Improving post-processing routines for GPS observations using prompted-recall data

Conference Paper

Author(s):
Schüssler, Nadine; Montini, Lara; Dobler, Christoph

Publication date:
2011

Permanent link:
https://doi.org/10.3929/ethz-a-006689506

Originally published in:
Arbeitsberichte Verkehrs- und Raumplanung 724

IMPROVING POST-PROCESSING ROUTINES FOR GPS OBSERVATIONS USING PROMPTED-RECALL DATA

9TH INTERNATIONAL CONFERENCE ON SURVEY METHODS IN TRANSPORT:
TERMAS DE PUYEHUE, CHILE, NOVEMBER 14-18, 2011

Nadine Schüssler
(Corresponding Author), IVT, ETH Zurich, CH-8093 Zurich
Lara Montini, IVT, ETH Zurich, CH-8093 Zurich
Christoph Dobler, IVT, ETH Zurich, CH-8093 Zurich

ABSTRACT

Travel diaries based on person-based GPS observations have become increasingly popular within and beyond the research community due to their manifold advantages compared to classic survey methods such as paper diaries or telephone interviews. However, there are still several open issues concerning the automated post-processing of these large datasets. Without a reliable post-processing, GPS-based studies require either a considerable amount of manual analysis, leading to costly surveys or extensive prompted-recall interviews with the respondents.

This paper reports on a travel diary study conducted using GPS devices and a web-based prompted recall element. The focus is on the survey methodology. The survey participants each carry the GPS devices for a week. Every night their data is uploaded via GSM, processed automatically and the results displayed on a web-site where the respondents can add, delete, review, confirm or correct their trip stages and activities that are depicted on digital maps and described in textual form. The paper presents and discusses the prompted recall tool along with

the newly improved post-processing routines. The post-processing routines use the standard GPS data as well as accelerometer data and the locations of public transport stops to derive stage start and end times and transportation modes.

1 INTRODUCTION AND RELATED WORK

In recent years, travel diaries based on person-based GPS observations have become increasingly popular within and beyond the research community. Nowadays, an increasing number of countries use – or consider the usage of – GPS observations in their National Travel Surveys due to their manifold advantages compared to classic survey methods such as paper diaries or telephone interviews. However, there are still several open issues concerning the automated post-processing of these large datasets. Without a reliable post-processing, GPS-based studies require either a considerable amount of manual analysis, leading to costly surveys, or extensive prompted-recall interviews with the respondents. Prompted-recall interviews place a lot of burden on the participants, thus, violating the promise of reducing participant burden made by researchers since the advent of GPS-based travel behaviour studies. Yet, prompted-recall surveys are the only way to establish reliable post-processing routines.

The post-processing routines that have lately been presented by researchers from all over the world are usually organised in sequential modules and contain the following five steps:

- Cleaning and smoothing
- Detection of trips, stages and activities
- Mode identification
- Activity purpose imputation
- Spatial matching

Out of these five steps, this paper focusses on data cleaning and smoothing, detection of trips, stages and activities and mode identification. A sound **cleaning of the data** is essential for meaningful results in the subsequent post-processing steps due to the variety of error sources of GPS measurements. The most commonly used filtering criteria are the number of satellites in view and the PDOP value (e.g. Wolf *et al.*, 1999; Ogle *et al.*, 2002). If these values are not sufficient or available, the stream of GPS points should be scanned for unrealistic position jumps (e.g. Schüssler and Axhausen, 2009). Minor deviations from the true position do not necessarily have to be filtered but it can help to smooth these positions (e.g. Ogle *et al.*, 2002; Chung and Shalaby, 2005; Jun *et al.*, 2007; Schüssler and Axhausen, 2009).

The **detection of trips and stages** can either be carried out top-down or bottom-up. Top-down in this context means to start with identifying trips and activities and subsequently breaking the trips down into stages (e.g. Tsui and Shalaby, 2006; Schüssler and Axhausen, 2009) whereas bottom-up approaches first determine stop points and afterwards classify them into activities and transfers (e.g. Moiseeva *et al.*, 2010; Marchal *et al.*, 2011). Three basic types of stop points can be distinguished: activities with signal loss, activities with ongoing GPS recording and mode transfers. Activities with signal loss are detected by finding time differences between two consecutive GPS points that are longer than a predefined threshold. Activities with ongoing GPS recording result in speeds close to zero (e.g. Schönfelder *et al.*, 2006; Tsui and Shalaby, 2006; Schüssler and Axhausen, 2009) or bundles of GPS points (e.g. Doherty *et al.*, 2001; Stopher *et al.*, 2005; Schüssler and Axhausen, 2009), i.e. sequences of GPS points positioned very closely to each other. Mode transfers are either characterised by one of the phenomena above or by a change between walking and another mode. These changes can be found using speed and acceleration characteristics of the recorded GPS points (Tsui and Shalaby, 2006; Schüssler and Axhausen, 2009).

Mode detection for person-based GPS can be done with a variety of methods and evaluation criteria. On the one hand, there are rule-based approaches (e.g. de Jong and Mensonides, 2003; Stopher *et al.*, 2005; Chung and Shalaby, 2005; Bohte and Maat, 2008; Marchal *et al.*, 2011) that use criteria such as average or maximum speed, duration of the stage, data quality or proximity to certain network elements (e.g. roads, bus stops or train stations) to derive deterministically the best fitting mode. On the other hand there are fuzzy logic approaches (Tsui and Shalaby, 2006; Schüssler and Axhausen, 2009) and Bayesian inference models (Zheng *et al.*, 2008; Moiseeva *et al.*, 2010) that use similar criteria but account for the fact that many modes have overlapping characteristics, particularly in urban settings, and can therefore only be distinguished with a certain probability.

The addition of a **prompted recall survey** to a GPS diary with automated post-processing serves three different purposes. First, the participants are given the opportunity to correct and validate the results of the post-processing procedures. Second, they are often asked to add information that cannot be imputed from the GPS data, e.g. the number of accompanying persons or the scheduling horizon. Third, the prompted recall survey can deliver input for the processing and for learning procedures. A first approach incorporating such learning procedures in the imputation of modes and activity purposes was presented by Moiseeva *et al.* (2010).

Regarding the format of the prompted recall survey, the researcher can choose between different options. It can either be conducted as a computer assisted personal interview (CAPI) or as a computer assisted telephone interview (CATI) or as a self-guided web-based interview. The recent trend is towards self-guided web-based prompted recall approaches (e.g. Auld *et al.*, 2009; Bohte and Maat, 2009; Clark and Doherty, 2010; Giaimo *et al.*, 2010) Typically, the data is transmitted via internet or mobile phone communication and the participants are able to review their processed data soon after they uploaded it and at a time most convenient for them. Moreover, the web-based format eases the addition of other survey elements such as stated preference experiments (Oliveira *et al.*, 2011) or attitudes and perceptions (Schüssler and Axhausen, 2011; Marchal *et al.*, 2011).

Due to these advantages, the GPS travel diary study reported in this paper also uses a web-based prompted-recall survey. The study participants are asked to carry a person-based GPS receiver for a week and to answer an online survey containing a questionnaire regarding their socio-demographic attributes, psychometric scales and the prompted-recall diary. The GPS data is transmitted via GSM when the respondents charge their device during the night. The transmitted data is automatically processed and the results fed into a database and displayed in the interactive prompted-recall diary. The participants can add, delete, review, confirm or correct their trip stages and activities that are depicted on digital maps and described in textual form. On the one hand, this ensures a travel diary as complete as possible. On the other hand, the corrections made by the respondents are used to improve the post-processing routines where necessary.

The goal of this paper is to discuss the new post-processing routines along with the design of the prompted recall tool and the execution of the survey. Beside an analysis of the influence of respondent burden on participation and completion rate, it is discussed how the pretest of the survey helped to improve and validate the automatic post-processing procedures. Special attention is paid to the characterisation of trip stages and activities that are not detected because either the GPS records are completely missing or they are misinterpreted by the post-processing routines.

The remainder of this paper follows the structure given by these goals. First, the new processing routines are presented. Then, the survey web-site and particularly the prompted recall interface are introduced. Subsequently, the survey design and the current status of the survey are described before the major findings are discussed. The paper closes with conclusions and an outlook on further work.

2 PROCESSING THE GPS DATA

The post-processing routines are based on earlier work by the first author (Schüssler and Axhausen, 2009). The routines for data cleaning, trip stage and activity detection and mode identification have been revised for this work. On the one hand, the routines now integrate information that is additionally available. On the other hand, the trip stage and activity detection as well as the mode identification have been checked and improved using data from this study and a different GPS dataset for which actual diary information was available (Flamm and Kaufmann, 2007).

In contrast to the work described in Schüssler and Axhausen (2009), the post-processing routines can access various sources and types of data in this study. First, there is the most basic information: three-dimensional position and timestamp. Second, the GPS devices used for this work do record the horizontal and vertical accuracy of the position as well as the number of satellites. Third, an accelerometer is integrated in the devices providing three-dimensional acceleration information in 10 Hz intervals. Fourth, the coordinates of public transport stops in Switzerland are known and used in the mode detection.

Another substantial change compared to the work described in Schüssler and Axhausen (2009) is the introduction of *stop points*. A stop point is defined as a place where the person performed an activity or changed the means of transport. The movement between two successive stop points is then defined as a stage conducted with a single means of transport. A stop point is either indicated by the person stopping at a location for a while or by a change between walk and another mode. This differs from the previous approach to first detect activities and subsequently mode changing points. The main argument for this adjustment is that short activities and mode changes cannot be reliably distinguished from each other without looking at the underlying mode chain. Therefore, it makes more sense to first detect stop points and stages and only subsequently merge the stages to trips and tours.

Taking into account these changes, the post-processing mainly consists of three steps:

- Data filtering and smoothing
- Stop point detection
- Mode identification

Its results are stored in an SQL data base and are subsequently presented to the respondents via the prompted recall web-interface. The remainder of this section follows the structure given by the three steps and elaborates the changes made compared to the processing routines described in Schüssler and Axhausen (2009). Only the parameter values valid for all the datasets tested are reported. It turned out that the parameters for many thresholds – e.g. point density, acceleration changes, required accuracy – strongly depend on the device series and have to be determined through careful calibration.

2.1 Data cleaning and smoothing

Little was changed regarding the filtering and smoothing of the data. The smoothing was not changed at all and the filtering was simply extended by an additional filter that uses the signal accuracy information now available. Opposed to standard GPS data not the HDOP and VDOP values were available but accuracy estimations by the device manufacturer in meters. In addition, a filter accounting for the number of satellites in view was tested. However, it was found to be redundant with the horizontal accuracy filter and thus, left out of the final setting. Currently, the accuracy threshold for the horizontal accuracy is set to 25 meters which corresponds more or less to an HDOP value of about 5. This implies a relative high tolerance for inaccurate points. But it was found that otherwise many valid points were unnecessarily deleted and it was more useful to rely on the other filters and the smoothing to take care of the few remaining erroneous points.

2.2 Stop point detection

The stop point detection is a combination of the previous activity detection and the trip segmentation step of the mode detection. Three types of stop point criteria are distinguished during the stop point detection: no movement, signal loss and changes from or to walking and in total, five detection criteria were used:

- Zero speed longer than a certain time threshold
- High density clusters
- Signal loss for longer than a certain time threshold
- Changes from and to walk based on speed and acceleration
- No movement recorded by accelerometer

For the first three criteria, the algorithms for the detection of potential activities by Schüssler and Axhausen (2009) are used. The detection of the changes from and to walk is also taken from Schüssler and Axhausen (2009) but improved by using the accelerometer data instead of the acceleration calculated from the GPS locations. Completely new is "no movement recorded by accelerometer" criterion. For this the median acceleration variation is calculated, which is defined as the median of the moving standard deviation of the length of the accelerometer vector, with a moving window size of 10 data points. If the median acceleration variation falls below a certain threshold for a certain time, a potential stop point is flagged. In the subsequent step potential stop points are joined to actual stop points as described in Schüssler and Axhausen (2009). All criteria apart from "changes from and to walk" are combined and a minimum duration of stop points and stages is enforced. Stop points are at least 120 seconds long and stages at least 60 seconds.

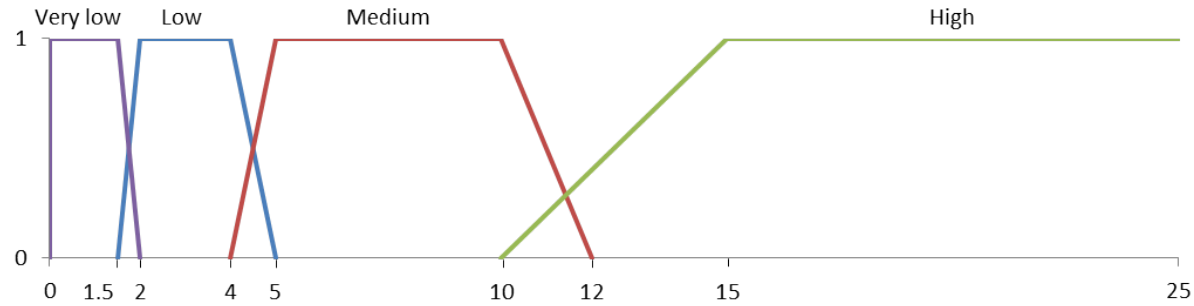
2.3 Mode identification

The mode identification is also based on the work by Schüssler and Axhausen (2009) and uses fuzzy logic to estimate the most likely modes. However, it was improved using additional information and a new system of rules. Five modes are distinguished: walk, bike, car, urban public transport (urban PuT) and rail. Regarding the fuzzy logic rules, it is the goal to keep them as simple and transferable as possible and, therefore, use as few fuzzy variables as possible. Four fuzzy variables were chosen. The median of the speed distribution of a stage is kept, including its four membership functions. The computed acceleration data is replaced by the median acceleration variation. The third criterion remains the 95th percentile of speed but now with only two membership functions. Last, as a new variable the squared distance of the end of a stage to the nearest public transport stop is introduced with two membership functions. It accounts for the issue that the speed and acceleration characteristics of busses, trams and cars are very similar.

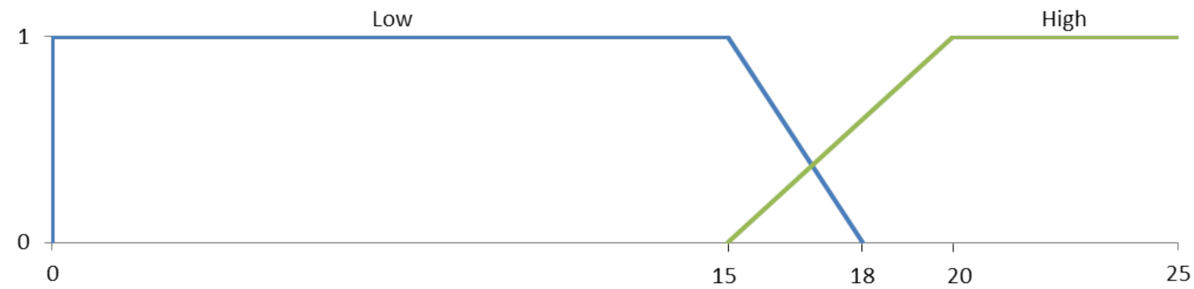
The accelerometer data shows significant differences for the respective modes. The median acceleration variation chosen as a fuzzy variable is a measure of how smoothly the acceleration changes. Typically these values are very high for walking stages, as the device is moved a lot in every direction without constant acceleration. Trains or trams on the other hand accelerate steadily, which results in low values. Four membership functions are used to represent this. All trapezoidal membership functions ar depicted in Figure 1. The corresponding set of 18 rules is shown in Table 1.

Figure 1: Membership functions of the fuzzy variables

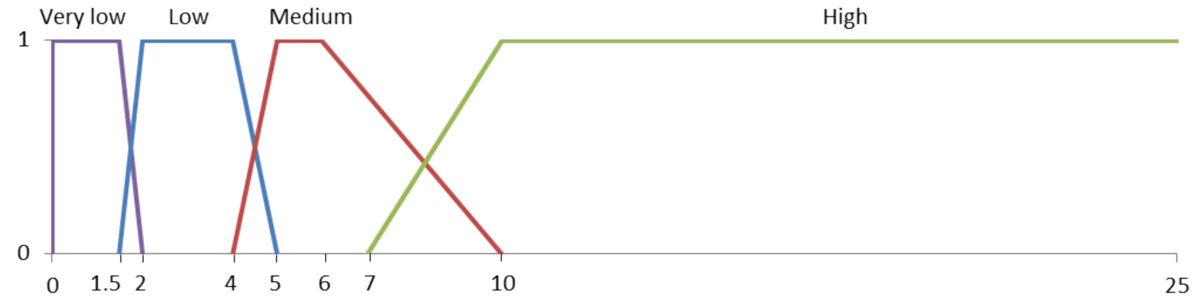
(a) Median speed $[m/s]$



(b) 95th-percentile of speed $[m/s]$



(c) Median acceleration variation $[m/s^2]$



(d) $(distance\ nearest\ public\ transport\ stop)^2$

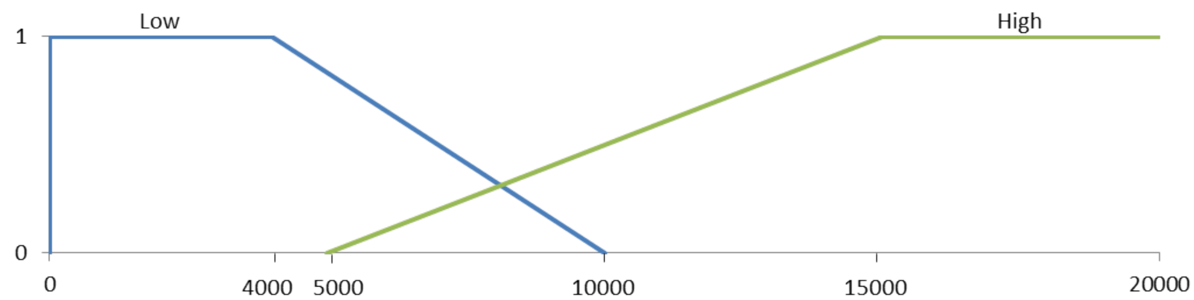


Table 1: Fuzzy rules for mode identification

Median speed	Median acceleration variation	Squared distance nearest PuT stop	95th-percentile speed	Mode
Very low	Very low	-	-	Urban PuT
Very low	Low	-	-	Bike
Very low	Medium	-	-	Bike
Very low	High	-	-	Walk
Low	Very Low	-	-	Urban PuT
Low	Low	-	-	Car
Low	Medium	-	-	Bike
Low	High	-	-	Walk
Medium	Very Low	-	-	Urban PuT
Medium	Low	Low	-	Urban PuT
Medium	Low	High	-	Car
Medium	Medium	-	-	Bike
Medium	High	-	-	Bike
Medium	-	-	High	Car
High	Very low	-	-	Car
High	Low	-	-	Car
High	Medium	-	-	Car
High	High	-	-	Rail

3 DESIGN OF THE ONLINE SURVEY TOOL

The web-based survey consists of three parts: a household and person questionnaire, three psychometric scales and the GPS-based prompted recall diary. In addition, there is a large help section that covers how to use the web elements especially the interactive map in the diary section and how to deal with the questions and the diary.

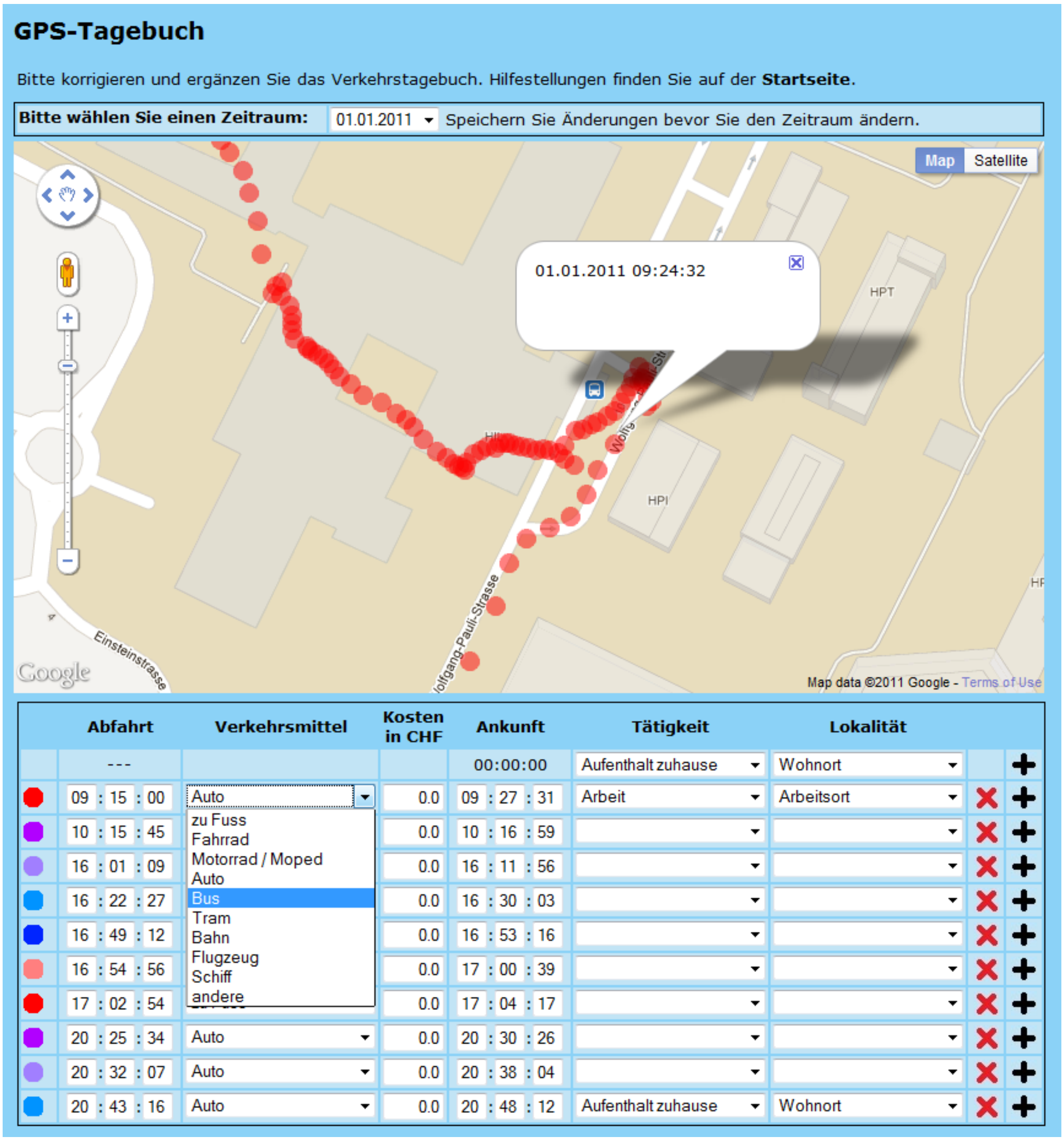
The household and person questionnaire asks for socio-economic attributes and frequently visited locations – home, work or education and the two most frequented grocery shops – and is usually filled out at the first meeting by the participants with the help of the interviewer. The psychometric scales are composed of 95 statements that have to be rated in terms of the level of agreement. They provide measures of the respondents attitude towards risk, environment and variety seeking. For more details on the psychometric scales see the parallel paper Schüssler and Axhausen (2011).

The last – and most extensive and burdensome – part is the GPS-based prompted recall diary. The data collected by the GPS logger is transmitted via GSM when the respondents charge their device. This is ideally done during the night. Every 4 hours, the transmitted data is read, cleaned and saved in a central SQL database. Once a day, currently early in the morning, the GPS post-processing routines are started and the resulting stages, modes and stop points are saved in the SQL database.

Once this is done, the participant can access the processing results via the interactive prompted-recall part of the survey website that is shown in Figure 2. Their task is then to check, correct and confirm the diary information. For this, they first have to choose the day they want to review from a drop-down menu. The GPS points for the chosen day are then presented on an interactive map; the timestamp for each point can be accessed by clicking on the respective point on the map. GPS points of the same stages are depicted in the same colour, stop points are shown in green.

In addition, the diary information is presented in a table below the map. Each diary entry consists of a stage and the subsequent stop point. The attributes for each diary entry contain the start and end time of the stage, the chosen mode, travel costs, the characterisation of the stop point – activity purpose or mode change – and the location of the stop point, chosen from a list of available locations. The list of locations is participant dependent. Locations can be added before or during filling out the diary. A location is defined by a description and its complete and geocoded address. If the address is not known it can be derived using reverse geocoding,

Figure 2: GPS diary user interface



i.e. by marking it on the interactive map.

The participants use the table to review and confirm or correct the diary entries and their attributes. They can delete and add diary entries, i.e. pairs of stage and stop points and change all the attributes provided. Once they hit the save button the corrected information is stored in the data base along with the results of the post-processing routines to allow also an analysis of the changes made by the respondents. Regarding the mode attribute they can select in addition to the five modes used in the mode detection the modes plane, ship or motorbike. Moreover, the mode urban public transport is subdivided into tram and bus, which are both common in the survey area. This is done to collect a more accurate picture of the modes actually used than the mode detection is currently able to deliver. Some experiments to differentiate between tram and bus were conducted but there was not enough difference in the speed and acceleration characteristics to achieve a satisfying results. The same applies for the differentiation between car and motorbike. Finally, the participants can leave free text comments for each day that might help the analyst to interpret their changes.

4 SURVEY DESIGN AND EXECUTION

Before the actual fieldwork was started, a pretest was conducted with 15 members and friends of the Institute. The participants carried the GPS device for approximately four weeks and filled out the prompted recall element of the survey for several days of those four weeks. Their feedback was used to improve the prompted recall interface and the processing algorithms. In addition, the authors analysed part the collected data by hand and extracted stages including the respective mode. The stages were chosen such that all modes of interest were represented, that is walk, bike, car, tram, bus and rail. A total of 322 stages was determined this way and used to improve stop point detection and mode identification.

The aim of the main study is to survey 600 participants within the next year. Participants have to be older than 18 and living within a radius of 22 kilometers around Zurich. The radius was chosen to just include the cities of Winterthur and Zug, which are the regional centers closest to Zurich. The first 1000 addresses were obtained from an address vendor including phone numbers and age of the respondents, the latter to ensure a representative age sample. The recruitment of and the interaction with the participants consists of five steps:

- Mailing of an introduction letter
- Recruitment phone call
- Delivery of the equipment and in-person introduction to the survey website
- Assistance during data collection via phone, email and if necessary in-person
- Returning of the GPS devices via mail in a self-addressed postpaid envelope
- Follow-up calls or mails only if necessary

The introduction letter explains the aim of the survey and announces the recruitment call. Ideally, the participants receive it 2 to 7 days before they are contacted by phone. All recipients are at least called 5 times over several days. After 5 unsuccessful calls the person is categorised as non-responsive.

If the call is successful it is first checked that the person answering the phone is the recipient of the letter. Referencing the introduction letter, the goals and design of the survey and particularly the contribution expected from the participants are explained in detail. No incentives are offered. If the person refuses participation the reasons for doing so is recorded whenever possible. If the person agrees to participate a time and location chosen by the participant are scheduled to deliver the survey box.

The survey box contains the GPS logger, the charging device, a self-addressed postpaid envelope, access information for the survey website and a brochure again explaining the motivation of the survey and the handling of the website elements. The meetings usually last half an hour to an hour. During the meeting, the handling of the GPS logger and the different parts of the website are introduced. The focus is on the usage of the GPS-based prompted recall diary that is demonstrated using an artificial example created for every participant.The example contains common errors like missing signal during rail trips or tunnels where stages have to be merged, wrong mode identification that have to be corrected and randomly occurring wrong points that were not filtered by the post-processing.

At the end of the interview, the participants are given a phone number and an email address where they can reach the survey team in case of any problems or difficulties. From the side of the survey team they are only contacted if necessary. Reasons for such a contact are for example that no new coordinates are uploaded or that the prompted recall diary is not filled out for several days which is monitored using forms directly connected to the central SQL database.

After the participants have completed their seven consecutive survey days they return the GPS device by priority mail service. For this, they simply have to put the device and the charger into the provided envelope and drop it into the next mailbox. The reception is confirmed with a final thank you letter, in which participants are also made aware of the parts of the survey they have not yet completed.

As already mentioned in the introduction, the main survey has just recently started. Thus, it is too early for any quantitative analysis on response patterns. Nevertheless, some initial insights were already gained. Up to now 130 letters have been mailed and 69 persons were reached by telephone. Sixteen persons accepted to participate. Six of those sixteen have already finished the survey, one person interrupted after a day due to health reasons and the rest have an appointment with a survey assistant. Sixty days of GPS data were collected by those participants who finished, of which 25 days are the same person who enjoyed supporting our research.

In two cases diaries were not completed well. Stages were not corrected, only the mode and location information was added. This made us realise, that on the one hand the introductory meeting has to emphasise the diary much more and that participants need more help and reminders during the data collection phase. One measure that is now implemented is to ensure that the participants complete the example data themselves under supervision of the assistant. So far, the interviewers did not insist on this if the participants stated that they understood how everything worked. Furthermore, the monitoring form is currently improved to make a lack of correction by the participants more obvious. Another issue concerns the representativeness of the sample. So far, no persons between the ages of 25 and 40 could be recruited. On the one hand, these participants are hard to reach by phone and if they are reached, they are not very willing to participate. Thus, measures to solve this issue have to be found.

In general we are satisfied with the willingness to participate, especially as the response burden is relatively high compared to other surveys conducted at the institute. Response burden depends on the number of trips participants undertake but also on the quality of the post-processing and therefore on how many corrections participants have to make. Assuming 10 diary entries per day and not yet accounting for imputed fields the burden is estimated at 1300 points according to the scale presented by Axhausen and Weis (2010). An improved estimation of response burden should be calculated after the survey when actual number of corrections and diary entries are known.

5 FINDINGS

This section summarises the lessons learned from analysing the pre-test data and qualitative evaluations of the data provided by the respondents. The focus is on insights that help to improve the methodology and not on examining the participants' transport behaviour. Basis for the validation is the data derived from the pretest described in the previous section. 322 fully annotated stages were obtained through the prompted recall answers of the pretest participants and manual examination of the GPS traces by the analysts.

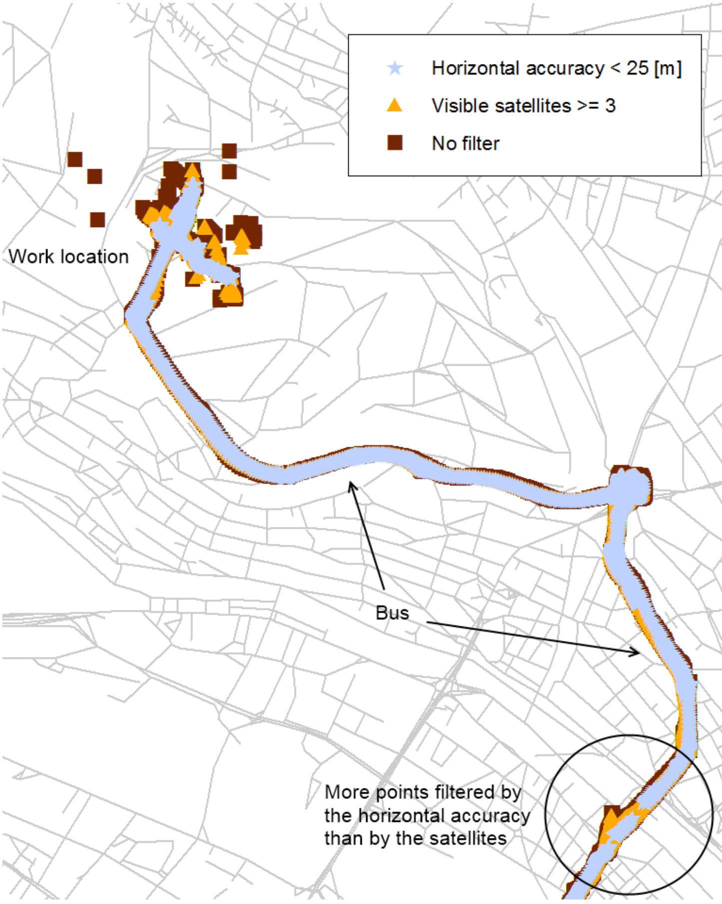
5.1 Data cleaning

The new filters using the number of satellites and the horizontal accuracy are both more restrictive than the altitude filtering. For most of the persons, the original jump point filter did not remove any more points after the maximum horizontal accuracy was set to 25. This accuracy is also enough to filter all points with less than 3 satellites, so the satellite filter is redundant. Altogether the horizontal accuracy filter is the most restrictive filter with only a third to half of the original points remaining after altitude and horizontal accuracy filtering compared to about three quarter of the points for the combination of altitude and jump point filtering. The stricter filtering is is actually an advantage because the subsequent processing steps are not misled by erroneous points and have to handle considerably fewer points leading to shorter computation times. Moreover, it can be seen in the example shown in Figure 3 that the relevant points remain in the dataset. The noisy points around the work location are removed while the the bus stages to and from work are still clearly visible.

5.2 Stop point detection

As a first indicator on the quality of the stop point detection, Figure 4 compares the duration distribution of the detected stages to the one of the actual stages reported in the pretest. Regarding the duration distribution of the actual stages, it can be seen that about two third of the stages are shorter than 10 minutes but that there are also some stages that are longer than an hour. The long stages are part of long-distance trips by car or rail while the majority of the shorter stages was undertaken in urban areas. Overall, the duration distribution of the detected stages follows the same pattern as the one of the actual stages. However, the stop point detection seems to detect slightly more stop points than there were mode transfers and activities resulting in a higher share of short stages. Moreover, several of the long duration stages are missing. But this is only partially caused by the stop point detection. Many of those were rail trips for which no valid

Figure 3: Unfiltered GPS points and remaining points after filtering



GPS points were recorded.

In order to get a better understanding of the wrongly detected stop points, a more detailed comparison between the actual and the detected stages is presented in Table 2. For this, each detected stage is assigned to the actual stage during which it took place. More than one detected stage can be assigned to each actual stage if there are some additional stop points wrongly detected. Since an exact match of reported and detected stages is very unlikely – not the least because the reported starting and end times might not always been precise – a tolerance buffer of 45 seconds around the start and end time of a stage is introduced. Table 2 shows the results of this matching categorised by the number of assigned detected stages, the accuracy of the start and end times and the duration of the actual stages.

It can be seen that 68% of all actual stages were matched by exactly one detected stage. Of

Figure 4: Distribution of stage durations for actual and detected stages

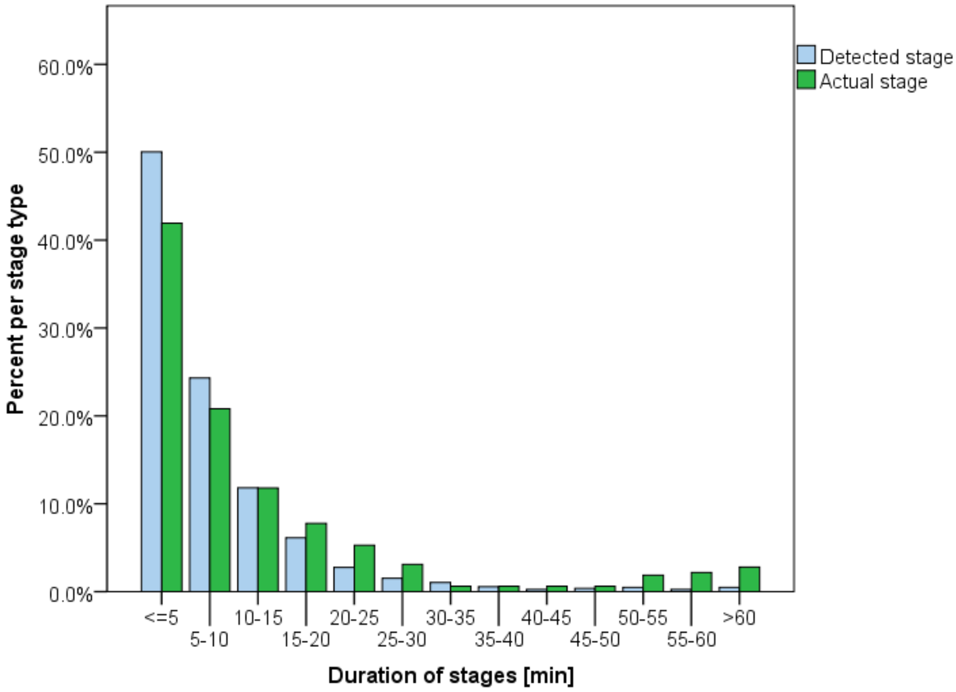


Table 2: Stop point stage identification

Real stage duration	$\lvert\Delta t_{start}\rvert$	$\lvert\Delta t_{end}\rvert$	Assigned detected stages 0	1	2	3	≥4
≤10 min	≤45 s	≤45 s	10	96	3	1	0
		>45 s	0	32	0	0	0
	>45 s	≤45 s	0	30	4	0	0
		>45 s	16	8	2	0	0
>10 min	≤45 s	≤45 s	0	27	16	0	7
		>45 s	0	9	9	3	4
	>45 s	≤45 s	0	15	6	3	2
		>45 s	3	2	6	6	2

those 96 short and 27 longer stages were a very good match as both the difference of start times and end times were within 45 seconds. Overall 86 stages either started or ended within 45 seconds of the actual stage but the other boundary of the stage is missed. In most cases a short stop point, such as a change of means of transport, was missed. In one example, two missed stop points were even short activities, i.e. dropping off other people on the way home. These stop points were very short and therefore not detected. As a result, the start of the first stage is detected correctly but the end is not, for the second stage neither start nor end are detected correctly and for the last stage only the end is correct. Consequently one very reasonably detected stage results in three mismatches. The opposite error occurs if an actual stage is split into several by the algorithm. Most cases are either signal loss in tunnels or on rail stages or are congestion detected as stop points due to the density criterion.

Missing GPS points did not only lead to wrongly split-up stages but also to completely missed stages. Most undetected stages with a duration shorter than 10 minutes are very short walking stages, for example from home to the bus stop, where due to the cold start of the device most of the walking stage is not recorded and the remaining GPS points are incorporated into the "waiting for the bus" stop point. This happens also because detected walking stages have to be at least 60 seconds long, but the analysed stages can be shorter. Regarding the 3 undetected stages that were longer than 10 minutes, one stage was a rail stage with no signal between stations and the other two cases were categorised as stop points due to high density of the GPS points caused once by a car in congestion and once by a promenading rather than a walking stage.

5.3 Mode identification

An overview about the detected and actual modes is given in Figure 5. The figure shows that all modes were covered in the sample data though some modes were less frequently used than others. The smallest number of stages per mode were 14 bike stages. The car stages cover urban trips as well as motorway trips and even though the mode identification does not differentiate between bus and tram and categorises both as urban public transport it was ensured that a fair amount of tram and bus stages are present. The sample also contains many short walking stages which are hard to detect and sometimes even too short for the given threshold of 60 seconds.

The mode identification depends on the characteristics of a stage but also on a correct detection of the stage boundaries. Therefore, stages constructed with the actual stage start and end points were used to assess the quality of the mode identification. Overall, the mode identification detects all modes and the distribution of detected and actual modes is rather similar as depicted in Figure 5, but there is a clear over-detection for bike and an under-detection for rail.

In Table 3, more precise detection rates are given along with a quantification of the misclassi-

Figure 5: Split of modes for actual stages and stages identified during post-processing

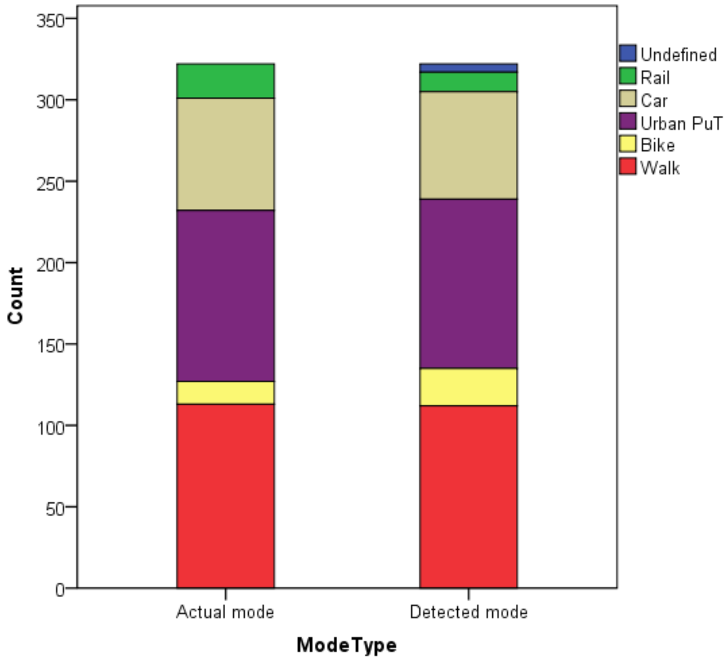


Table 3: Mode identification results

		Mode identified by the post-processing					
		Walk	Bike	Urban PuT	Car	Rail	Undefined
Actual Mode	Walk	108	4	1	0	0	0
	Bike	1	12	1	0	0	0
	Urban PuT	2	2	77	15	9	0
	Car	0	2	9	51	4	3
	Rail	1	3	5	0	10	2

fications. Overall, 83% of the modes were identified correctly. The highest success rate was obtained for walking with 95% of all walking stages correctly detected, the lowest rate was for car stages with 74%. The biggest misclassification problems occur between car and urban public transport which was expected given the very similar speed and acceleration patterns of these two modes. Even the integration of the public transport stops could not solve this issue completely, mainly because quite a number of car stages end close to them.

6 CONCLUSION AND OUTLOOK

This paper reports on a travel diary study conducted using GPS devices and a web-based prompted recall element. The focus is on the survey methodology, i.e. the automatic post-processing routines for the GPS data and the survey design and execution. Since the survey itself just started, only qualitative evaluations of the experiences with and the data provided by the respondents can be made. However, a relatively large pre-test was conducted that allows to conduct some initial analysis of the post-processing routines.

Looking at the response patterns observed so far, the overall response rate is good but some of the respondents seem to have trouble completing the diaries the way it was intended. While the correction of modes and the addition of location information works well the correction of the stage start and end times is not always satisfactory. This will be addressed by an improved training of the respondents during the introductory meeting. Moreover, special attention will be paid to the recruitment of persons between the ages 25 and 40 because they have so far been very hard to reach.

The results of the stop point detection are promising but there is still room for some improvement. A procedure to join stages could offer a significant potential. According to Table 2 27 stages would be completely detected if the detected stages were joined and another 31 would be partially detected by one stage. In many cases the detected stages that could be joined are of the same mode, and if also e.g. bus-car or tram-rail combinations are used as potential joins, which is reasonable as mode detection is good but not perfect, almost all cases are covered. However, some research is still necessary regarding the criteria of when stages should be joined and when they should not. In this process, it will also be useful to have a look at the cold start problem and ways to remedy it.

Another issue that will be looked at are the completely missing rail stages. One approach is to use the gap speed as in Marchal *et al.* (2011) and accelerometer data to confirm if the gap is a stop point or rather a stage. In combination with the location of the rail stations this could probably improve the rail detection.

Since the ultimate goal of the study is to model urban public transport choices, the next step following the survey will be the detection of not only the mode but the precise public transport connection. A combination of map-matching to the public transport network and look-ups on the timetable will be used to achieve this. Therefore, precise information about public transport routes and timetables are essential. Eventually, this work can also be used via a feedback loop to improve the mode detection and enable a differentiation between bus and tram already in the mode detection stage. However, the map-matching is still rather time consuming and for survey purposes the data has to be processed rapidly. Thus, some performance improvements for the map-matching algorithm might be needed as well.

7 ACKNOWLEDGEMENTS

The authors would like to thank the Swiss State Secretariat for Education and Research for funding this research that is part of the research project "Route choice in urban public transport systems" within the COST Action "TU0603 - Buses with high levels of service". Special thanks goes to Ilona Imoberdorf, Patrice Frei and Nathalie Schenk for their support with the field work.

REFERENCES

Auld, J., C. Williams, A. K. Mohammadian and P. Nelson (2009) An automated GPS-based prompted recall survey with learning algorithms, *Transportation Letters*, **1** (1) 59–79.

Axhausen, K. W. and C. Weis (2010) Predicting response rate: A natural experiment, *Survey Practice*, **3** (2).

Bohte, W. and K. Maat (2008) Deriving and validating trip destinations and modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands, paper presented at the *8th International Conference on Survey Methods in Transport*, Annecy, May 2008.

Bohte, W. and K. Maat (2009) Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands, *Transportation Research Part C: Emerging Technologies*, **17** (3) 285–297.

Chung, E.-H. and A. Shalaby (2005) A trip bases reconstruction tool for GPS-based personal travel surveys, *Transportation Planning and Technology*, **28** (5) 381–401.

Clark, A. F. and S. T. Doherty (2010) A multi-instrumented approach to observing the activity rescheduling decision process, *Transportation*, **37** (1) 165–181.

de Jong, R. and W. Mensonides (2003) Wearable GPS device as a data collection method for travel research, *Working Paper*, **ITS-WP-03-02**, Institute of Transport Studies, University of Sydney, Sydney.

Doherty, S. T., C. Noel, M. E. H. Lee-Gosselin, C. Sirois, M. Ueno and F. Theberge (2001) Moving beyond observed outcomes: Integrating Global Positioning Systems and interactive computer-based travel behaviour surveys, *Transportation Research E-Circular*, **C026**, 449–466.

Flamm, M. and V. Kaufmann (2007) Combining person based GPS tracking and prompted recall interviews for a comprehensive investigation of travel behaviour adaptation processes during life course transitions, paper presented at the *11th World Conference on Transportation Research*, Berkeley, June 2007.

Giaimo, G., R. Anderson, L. Wargelin and P. R. Stopher (2010) Will it work? Pilot results from the first large-scale GPS-based household travel survey in the United States, *Transportation Research Record*, **2176**, 26–34.

Jun, J., R. Guensler and J. Ogle (2007) Smoothing methods to minimize impact of Global Positioning System random error on travel distance, speed, and acceleration profile estimates, *Transportation Research Record*, **1972**, 141–150.

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.

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.

Ogle, J., R. Guensler, W. Bachman, M. Koutsak and J. Wolf (2002) Accuracy of Global Positioning System for determining driver performance parameters, *Transportation Research Record*, **1818**, 12–24.

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, paper presented at the *90th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2011.

Schönfelder, S., H. Li, R. Guensler, J. Ogle and K. W. Axhausen (2006) Analysis of commute Atlanta instrumented vehicle GPS data: Destination choice behavior and activity spaces, paper presented at the *85th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 2006.

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

Schüssler, N. and K. W. Axhausen (2011) Combining GPS travel diaries with psychometric scales, paper presented at the *9th International Conference on Survey Methods in Transport*, Termas de Puyehue, November 2011.

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.

Tsui, S. Y. A. and A. Shalaby (2006) An enhanced system for link and mode identification for GPS-based personal travel surveys, *Transportation Research Record*, **1972**, 38–45.

Wolf, J., S. Hallmark, M. Oliveira, R. Guensler and W. Sarasua (1999) Accuracy issues with route choice data collection by using Global Positioning System, *Transportation Research Record*, **1660**, 66–74.

Zheng, Y., L. Liu, L. Wang and X. Xie (2008) Learning transportation mode from raw GPS data for geographic applications on the web, paper presented at the *17th World Wide Web Conference*, Bejing, April 2008.