ETH *zürich*

Processing raw data from global positioning systems without additional information

Journal Article

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

Processing GPS Raw Data Without Additional Information

Nadine Schuessler
(Corresponding Author)
IVT
ETH Zurich
CH-8093 Zurich
phone: +41-44-633 30 85
fax: +41-44-633 10 57
schuessler@ivt.baug.ethz.ch

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

ABSTRACT

Since the first GPS studies in the mid-1990s, this way of surveying individual travel behaviour has gained increasing attention in transport research. Compared to classic travel survey methods researchers benefit from more accurate and reliable information. At the same time, the participants' burden is reduced substantially if the GPS data collection is not accompanied by elaborate questioning. However, without additional information, such as modes and trip purposes, extensive post-processing is required to derive data that can be used for analysis and model estimation. The corresponding procedures are still an ongoing research issue.

This paper describes a post-processing procedure that needs no other input than the most basic GPS raw data: three-dimensional positions and the corresponding timestamps. First, the data is thoroughly cleaned and smoothed. Second, trips and activities are determined. Third, the trips are segmented into single-mode stages and the transport mode for each of the stages is identified. The procedure is applied to GPS records collected in the Swiss cities of Zurich, Winterthur and Geneva. 4882 participants carried an on-person GPS-receiver for 6.65 days on average. The results are compared to the Swiss Microcensus 2005 to demonstrate that derived data is ready for further applications, such as discrete choice model estimations.

INTRODUCTION AND RELATED WORK

Since the first GPS studies in the mid-1990s *(e.g. 1, 2, 3, 4, 5, 6)*, this new way of surveying individual travel behaviour has gained increasing attention in transport research. Compared to classic travel survey methods researchers benefit from more accurate and reliable information about times, geographic locations, and routes as well as from the prevention of trip under-reporting. At the same time, the participants' burden is reduced substantially if the GPS data collection is not accompanied by elaborate questioning to derive additional information, such as trip purposes and transport modes. However, without additional information extensive data processing is required to derive data that can be used for analysis and model estimation. Accordingly, current research focusses on the development of GPS post-processing procedures that allow the researcher to derive all necessary information, such as start and ending time, mode, and trip purpose directly from the GPS records. Moreover, continuously growing sample sizes lead to an increasing demand for automated procedures with low computational cost.

The choice of the approach strongly depends on whether the GPS data was collected vehicle-based or person-based. In vehicle-based studies *(e.g. 7, 8, 9, 10)*, the participants' vehicles are usually equipped with GPS loggers that record only, when the engine of the vehicle is running. Accordingly, the detection of individual trips is relatively easy using the time differences between the recorded points. In addition, short stops during which the engine is not turned off can be found fairly reliably by identifying times when the speed of the vehicle is zero. However, there are also some shortcomings related to vehicle-based data. First and foremost, all other modes are omitted, even though they are essential for the analysis of transport behaviour in an urban environment. Second, the real trip origins and destinations have to be guessed since only vehicle movements are recorded.

Therefore, person-based GPS studies have recently become more popular, although they raise the requirements for the post-processing procedures considerably. In addition to data filtering, trip detection and map-matching, the analyst has to detect the modes used by the participant. Moreover, the method for trip detection needs refinement and the map-matching has to be done either on a multi-modal network or, if that is not available, multiple networks.

As summarised in Stopher *(11)* a couple of authors have started to address these problems *(e.g. 12, 13, 14, 15, 16, 17)*. Basically, all approaches contain individual modules accounting for:

- Data filtering
- Detection of trips and activities
- Mode stage determination
- Mode identification
- Map-matching

Some authors include additional features such as the merging of stages after the mode detection *(17)* or a feedback between the map-matching and the mode detection *(15)*. Yet, all of these methods have so far only been tested on small samples or test scenarios and most still require manual intervention, particularly in the mode detection.

Relying on manual interaction is not feasible for the data set at hand, which contains about 32,000 person-days recorded in the Swiss cities of Zurich, Winterthur and Geneva. The original study was conducted by a private sector company with the aim to explore whether or not participants pass certain billboards. 4882 participants were asked to carry an on-person GPS logger for 6.65 days on average. No additional information, such as modes or trip purposes, were collected. Due to the large amount of data it was decided not to work within a GIS environment but to implement our own procedures in JAVA. Its design, advantages, and shortcomings are

presented in this paper, which extends an earlier paper of the authors *(18)*.

The remainder of this paper is structured as follows: The next section focusses on data cleaning and smoothing. Subsequently, it is described how the GPS records are subdivided into trips and activities. Since each trips can still contain more than one mode, the mode detection, which is presented afterwards, starts with a segmentation of the trips into single-mode stages. The actual mode detection is then executed using a fuzzy logic approach based on the speed and acceleration characteristics of the stages. Before presenting the conclusions and the outlook on future work, the results of the proposed approach are compared to the Swiss Microcensus 2005 since no validation data was available from the study itself.

DATA CLEANING AND DATA SMOOTHING

The positioning accuracy of GPS receivers under ideal conditions lies between five and ten metres *(19)*. In reality, however, it is usually much worse due to several error sources. For instance, there might be less than the four satellites in view that are required to precisely calculate a three-dimensional position. Even if there are enough satellites in view, they might not be ideally positioned, which is expressed by a high position dilution of precision (PDOP) value *(20)*. While this leads to GPS positions that are completely different from the actual position of the receiver, the so-called *warm start/cold start problem* results in missing GPS points at the beginning of the trip due to the time the GPS receiver needs to acquire the position of at least four satellites in view *(21)*.

In addition, there are random errors caused for example by satellite or receiver issues, atmospheric and ionospheric disturbances, multi-path signal reflection or signal blocking *(22)*. Especially burdensome are multi-path errors, also called *urban canyoning errors* because they typically appear in urban canyons. The GPS signal is reflected by buildings, walls or surfaces and the corresponding GPS positions jump and are often scattered around the actual position of the receiver. Signal blocking, on the contrary, leads to missing GPS points and is of special importance for person-based GPS surveys since it varies for the different means of transport. While GPS reception is generally good when the participant is walking, cycling and or travelling by car, it varies considerably for public transport journeys, depending on the proximity of the person to the nearest window *(4, 17)*.

There are several ways to overcome the problems caused by the GPS errors described above. Data filtering, for instance, takes care of systematic errors while data smoothing removes random errors. All these approaches, however, depend on the information available in the study. Previous studies *(e.g. 20, 9)* showed that the number of satellites in view and the PDOP value are fairly efficient measures to determine systematic errors. Unfortunately, they are not accessible here. Therefore, other criteria to identify erroneous data points had to be developed. One of these criteria is the altitude value. Considering the Swiss topology, all points with an altitude value of less than 200 and more than 4200 metres above sea level are removed.

Another criterion are sudden jumps in the position. Position jumps are detected by comparing the distance between two consecutive GPS points with the distance the person could have travelled in the time interval assuming a maximum speed of $50\ m/s$ and a random error buffer of 30 metres. The GPS points are split into so-called quality segments each of which comprises all points between two consecutive position jumps. Subsequently, every two adjacent quality segments are compared and the GPS points of the shorter one are deleted until the end of the quality segment is reached or the distance between every two consecutive GPS points is smaller the threshold defined above. The whole procedure is repeated until all positional jumps are removed. It is important to note, that for this filter the three-dimensional distance is used, because errors in longitude and latitude are often accompanied by fairly strong jumps in the

altitude, while in the subsequent trip and mode detection only the two-dimensional positions are considered.

Concerning the random errors, several approaches were reviewed to select the appropriate smoothing technique. Since no speeds from Doppler measurements are available, speed and acceleration have to be calculated directly from the position and the timestamp of the GPS points. Hence, the position of the GPS points is smoothed rather than the speed. A Gauss kernel smoothing approach was implemented. For each coordinate dimension $c \in x, y, z$ the smoothed value $\tilde{c}(t)$ at time t is individually calculated as

$$\tilde{c}(t) = \frac{\sum_j (w(t_j) \cdot c(t_j))}{\sum_j w(t_j)} \tag{1}$$

with $c(t_j)$ being the raw value of the coordinate c at time t_j and $w(t_j)$ the Gaussian Kernel function computed for each point of time t_j by

$$w(t_j) = \exp{-\frac{(t - t_j)^2}{2\sigma^2}} \tag{2}$$

The Kernel bandwidth, represented by σ, is set to 10 seconds, which results in a 15 second smoothing range. This is assumed to be a reasonable time frame for real behavioural changes as opposed to signal jumps. Accordingly, the directional speed for each coordinate c is the first derivative with respect to t of the smoothed position and the acceleration the second derivative with respect to t.

TRIP AND ACTIVITY DETECTION

The filtered and smoothed GPS points then have to be subdivided into trips and activities. Since the GPS points have been collected person-based, two basic types of activities are considered: activities with ongoing GPS recording and activities with signal loss. Activities with ongoing GPS recording are either characterised by speeds that are close to zero *(e.g. 8, 15)* or by so-called *bundles of GPS points (e.g. 23, 21)*. A bundle is a sequence of GPS points positioned very close to each other, i.e. within a diameter of about 30 metres which equates to approximately thrice the standard deviation of the measurement accuracy *(21)*. Therefore two criteria to detect activities with ongoing recording were established. The first one flags an activity, if the speed is lower than 0.01 m/s for at least 120 seconds and the second, if the *point density* in a sequence of GPS points is higher than 15 for at least 2/3 of the points and the sequence lasts for at least 10 points or 300 seconds. Thereby, the point density is determined by counting how many of the 30 preceding and succeeding GPS points are positioned within a 15 metres radius around the GPS point in question.

Activities with signal loss are detected by means of the time difference between two consecutive GPS points. The threshold beyond which it is assumed that an activity took place varies in the literature between 45 *(6)* and 300 seconds *(e.g. 24, 23)*, whereas most studies apply 120 seconds. In this study, however, a 900 second threshold is used. Compared to former studies this is a high value. But the examination of the GPS points lead to the conclusion that a shorter dwell time would lead to too many wrongly detected activities due to bad reception during trips. As these reception losses will be handled by the map-matching algorithm, they do not need to be considered here.

Each of the three criteria is used individually to determine *potential activity start* and *potential activity end* points. It is important to note, that each activity can be detected by more than one criterion. Consequently, the potential activity start and end points are joined in the way

that the outermost potential activity start and end points are considered to be the true activity start and end points regardless of the criterion they belong to. Moreover, if a new activity starts shortly (maximum 15 GPS points) after the last one has ended, the two activities are joined. This rule on the one hand accounts for measurement errors and on the other hand considers that trips of less than 15 GPS points cannot be reasonably used for route choice modelling. After finding all activity start and end points, activity and trip objects are generated and stored in separate lists. Only the trip objects are used in the subsequent analysis. The activity objects will be used later on to analyse trip purposes as well as trip and activity chains.

Finally, the results of the trip and activity detection were compared to the results of a human analyst on a small sample of persons, since no information about the actual activities was available. 97% of the activities could be detected by at least one of the criteria and none was falsely detected.

MODE DETECTION

So far, only a few approaches have been published that implement an automated mode detection although this is crucial to make person-based GPS usable in large-scale applications. Most approaches comprise two steps: a segmentation into single-mode stages and the mode assignment for each of these stages. In addition, some authors *(e.g. 25, 17)* introduced a third step that accounts for the reasonability of the derived mode chains.

The segmentation of trips into single-mode stages implements the definition, that walking is required for every mode change. The procedure, which follows the mode detection method presented by Chung and Shalaby *(16)* and Tsui and Shalaby *(15)*, exploits the uniqueness of the walk mode with consistently low speeds and accelerations. In addition, a new stage is created if a signal loss, i.e. a gap, of more than 120 s occurs. Three types of *potential mode transfer points (MTP)* are detected: *end of walk (EOW)*, *start of walk (SOW)*, and *end of gap (EOG)* points. The speed and acceleration thresholds for walking are 2.78 m/s and 0.1 m/s^2 , respectively. Consequently, the potential MTPs are aligned, to ensure that each walking stage is enclosed by exactly one SOW (or EOG) and one EOW (or EOG) point, that the speed in a walking stage never exceeds 2.78 m/s, and that the derived stages are sufficiently long. For a walking stage the minimal duration is 60 s and for all other modes 120 s.

Subsequently, the mode for each stage has to be derived. Former studies have applied different approaches to achieve this. Bohte and Maat *(12)*, Stopher *et al. (21)*, Chung and Shalaby *(16)*, and de Jong and Mensonides *(17)* use rules based, for instance, on average and maximum speed, proximity to certain network elements (e.g. bus stops or train stations), or the deviation from the street network. Zheng *et al. (25)* evaluate four inference models (Decision Trees, Bayesian Networks, Support Vector Machines and Conditional Random Fields) and obtain best results for the Decision Tree model. In this study, however, a fuzzy logic approach based on speed and acceleration characteristics is employed, as it has been introduced by Tsui and Shalaby *(15)*. Five modes have been distinguished:

- walk
- cycle
- car
- urban public transport (i.e. bus and tram)
- rail

An open source fuzzy engine *(26)* is used. Three fuzzy variables were chosen: the median of the speed distribution with four membership functions, and the ninety-fifth percentiles of the speed and acceleration distributions, each with three membership functions. These statistical

FIGURE 1   Membership functions of the fuzzy variables

(a) Median Speed

(b) 95 Percentile Acceleration

(c) 95 Percentile Speed
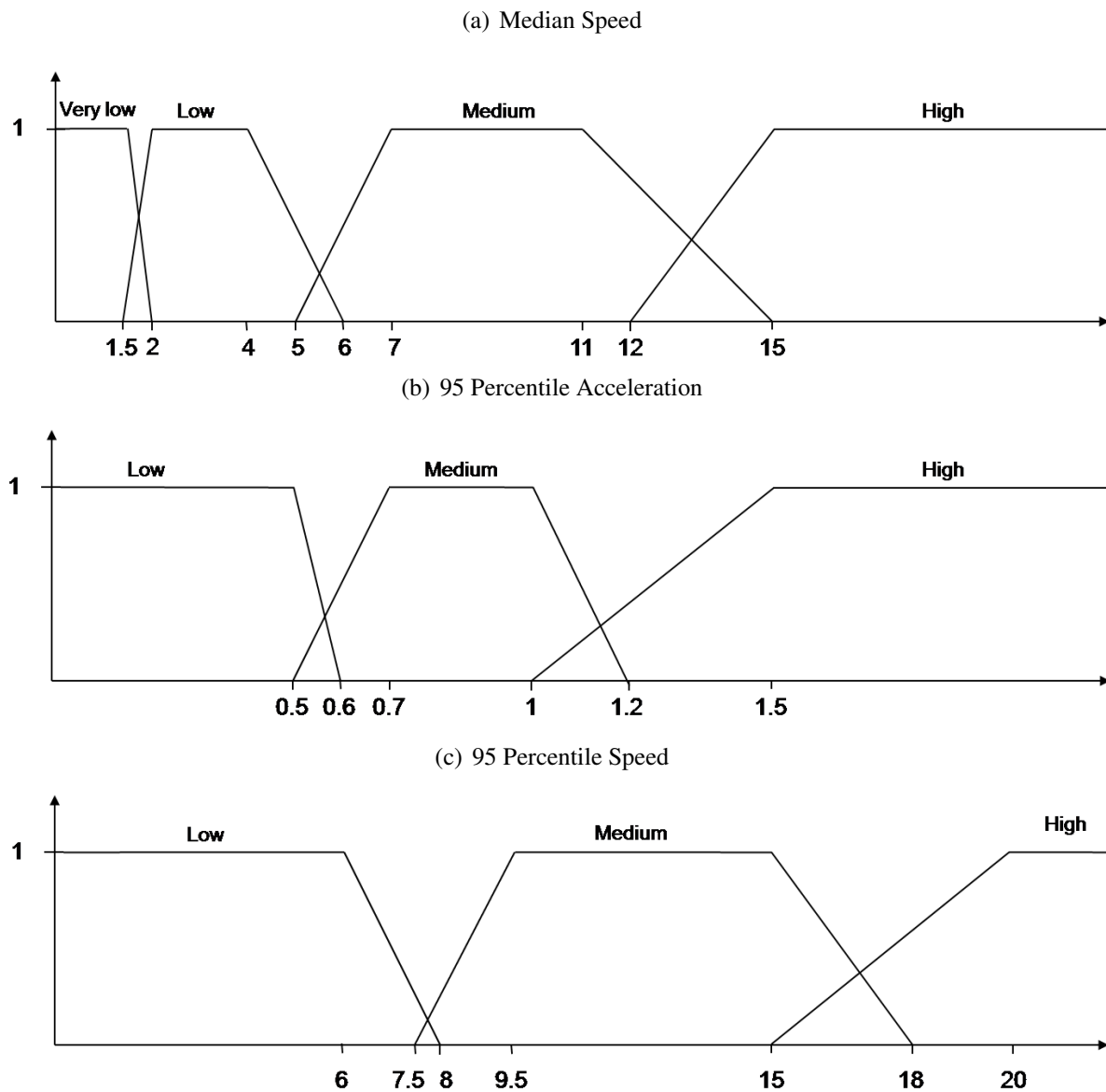
location parameters were explicitly chosen over the average speed or the maximum speed and acceleration to make the algorithm more robust against outliers. The trapezoidal membership functions are depicted in Figure 1. They are defined by four key points: Start point, left top corner, right top corner, and end point. The values for these points were chosen after an analysis of the available modes and the speed and acceleration characteristics in the GPS data.

Having established the membership functions, fuzzy rules are derived. They characterise the different modes with regard to the fuzzy variables. As depicted in Table 1, each mode is described by at least one rule. Since the ranges of the membership functions overlap, more than one rule can apply to the same stage. Hence, multiple modes can be assigned to that stage. These ambiguities are deliberately included in the outcome of the mode detection and allow for feedback loops with the map-matching and other correction methods. The defuzzify method combines the membership values for each individual mode using the AND operator. Thus, the final score for each mode equals the minimum membership value amongst all its

TABLE 1   Fuzzy rules for mode detection

Median speed	95 perc. acceleration	95 perc. speed	Mode
very low	low	—	Walk
very low	medium	—	Cycle
very low	high	—	Cycle
low	low	low	Cycle
low	low	medium	UrbanPuT
low	low	high	Car
low	medium	—	UrbanPuT
low	high	low	UrbanPuT
low	high	medium	Car
low	high	high	Car
medium	low	—	UrbanPuT
medium	medium	—	Car
medium	high	—	Car
high	low	—	Rail
high	medium	—	Car
high	high	—	Car

rules. Subsequently, the likelihood for each mode is calculated based on all the mode scores of a stage.

In a third step, the reasonability of the derived mode chains is investigated. Due to the design of the stage generation process described above, every two non-walking stages are either separated by a walking stage or by a time gap of at least 120 seconds. The latter accounts for the fact, that the analyst cannot know for certain whether a mode change has or has not occurred during that time. This, however, sometimes leads to relatively unlikely mode transitions, for example a direct transfer from car to train. Therefore, the characteristics of the time gap are scrutinised, in particular the average speed during the time gap. If the average speed is higher than walking speed and the two neighbouring stages are both non-walk stages, it is assumed that no mode transfer happened during the time gap. The stages are joined and the mode detection is repeated for the joined stage. In addition, two adjacent walk stages are joined if the average speed in the time gap is below the walk speed threshold. The threshold for walking speed is thereby set to 2 m/s. Thus, not only more realistic mode chains were derived, but also the average distance and duration of the stages were increased, which had been too short in the earlier version of the algorithm presented by Schüssler and Axhausen *(18)*.

Furthermore, the mode detection was embedded in a bigger framework that allows for analyses and corrections of the resulting mode chains. One correction, that will be implemented in the near future, is the deviation from the mode-specific transport network. This will be used to better distinguish for example inter-urban rail trips from car trips. Another potential approach considers the likelihood of the mode transitions and mode transition sequences, as it was employed by Zheng *et al. (25)*. This, however, requires further research, not least because available survey data is not detailed enough.

TABLE 2   Overall statistics of the GPS study compared to the Microcensus 2005

|	Zurich	Winterthur	Geneva	MZ 2005
Number of persons	2 435	1 086	1 361	3199
Number of days per person	6.99	5.96	6.51	1.00
Number of trips per day	4.50	3.40	4.26	3.65
Average trip distance [km]	7.72	7.37	7.19	8.79
Average daily mileage [km]	34.74	23.20	29.25	32.13
Average trip duration [min]	15.17	13.71	15.05	26.21
Average number of stages per trip	1.40	1.31	1.47	1.68

ANALYSIS OF THE RESULTS

Since no information about the actual trips and activities of the participants is available, the Swiss Microcensus on Travel Behaviour 2005 (MZ 2005) *(27)* is used as the basis for the validation of the post-processing procedure. The Swiss Microcensus on Travel Behaviour (MZ) is conducted every five years and delivers a representative and detailed insight into the travel patterns of the Swiss population. In 2005, 33,390 individuals reported in the course of a computer-assisted telephone interview (CATI) on their socio-economic background, their mobility tools, and their, stage-based recorded, trips and activities on the reporting day. In the following, the results of the post-processing procedure are compared to a sub-sample of the Microcensus 2005, which comprises the respondents living in Zurich, Winterthur or Geneva.

The overall statistics for the three sub-studies for Zurich, Winterthur and Geneva, and the corresponding sample in the MZ 2005 are represented in Table 2. It can be seen that the values vary between the cities. The number of days observed per participant, for example, is about seven in Zurich compared to approximately six days in Winterthur. Overall, however, the numbers reveal the same trend. Except for Winterthur, the number of trips per day is higher than the one reported in the MZ 2005 and the average trip distance and duration are smaller. This effect was expected because several previous studies *(e.g. 7, 28, 3, 29)* could demonstrate that in particular shorter trips and activities are under-reported in recall-based surveys.

The distribution of the number of trips per day is depicted in Figure 2. It can be seen that the distributions are similarly skewed positive although the tail of the one derived from the GPS data is slightly longer and the two distinct peaks for two and four trips per day that are present in the MZ 2005 are not replicated in the GPS data. Both effects were expected since the omission of shorter trips and intermediate activities, e.g. shopping on the way home from work, in recall-based surveys leads to fewer and shorter trip chains.

The distributions of the trip lengths and durations presented in Figure 3 reveal similar patterns in the MZ 2005 and the GPS data. This confirms that the trip and activity detection works properly. Moreover, the graphs reveal that the rounding of times and distances is an important issue in recall-based surveys. Estimating distances and times accurately is a real challenge for the respondents. In the Swiss Microcensus, distances are frequently rounded to km values and durations to quarter hours. This underlines once more the advantage of GPS over recall-based surveys with respect to temporal and spacial accuracy.

In Figure 4, the distribution of the mode per stage is compared to the MZ 2005. Since the fuzzy logic algorithm delivers probabilities and not crisp values, for each stage the mode with the highest probability is chosen. The walk stages are excluded because they are only partially comparable. On the one hand, around 15% of the mode transitions in the MZ 2005

FIGURE 2   Trips per day compared with the Swiss Microcensus 2005
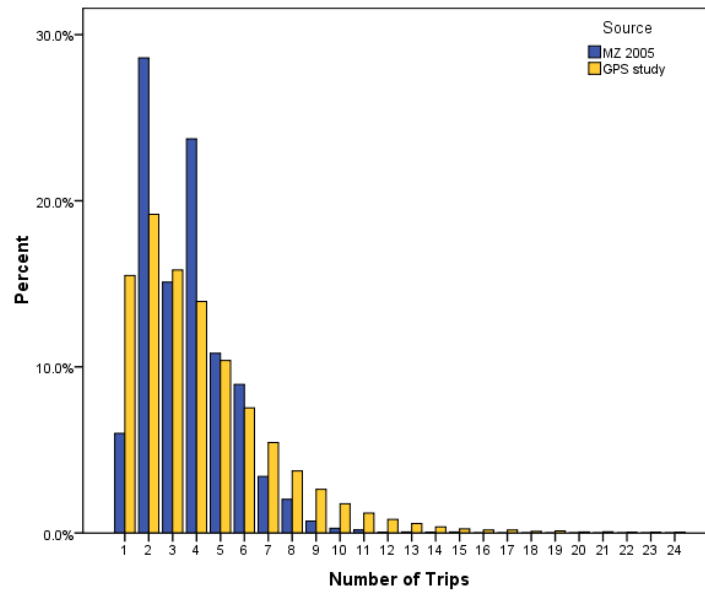
FIGURE 3   Trip distance and duration distributions compared with the Swiss Microcensus 2005

(a) Trip distance distribution

(b) Trip duration distribution
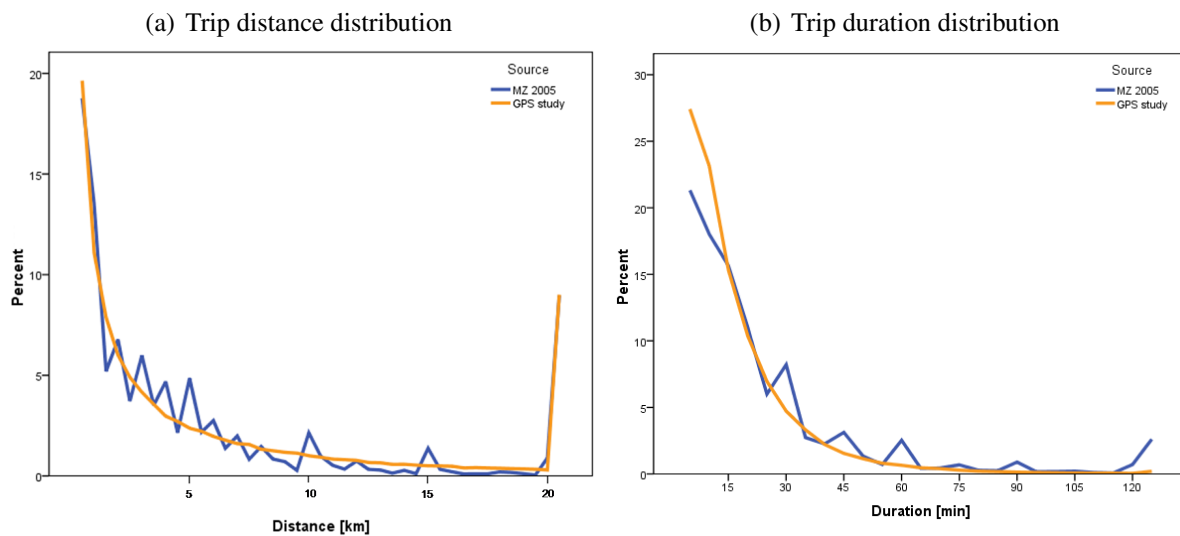
do not meet the basic assumption that every two non-walking stages should be separated by a walk stage. On the other hand, the trips derived from the GPS data suffer from the warm start/cold start problem. Accordingly, 43% of the GPS trips start with a car, rail or urban public transport stage as opposed to 31% in the MZ 2005. Figure 4 reveals that rail stages are fairly under-represented at this phase of the analysis. This might be due to the design of the study which focussed on passing urban bill-boards and not on inter-urban travel. Another problem is the generally bad GPS signal reception in trains. However, some rail trips are most probably misclassified as car stages, likewise some of the missing urban public transport stages. Two

FIGURE 4   Distribution of stage modes compared to the Swiss Microcensus 2005
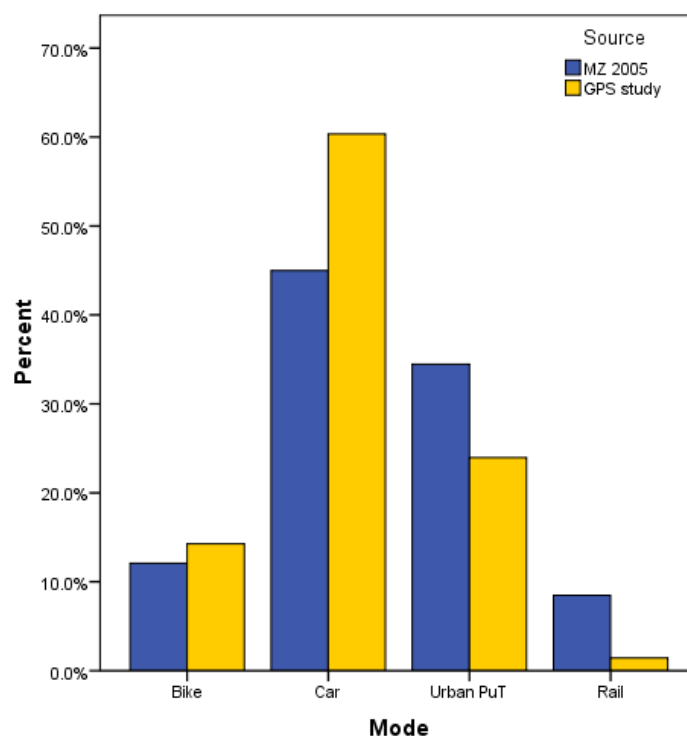
approaches to correct this are envisaged in the future. First, the mode detection parameters will be further optimised. Second, a feedback-loop between the map-matching and the mode-detection will be implemented.

Even closer to the MZ 2005 is the distribution of the stage distance per mode, as depicted in Figure 5. In both surveys, this distributions follow reasonable patterns, e.g. walk trips are rather short, while car and rail are predominantly used for longer trips. Overall, the distance distributions of the individual modes derived from the GPS data are very similar to the ones reported in the MZ 2005. This underlines the quality of the mode detection procedure, especially since the stage distance is not used in the procedure. Only the detected bike trips are quite short compared to the MZ 2005. This might indicate that some of the bike trips are in reality misclassified walk trips. This has to be further analysed in the future.

COMPUTATIONAL PERFORMANCE
One of the main reasons for implementing our GPS post-processing procedure in Java and not in a GIS environment was the requirement to handle the large amount of data in this study in a reasonable computation time. This objective was fully achieved. The computation was performed on a machine with 4 dual-core AMD Opteron 2 GHz CPUs, 4 GB RAM and a Debian Linux version 2.618-5-amd64. The Java 1.5 program processed the complete data set containing about 64.5 million GPS points using one CPU and 1 GB allocated memory in 7091 seconds resulting in an average speed of about 9100 GPS points per second.

CONCLUSION AND OUTLOOK
The data obtained from the post-processing procedure described in this paper will primarily be used to develop route and destination choice models. The focus will be on the different

FIGURE 5 Distance distribution per mode compared to Microcensus 2005

(a) Walk

(b) Bike

(c) Car

(d) Urban public transport

(e) Rail
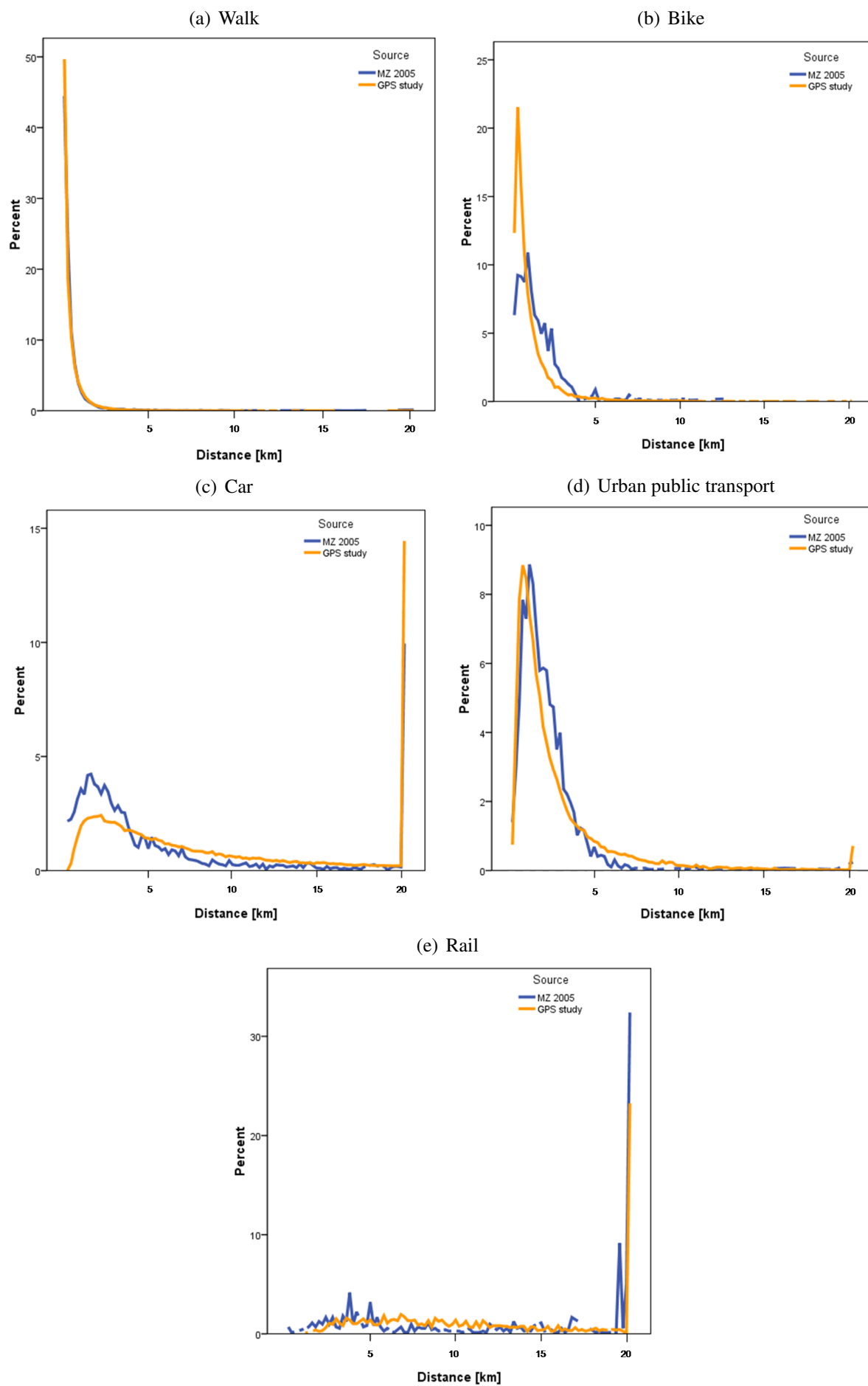
ways to account for similarities in discrete choice modelling. The person-based GPS data set at hand is ideal for this purpose. It covers complete trip and activity chains over several days, thus describing the participants travel behaviour in the most comprehensive way. In addition, the sample size is large enough to obtain stable choice models and is now usable since the procedure presented in this paper delivers sound results in a reasonable computation time even though only the most basic information was available.

A key success factor for the post-processing procedure is the implementation of appropriate filtering and smoothing techniques. Finding the right approach was essential in the study at hand because no information about the numbers of satellites in view or their positioning was available. Though filtering based on altitude levels and unrealistic positional jump was essential, it was not sufficient. Therefore, a Gauss kernel smoothing had to be applied as well.

The activity dwell time of 900 seconds employed in the trip and activity detection is rather high. Shorter dwell times, however, led to too many false trip ends. Instead, the detection of shorter activities is based on the identification of point bundles and zero speeds. With these criteria, the trip and activity detection delivers results very similar to those derived manually, which had to be the frame of reference since no information about the actual trips and activities was available.

For the same reason, the results of the fuzzy logic mode detection could only be compared to the MZ 2005. The comparison showed that mode detection yields realistic results. Especially, the distance distributions per mode are very satisfactory. However, a more detailed validation of the mode detection will be carried out soon with the handcoded GPS data presented by Flamm and Kaufmann *(14)*. First, the mode detection parameters alone will be investigated. Then the whole trip will be examined with regard to the stage generation as well as the actual mode detection. A method for the first analysis has already been implemented by Upadhyay *et al. (30)* and tested for an earlier version of the post-processing procedure. For the second step, an advanced multi-dimensional sequence alignment method developed by Wilson *et al. (31)* will be used.

The map-matching will be done with the algorithm described in Marchal *et al. (32)* because it produces accurate results and efficiently handles large data volumes. For a start, the map-matching will be employed externally and the results fed back into the mode detection. Due to the modular framework of the mode chain correction, the deviation from the mode-specific transport network can be straightforwardly used to distinguish for example inter-urban rail trips from car trips. Eventually, the map-matching ought to be integrated into the post-processing procedure.

The last step of the post-processing procedure will be the determination of activity purposes. A couple of researchers have already worked on that *(e.g. 33, 34, 24, 21)*. Their input will be reviewed to derive an approach suitable for the data available in this study. Key criteria will be observed activity patterns such locations that are visited multiple times and the time of day and duration of activities. In addition, detailed land-use date will be used.

REFERENCES

1. Wagner, D. P. (1997) Lexington area travel data collection test: GPS for personal travel surveys, *Final Report*, Office of Highway Policy Information and Office of Technology Applications, Federal Highway Administration, Battelle Transport Division, Columbus, September 1997.

2. Casas, J. and C. Arce (1999) Trip reporting in household travel diaries: A comparison to GPS-collected data, paper presented at *the 78th Annual Meeting of the Transportation Research Board*, Washington, D.C., January 1999.

3. Yalamanchili, L., R. M. Pendyala, N. Prabaharan and P. Chakravarty (1999) Analysis of Global Positioning System-based data collection methods for capturing multistop trip-chaining behavior, *Transportation Research Record*, **1660**, 58–65.

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

5. Wolf, J. (2000) Using GPS data loggers to replace travel diaries in the collection of travel data, Ph.D. Thesis, Georgia Institute of Technology, Atlanta.

6. Pearson, D. (2001) Global Positioning System (GPS) and travel surveys: Results from the 1997 Austin household survey, paper presented at *8th Conference on the Application of Transportation Planning Methods*, Corpus Christi, April 2001.

7. Du, J. and L. Aultman-Hall (2007) Increasing the accuracy of trip rate information from passive multi-day GPS travel datasets: Automatic trip end identification issues, *Transportation Research Part A: Policy and Practice*, **41** (3) 220–232.

8. 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.

9. 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.

10. Biding, T. and G. Lind (2002) Intelligent Stöd för Anpassning av hastighet (ISA), Resultat av storskalig försöksverksamhet i Borlänge, Lidköping, lund och Umea under perioden 1999–2002, *Research Report*, Vaegverket, Borlaenge.

11. Stopher, P. R. (2008) Collecting and processing data from mobile technologies, paper presented at *8th International Conference on Survey Methods in Transport*, Annecy, May 2008.

12. 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 *8th International Conference on Survey Methods in Transport*, Annecy, May 2008.

13. Stopher, P. R., Q. Jiang and C. FitzGerald (2007) Deducing mode and purpose from GPS data, paper presented at *11th TRB National Transportation Planning Applications Conference*, Daytona Beach, May 2007.

14. 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.

15. 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.

16. 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.

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

18. Schüssler, N. and K. W. Axhausen (2008) Identifying trips and activities and their characteristics from GPS raw data without further information, paper presented at *8th International Conference on Survey Methods in Transport*, Annecy, May 2008.

19. Wolf, J. (2006) Applications of new technologies in travel surveys, in P. R. Stopher and C. C. Stecher (eds.) *Travel Survey Methods - Quality and Future Directions*, 531–544, Elsevier, Oxford.

20. 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.

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

22. 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.

23. 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.

24. Wolf, J., S. Schönfelder, U. Samaga, M. Oliveira and K. W. Axhausen (2004) Eighty weeks of Global Positioning System traces, *Transportation Research Record*, **1870**, 46–54.

25. 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.

26. Sazonov, E. S., P. Klinkhachorn, H. V. GangaRao and U. B. Halabe (2002) Fuzzy logic expert system for automated damage detection from changes in strain energy mode shapes, *Nondestructive Testing and Evaluation*, **18** (1) 1–17.

27. Swiss Federal Statistical Office (2006) *Ergebnisse des Mikrozensus 2005 zum Verkehrsverhalten*, Swiss Federal Statistical Office, Neuchatel.

28. Wolf, J., M. Oliveira and M. Thompson (2003) Impact of underreporting on mileage and travel time estimates - results from Global Positioning System-enhanced household travel survey, *Transportation Research Record*, **1854**, 189–198.

29. Bricka, S. and C. R. Bhat (2006) A comparative analysis of GPS-based and travel survey-based data, *Transportation Research Record*, **1972**, 9–20.

30. Upadhyay, D., N. Schüssler, K. W. Axhausen, M. Flamm and V. Kaufmann (2008) Optimal parameter values for mode detection in GPS post-processing: An experiment, *Working Paper*, **509**, IVT, ETH Zurich, Zurich.

31. Wilson, C., A. S. Harvey and J. Thompson (2005) Clustalg: Software for analysis of activities and sequential events, paper presented at *the Workshop on Sequence Alignment Methods*, Halifax, October 2005.

32. 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 Zurich, *Transportation Research Record*, **1935**, 93–100.

33. Wolf, J., R. Guensler and W. Bachman (2001) Elimination of the travel diary - experiment to derive trip purpose from Global Positioning System travel data, *Transportation Research Record*, **1768**, 125–134.

34. Schönfelder, S. and U. Samaga (2003) Where do you want to go today? - more observations on daily mobility, paper presented at *the 3th Swiss Transport Research Conference*, Ascona, March 2003.