are chosen. This does, however, presume that the schedule is available and that the public transport lines are running according to schedule. Thus, if the public transport system is unreliable, the approach would require real time vehicle location data.

The last approach is to randomly draw line out of the remaining set of line candidates. This approach was primarily implemented for application cases where the first approach did not suffice to reduce the set of line candidates down to one line and no schedule or real time vehicle location data is available to use the second approach. If the network coding and the GPS traces are accurate and complete, this will can still produce a reasonable selection.

In the tests described in this paper, the line selection utilises all three approaches in the

FIGURE 2 Distance-based line selection

sequence described here. First, the set of line candidates is reduced using the distance-based approach. Then, if the set of line candidates still contains more than one candidate, the schedule-based selection is employed. Only if this is still not sufficient to single out only one line the random draw is used.

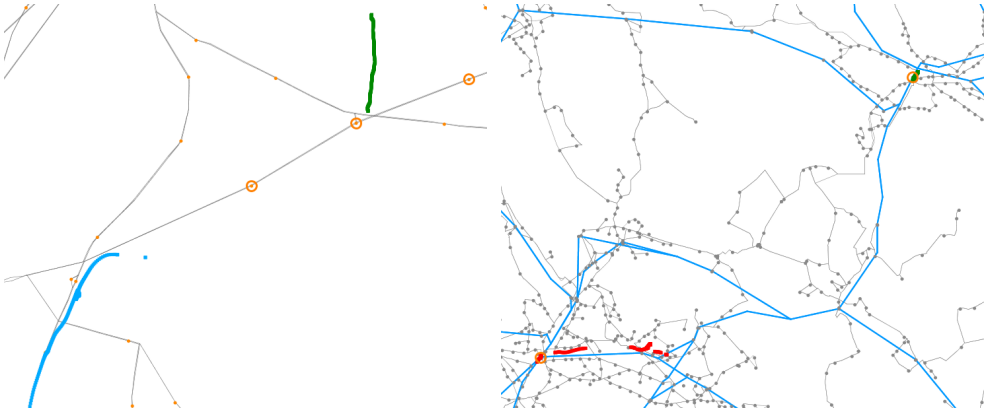
Determination of boarding and alighting stops

The fourth step in the identification of the public transport connection is the determination of the boarding and alighting stops. For the majority of the stages, this is done by finding the stops that are closest to the first and last GPS point of the stage and are served by the selected line. However, in two cases special procedures are necessary. The first case encompasses tram stages that start or end at stops that are located underground since for these stops no GPS signal is available. The second case comprises the train stages because GPS reception is habitually poor in trains.

For the tram stages that potentially start or end at underground stops, a preliminary solution was used that requires the analyst to specify which stops served by which lines are underground stops. In the Zurich case, this only concerns two tram lines and three tram stops that are marked with orange circles in Figure 3(a) which shows an example for a tram stage that ends at an underground stop. The blue points are the GPS points that are available for the tram stage and the green points the GPS points available from the subsequent walk stage. The basic idea derived from this is to use the subsequent (or preceding) walk stage to determine the correct alighting (or boarding) stop.

First, the procedure has to determine whether the stage potentially contains an underground alighting or boarding. This is done by checking if the chosen line is one of the concerned lines, the last valid stop is a stop before the tunnel, there is a suitable gap – larger than 500m but no longer than 15 minutes – between the public transport stage and the subsequent (or preceding) stage is a walk or bike stage. Then, the underground stop that is closest to the start (or end) of the walk or bike stage is determined as the alighting (or boarding) stop.

The procedure for identifying the boarding and alighting stops for train stages is similar but not entirely the same. First, not only walk and bike stages but also public transport stages



(a) Tunnel stops (b) Train stages

FIGURE 3 Boarding and alighting stop determination for tunnel stops and train stages

are allowed to be subsequent or preceding stages. Second, there are no fixed distance and time thresholds since recording gaps in train stages are often very large as can be seen in Figure 3(b). Instead, the gap speed, calculated based on the distance and time difference between the last (or first) GPS point of the train stage and the first (or last) GPS point of the subsequent (or preceding) stage are evaluated. If this gap speed is at least 10 m/s, the end (or start) stop of the train stage is the stop served by the chosen train line, that is closest to the start (or end) of the subsequent (or preceding) stage.

Joining stages to trips

Finally, after the successful identification of the public transport lines and the boarding and alighting stops for each stage, the stages are joined to trips that should contain the entire door-to-door journey including the access and egress walk or bike stages, all public transport stages and the transfers and waiting times in between.

A key element in the joining process is to decide whether two succeeding stages should be joined to a trip. This is done based on a set of assumptions regarding the maximum allowed time difference and distance between walking or biking and public transport stages or two subsequent public transport stages. In this application, comparably high thresholds of 500 m and 900 s produced good results as can be seen in the next section. However, depending on the study area different thresholds might be applicable and can be configured. The analyst even has the option to define different time thresholds depending on whether the interaction in question is boarding, alighting or transfer because one can assume that people rarely wait around long after alighting from their last public transport stage, but often wait before boarding and during transfers.

Similar to the determination of the boarding and alighting stops, a special treatment is necessary for stages that end in a tunnel and train stages. For tunnel stages, a higher distance threshold of 2500 m is used to account for the issue that the boarding stop within the tunnel is not always detected properly because the GPS signal needs too much time to pick-up again after the tunnel. For train stages the time difference is set to 60 minutes because the recording gap is often very large.

EVALUATION OF THE MAP-MATCHING RESULTS

This section gives a brief overview about the performance of the map-matching in the sense of how well it was able to reconstruct the chosen lines, the stops used for boarding, alighting and transfer and the overall trips made by the participants. The major part of the analysis was done by manually comparing the results of the map-matching with the available trip records and GPS points and classify the cases in which one or more parts of the procedure failed. To account for the different characteristics of the two samples and the particular issues associated with the low GPS reception in trains, the analysis differentiates between the two available samples and between train and bus/tram stages.

Of the 102 bus and tram stages in the pretest sample, the public transport map-matching was successful and correct for 92 stages. For nine stages, no valid line was found and for one stage the detected line was wrong, but this was caused by a line detour due to on-going construction. The success rate of 90% is satisfactory. For the tram and bus stages in the main study data, the results are slightly worse with 183 of the 240 stages being successfully matched and for 57 stages the map-matching failed. The reasons for public transport map-matching failures are similar for both samples. The main reasons were:

- No, sparse or bad GPS reception (3/22 cases)
- Trips outside the study area (3/16 cases)
- Errors in the line coding (2/8 cases)
- Detours due to construction (2/3 cases)

As can be seen from this list the map-matching failures were due to external constraints such as bad GPS reception, missing network coverage or errors in the network representation. In the main study sample, in particular the first two issues occurred more frequently than in the main study leading to the lower success rate for this sample.

As expected, the issue of no, sparse or bad GPS reception was even worse for the rail stages. Accordingly the success rates of the public transport map-matching for rail trips were lower. The public transport map-matching was unable to identify a correct public transport connection for 18 of the 33 rail stages in the pretest sample and 87 of the 173 rail stages of the main study sample. In 14 and 74, respectively, cases this was caused by no, sparse or bad GPS reception. Other reasons were trips outside the study area (1/4 cases), misclassified stages (1/6 cases) and wrong line codings (1/2 cases).

Subsequently, the public transport stages for which a valid line and valid start and end stops were identified were joined to 80 trips that include 58 walk stages for the pretest sample. The relative low number of walk stages in the pretest sample trips is mainly caused by missing identification of access and egress walk stages through the participants and not by a problem in the trip joining procedure or systematic recording gaps. A manual analysis of the 80 pretest sample trips showed that 67 of them were joined correctly but 12 were missing at least one stage. Most of these stages were train stages and all of them were stages for which the public transport map-matching had failed.

For the main study sample, 203 trips including 218 walk and 43 bike stages for the main study sample were detected. 142 of these trips were found to be correct. 13 trips are actually part of another trip, but the trip was split because the public transport map-matching failed for an intermediate stage. For the remaining stages, at least one stage at the beginning or the end was missing due to failed map-matching, mainly of train stages.

In terms of computation time, the map-matching was run on a single CPU of a desktop

Windows 7 PC with a 3 GHz dual-core AMD Opteron 2222 processor and 1 GB RAM. On this set-up the public transport map-matching was completed in 637 seconds for the 135 stages of the pretest sample and in 1232 seconds for the 413 stages of the main study.

CONCLUSION AND OUTLOOK

The increasing popularity of person-based GPS devices for travel behaviour studies opens up new possibilities to observe different aspects of travel behaviour. One of these aspects is public transport connection choice with all its attributes such as chosen lines, access and egress stages, transfer points or waiting times. However, so far a procedure has been missing that can automatically extract the chosen public transport connection from the GPS traces.

This paper presents such as procedure called public transport map-matching and evaluates its performance using data from an ongoing GPS study and its pretest in Zurich – an area in Switzerland with a very dense public transport network. In terms of accuracy, the public transport map-matching performs very satisfactory for tram and bus stages with a success rate of about 90% and the main reasons for matching failures being not enough GPS points or trips outside the area for which public transport network data was available. These were also the main reasons for matching failures for the train stages, but since GPS reception is much worse in trains than in buses or trams, only half of the train stages could be matched satisfactorily. Therefore, future work will explore different approaches to identify at least the start and end points of rail stages based on the access and egress stages.

The procedure to join the stages into trips also works reliably as long as all public transport stages within the trip have been matched successfully. Along with new approaches for rail stages, options for joining trips with missing intermediate stages will be considered.

Apart from these shortcomings, the trips that have been joined successfully are now ready for further analysis. The next step towards model estimation will be the generation of choice sets. A variety of efficient time-table based choice set generation procedures will be tested to investigate the influence of the choice set composition on choice estimates. The basic idea is to generate a routing network that contains the entire schedule and generate routes using least-cost path based choice set algorithms on this network. In Rieser-Schüssler *et al. (35)*, a proof-of-concept study for this approach has shown promising results. The next step will be testing different procedures.

ACKNOWLEDGEMENTS

This research was funded by the Swiss State Secretariat for Education and Research as part of the project "Route choice in urban public transport systems" within the COST Action "TU0603 - Buses with high levels of service" and by the European Union as part of the project "Peacox - persuasive advisor for CO2-reducing cross-modal trip planning" within the Seventh Framework Programme (FP7).

REFERENCES

1. Hoogendoorn-Lanser, S. (2005) Modelling travel behaviour in multi-modal networks, Ph.D. Thesis, Technical University Delft, Delft.

2. Larsen, M. K. and T. K. Rasmussen (2010) Matching observed public route choice data to a GIS network, paper presented at the *Trafikdage på Aalborg Universitet*, Aalborg, August 2010.

3. Eluru, N., V. Chakour and A. El-Geneidy (2012) Travel mode choice and transit route choice behavior in Montreal: Insights from McGill University members commute patterns, paper presented at the *91st Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2012.

4. Wilson, N. H. M., J. Zhao and A. Rahbee (2009) The potential impact of automated data collection systems on urban public transport planning, in N. H. M. Wilson and A. Nuzzolo (eds.) *Schedule-Based Modeling of Transportation Networks: Theory and Applications*, 75–99, Springer, New York.

5. Munizaga, M. A., C. Palma and D. Fischer (2011) Estimation of disaggregate multimodal public transport OD matrix from passive SmartCard data from Santiago, Chile, paper presented at the *90th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2011.

6. Devillaine, F., M. A. Munizaga and M. Trépanier (2012) Detection of public transport user activities through the analysis of smartcard data, paper presented at the *91st Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2012.

7. Chakirov, A. and A. Erath (2011) Use of public transport smart card fare payment data for travel behaviour analysis in singapore, paper presented at the *16th international conference of Hong Kong Society for Transportation Studies*, Hong Kong, December 2011.

8. Draijer, G., N. Kalfs and J. Perdok (2000) Global Positioning System as data collection method for travel research, *Transportation Research Record*, **1719**, 147–153.

9. Wolf, J. and M. Oliveira (2008) Metropolitan Washington, D.C., household travel survey Global Positioning System pretest: Results and applications for large-scale regional survey, paper presented at the *87th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2008.

10. Bricka, S., J. Zmud, J. Wolf and J. Freedman (2009) Household travel surveys with GPS: An experiment, paper presented at the *88th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2009.

11. Marchal, P., J.-L. Madre and S. Yuan (2011) Post-processing procedures for person-based GPS data collected in the French National Travel Survey 2007-2008, paper presented at the *90th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2011.

12. Oliveira, M., P. Vovsha, J. Wolf, Y. Birotker, D. Givon and J. Paasche (2011) GPS-assisted prompted recall household travel survey to support development of advanced travel model in Jerusalem, Israel, *IEEE Transactions on Power Systems*, **2246**, 16–23.

13. Stopher, P. R., Q. Jiang and C. FitzGerald (2005) Processing GPS data from travel surveys, paper presented at the *2nd International Colloqium on the Behavioural Foundations of Integrated Land-use and Transportation Models: Frameworks, Models and Applications*, Toronto, June 2005.

14. Schüssler, N. (2010) Accounting for similarities between alternatives in discrete choice models based on high-resolution observations of transport behaviour, Ph.D. Thesis, ETH Zurich, Zurich.

15. Chen, C., H. Gong, C. Lawson, E. Bialostozky and J. Muckell (2010) Evaluating the feasibility of a passive travel survey data collection in a complex urban environment: A case study in New York City, paper presented at the *89th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2010.

16. Moiseeva, A., J. Jessurun and H. J. P. Timmermans (2010) Semi-automatic imputation of activity travel diaries using GPS-traces, prompted recall and context-sensitive learning algorithms, *Transportation Research Record*, **2183**, 60–68.

17. Jan, O., A. J. Horowitz and Z.-R. Peng (2000) Using Global Positioning System data to understand variations in path choice, *Transportation Research Record*, **1725**, 37–44.

18. Nielsen, O. A. (2004) Behavioral responses to road pricing schemes: Description of the Danish AKTA experiment, *Journal of Intelligent Transportation Systems*, **8** (4) 233–251.

19. Li, H., R. Guensler and J. Ogle (2005) Analysis of morning commute route choice patterns using Global Positioning System-based vehicle activity data, *Transportation Research Record*, **1926**, 162–170.

20. Frejinger, E. (2008) Route choice analysis: Data, models, algorithms and applications, Ph.D. Thesis, EPF Lausanne, Lausanne, April 2008.

21. Schüssler, N. and K. W. Axhausen (2009) Accounting for route overlap in urban and suburban route choice decisions derived from GPS observations, paper presented at the *12th International Conference on Travel Behaviour Research (IATBR)*, Jaipur, December 2009.

22. Menghini, G., N. Carrasco, N. Schüssler and K. W. Axhausen (2010) Route choice of cyclists in zurich, *Transportation Research Part A: Policy and Practice*, **44** (9) 754–765.

23. Broach, J., J. P. Gliebe and J. Dill (2011) Bicycle route choice model developed using revealed preference GPS data, paper presented at the *90th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2011.

24. Pyo, J.-S., D.-H. Shin and T.-K. Sung (2001) Development of a map matching method using the multiple hypothesis technique, paper presented at the *Intelligent Transportation Systems Conference (ITSC)*, Oakland, August 2001.

25. Marchal, F., J. K. Hackney and K. W. Axhausen (2005) Efficient map matching of large Global Positioning System data sets: Tests on speed-monitoring experiment in Zürich, *Transportation Research Record*, **1935**, 93–100.

26. Quddus, M. A., W. Y. Ochieng, L. Zhao and R. B. Noland (2006) A high accuracy fuzzy logic-based map matching algorithm for road transport, *Journal of Intelligent Transportation Systems*, **10** (3) 103–115.

27. Dalumpines, R. and D. M. Scott (2011) GIS-based map matching: Development and demonstration of postprocessing map-matching algorithm for transportation research, paper presented at the *90th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2011.

28. Rieser-Schüssler, N., L. Montini and C. Dobler (2011) Improving automatic post-processing routines for GPS observations using prompted-recall data, paper presented at the *9th International Conference on Survey Methods in Transport*, Termas de Puyehue, November 2011.

29. MGE DATA (2012) MobiTest GSL, webpage, May 2012, `http://www.mgedata.com/de/hw-und-sw-produkte/custom-produkte/mobitest`.

30. Amt für Verkehr, Volkswirtschaftsdirektion Kanton Zürich (2011) Gesamtverkehrsmodell, webpage, March 2011, `http://www.afv.zh.ch/internet/volkswirtschaftsdirektion/afv/de/home.html`.

31. PTV (2009) Ptv, webpage, `http://www.ptvag.com`.

32. MATSim (2012) Multi Agent Transportation Simulation, webpage, `http://www.matsim.org`.

33. POSDAP (2012) Position Data Processing, webpage, `http://sourceforge.net/projects/posdap/`.

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

35. 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.