Exploring the potentials of automatically collected GPS data for travel behaviour analysis
a Swedish data source

Author(s): Schönfelder, Stefan; Axhausen, Kay W.; Antille, Nicolas; Bierlaire, Michel

Publication Date: 2002

Permanent Link: https://doi.org/10.3929/ethz-a-004403386

Rights / License: In Copyright - Non-Commercial Use Permitted
Exploring the potentials of automatically collected GPS data for travel behaviour analysis – A Swedish data source

Stefan SCHÖNFELDER and Kay W. AXHAUSEN
Institut für Verkehrsplanung, Transporttechnik, Strassen- und Eisenbahnbau (IVT), ETH Zürich

Nicolas ANTILLE and Michel BIERLAIRE
Chaire de recherche opérationelle, Département de Mathématiques, EPF Lausanne

ABSTRACT

Understanding the regularity and the variability of individual travel behaviour over time has been one of the key issues in travel behaviour research for three decades. A deeper insight into the long-term mobility patterns of persons and households has been so far restricted by the limited availability of longitudinal data, though.

This paper presents an innovative approach to gain longitudinal travel behaviour data by means of Global Positioning Systems (GPS) and furthermore, describes the outline of post-data-processing work as well as possible analysis and modelling directions.

1. TEMPORAL ASPECTS OF TRAVEL BEHAVIOUR: CONTEXT AND DATA NEEDS

Transportation research has so far mainly focused on the traditional cross-sectional analysis of persons’ and households’ mobility patterns. The main interest has been on differences of travel behaviour between travellers or groups of travellers (inter-personal level) which can be covered easily by one-day or few-days travel diary data (e.g. KONTIV). Hence, mobility patterns observed on single days have been interpreted as the optimal decision of the traveller and as a state of behavioural equilibrium – which is assumed to be existent for any point of time and any situation. The investigation of the intra-personal level of travel (see Figure 1) which describes the variability of single persons’ and households’ mobility patterns over time has been restricted so far by the absence of survey data longer than one week and suitable methodology to treat such data. From a
scientific point of view but especially taking into account the planners’ legitimate requirements of data and results concerning the temporal aspects of travel, the available knowledge is insufficient. This led to the inadequate explanation of travel motives and determinants and therefore to problems with the implementation of transport policies.

Fig. 1: Inter-personal and intra-personal level of travel behaviour

Based on the experiences made with earlier approaches to collect longitudinal mobility data, the research project Mobidrive (1998-2000) – funded by the German Federal Ministry of Education and Research (BMBF) – was established to address the described research deficit and to update data availability (Axhausen, Zimmermann, Schönfelder, Rindsfüser and Haupt, 2002). With the implementation of a continuous six-week travel diary as core of Mobidrive (see Axhausen et al., 2002), a current data set of long-term individual travel behaviour is now available for analysis. The extensive investigation of the data during the last years has led to the development and the adoption of a range of analysis and modelling approaches for long-term travel (see Zimmermann et al., 2001 for a comprehensive overview).

2. The Potentials of Automatically Collected GPS Data

Collecting long-term mobility data on the person and household level remains a challenge for travel behaviour research. Despite the Mobidrive success in data collection, it is questionable if for bigger samples the same extent of intensive support of the respondents could be guaranteed. The
potential drawbacks of travel surveys based on self-administered paper based survey designs are well known, and are especially worth considering for the observation of long-term mobility patterns. These include among others

- the limited pool of respondents for long-term travel behaviour studies
- item and unit non-response
- reporting errors / inaccuracy (especially times/durations and distances)
- fatigue effects in longitudinal surveys

Besides, most travel survey data lack information for the route choice decisions of motorised individual transport which is a substantial pillar of travel demand modelling.

This has raised the interest in longitudinal data bases – eventually covering even longer periods of time than in Mobi\textit{drive} – without the high expenses for data collection, though. Possible approaches to meet this requirement are a) to reduce the respondents burden in ordinary paper based instruments by frequent activities elements (see Massot, Madre and Armoogum, 2000; Schlich and Schönfelder, 2001 for examples) or b) the use of Computer Assisted Data Collection (CADAC) methods which has been promoted extensively in the last years (see Leeuw and Nicholls, 1996).

An even further reaching technical approach is the automatic collection of travel behaviour data by in-vehicle or on-person GPS applications. GPS provides specially coded satellite signals that can be processed in a GPS receiver, enabling the receiver to compute location, speed and time. Trip data generated by GPS is getting especially appealing for travel analysis if it is connected with GIS applications which offer digital mapping (e.g. including detailed land use information). The combination of the two methodologies promises to accurately track movements of vehicles or individuals which may supplement or even substitute ordinary travel diary survey designs. The technique offers data accuracy and comprehensiveness which would never be reached by ordinary paper surveys, especially in the detection of micro space-time details (short trips, speeds etc.), obtaining route choice behaviour information and the extension of the observation period (Lee-Gosselin, 2002). Normally, the positioning data is accurate to only few meters.
Figure 2 shows an example of a technical implementation of automated GPS data collection in a Dutch pilot study (Draijer, Kalfs and Perdok, 2000). Similar settings have been applied elsewhere. Here, the mobile data collection system consisted of a GPS receiver with RDS/FM correction\(^1\), a data storage device / micro-computer (called GEOMATE) and mobile power supply. The equipment used in this study was portable, and had the total size of a video-camera, whereas in several other studies the instruments were installed permanently in cars. For each trip (irrespective of made by car, foot or public transport), the respondents switched on the system independently which started data transmission to the computer (GEOMATE) in intervals of 4 to 10 seconds. After data collection, the data was transferred from GEOMATE to a conventional PC for processing.

![Diagram of Automated GPS Data Collection](image.png)

Fig. 2: Automated GPS data collection: Example of technical set-up; adopted from: Draijer et al. (2000) 148

2.1 Recent experiences

Since mid 1990s, GPS data loggers in connection with GI systems have been applied to data collection in several initial feasibility studies (see Table

---

\(^1\) RDS/FM: Radio Data System, FM: Frequency Modulation; RDS/FM signals correct the various inaccuracies in the GPS system which finally leads to extremely high spatial accuracy compared with conventional GPS. This concept is called Differential Global Positioning System (DGPS).
Exploring the potentials of automatically collected GPS data for travel behaviour analysis

1 for an overview). Lee-Gosselin (2002) summarises the existing studies by the categories

- “Imitating” traditional travel diaries,
- Passive monitoring and
- Hybrid approaches.

The first category may be defined as a mixed data collection approach: Travellers are equipped with a GPS device (mostly fitted in the person’s vehicle; for an exception see Draijer et al., 2000) and a hand-held computer (e.g. Personal Digital Assistant, PDA). The latter device is provided in order to substitute the ordinary travel-diary form by the digital/computer assisted input of further trip attributes. Hence, there are two base sources of information: (1) the GPS tracking data covering start and end times, position and speed recorded in short intervals (mainly 1 to 10 seconds) as well as travel distances and (2) activity purposes, car occupancies etc. reported by the travellers themselves via PDA input.

Passive monitoring describes the collection of trip data by passive in-vehicle GPS systems without any further information input by the drivers (i.e. no driver-computer-interaction). The data collection of those studies is mainly traffic safety driven, i.e. the research focus lies on the investigation of the relationship between driving behaviour (speeds) and crash risks.

Finally, hybrid approaches of passive monitoring plus interim or subsequent contacts with the respondents have been developed in the last few years. In a Canadian study which aims to obtain deeper insight into activity-scheduling processes of travellers, personal traces collected by in-vehicle GPS devices was made available to respondents by means of an inter-active GIS application (Doherty and Lee-Gosselin, 2000). The idea was to “help” the respondents – by visualising their own spatial movements patterns – to optimise their daily activity patterns and to eventually provide recommendations for a change of spatial behaviour.
Tab. 1: GPS and travel behaviour analysis: Selection of recent studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Technical details / collection procedure</th>
<th>Main research objectives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexington Area Travel Data Collection Test 1996 (Batelle, 1997)</td>
<td>Passive recording plus interactive input by the drivers PDA; post-usage interviews (see above)</td>
<td>Acceptance of using automatic collection devices; test of passive and interactive reporting; analysis of route choice information</td>
</tr>
<tr>
<td>1997-1998 Austin Household Survey (Pearson, 2001)</td>
<td>Passive vehicle-based GPS recording; 200 vehicles; additional paper and pencil diary</td>
<td>Feasibility; underreporting of trips in ordinary paper surveys; identification of trip ends in GPS data</td>
</tr>
<tr>
<td>Transport Research Centre (AVV) experiment 1997, (Draijer et al., 2000); several cities in the Netherlands</td>
<td>Mixed design: Passive GPS recording by mobile equipment plus paper-pencil diary as well as GPS/pencil diary only; total sample size: 150</td>
<td>Acceptance of survey methodology; test of mobile GPS devices (in hand-held computer); test of suitability for all travel modes</td>
</tr>
<tr>
<td>Georgia Tech experiment 2000 (Wolf, 2000; Wolf, Guensler and Bachman, 2001); Atlanta / Georgia</td>
<td>Passive in-vehicle GPS system plus paper trip diary for part of the sample; 30 respondents</td>
<td>Possibility of total substitution of paper travel diaries; post-data processing issues</td>
</tr>
<tr>
<td>SMARTRAQ / Drive Atlanta, start: 2002, Atlanta (Wolf, Guensler, Frank and Ogle, 2000; Sanders, 2002)</td>
<td>Passive monitoring of about 1.100 vehicles, up to two-years monitoring period plus paper travel diaries</td>
<td>(1) Traffic safety and travel behaviour issues; (2) Physical activity of the respondents; (3) Air quality issues</td>
</tr>
</tbody>
</table>

Summarising the outcomes of the recent feasibility studies, they prove the feasibility of automatic data collection for studying travel behaviour, but also technical and operational difficulties. The advantages of data collection by GPS are manifold (see e.g. Draijer et al., 2000; Wolf et al., 2001), such as

- the reduction (or even elimination) of respondents’ burden
- the availability of path choice information
- the high level of spatial accuracy
- the fact that data is generated in digital format which allows direct analysis.
As potential drawbacks, the following aspects were identified:

- the possibility of sporadic or even systematic technical problems of transmission, eventually leading to total loss of information (e.g. certain warm-up times before receiving signals)
- costly post-processing of the GPS data, i.e. trip end, trip purpose and street address detection
- still relatively high equipment costs
- unlike most ordinary paper-pen surveys, no motives for travel are queried.

Generally, all studies have in common that the GPS tracking is only one part of the overall survey structure. GPS monitored and therefore passively collected travel behaviour data needs to be framed by socio-demographic attributes of the travellers but especially by further information on trip purposes and the size of the company. Some methodological approaches to detect trip purposes – based on recent studies – are described in section 4. Future GPS applications will be possibly combined with more interactive techniques (such as the usage of PDAs to state trip purposes and car occupancies) to obtain a broader picture of the observed trips or travel patterns (Doherty, Nöel, Lee Gosselin, Sirois, Ueno and Theberge, 1999). The level of user interaction is believed to be an important issue for the development of future survey design incorporating GPS data collection elements.

3. **The Rätt Fart Borlänge GPS Data Set**

Based on contacts with transport psychologists from the universities of Dalarna and Uppsala (Sweden), the Institute of Transport Planning, Traffic, Highway and Railway Engineering (IVT) obtained access to the GPS data set *Rätt Fart* which promises to match nicely the research directions of Mobidrive.

3.1 **Background**

Rätt Fart, based in the town of Borlänge 200 km northwest of Stockholm (Right Speed; see http://www.rattfart.com), is one of the sub-projects of the Swedish National Road Administration initiated approach *Intelligent Speed Adaptation* (ISA). ISA aims to influence car and truck drivers’ behaviour by in-vehicle information (see Vägverket, 2000a).
The ISA sub-projects are designed to analyse the responses of drivers, ways of integrating interactive technologies into vehicles and the effect of intelligent speed adaptation systems on road safety and the environment. The technical set-up of ISA was to equip about 5000 vehicles (private and commercial) Sweden-wide with GPS receivers and digital mapping for three main purposes:

- to provide the drivers with information about speed limits during the drive (via a special display)
- to actively give support to the drivers when exceeding speed limits in urban areas (i.e. by warning signals or devices which prevent the drivers from further acceleration)
- to reinforce self-commitments made by public transport and goods delivery companies to respect speed limits by monitoring the drivers’ behaviour.

Rätt Fart in Borlänge had its foci on information provision and quality assurance. The study was conducted from 1999 to 2001. The project team managed to equip approximately 400 private and commercial cars with ISA systems over the period of up to 2 years. All speed limits of the Borlänge region road network links were assigned to a digital map which was integrated in the vehicles’ equipment. Through the GPS device the drivers were informed about the speed limits of their actual position and were eventually warned if the speed limit was exceeded (see Figure 3). The drive’s characteristics such as speed, acceleration, actual time, position etc. was stored internally for analysis in logs every second respectively every tenth second depending to the road link used (see below). The GPS receiver itself did not transmit any signal of its own which prevented external sources to get access to the car’s location or other information. Hence, the data protection regulations were met.
3.2 Data availability and further data acquisition

The data logs from the vehicles as well as supplementary data concerning the acceptance of the ISA device are analysed right now in terms of traffic safety by researchers of the two Swedish universities. Some of the test drivers accepted to participate in a post-in-depth interview which yielded further information on the socio-demographics of the test drivers and their fixed commitments (obligatory and voluntary).

For the travel behaviour analysis purposes, the data logs of the first 14 months of monitoring and further vehicle respectively driver information were made available to the IVT in 2001 – a second part covering further 12 months of GPS data is expected for the summer 2002. Besides, additional useful data sources could be explored to support the identification of further trip attributes such as trip purposes and route choice. Table 2 gives an overview over the data available.

The core of the GPS mobility data base are the trip and the logs-files stored in a SQL server system. The size of the files is enormous due to the richness of the data and the prolonged monitoring period, with e.g. 20 GB for the logs file. Consequently, the extraction, processing and analysis of relevant information will be determined by the hardware performance.

The movement file provides accurate trip-specific information such as times, positions speeds and path choices (see below). The area for detailed monitoring was limited, though, to the town of Borlänge plus some surrounding region – an area with a radius of about 20 km around the town centre of Borlänge.
According to the project objectives, the Rätt Fart test drivers were recruited by traffic user groups, type of owner and (related) vehicle types to reach a certain composition for the total sample. The traffic groups were ‘private cars’, ‘delivery / commercial vehicles’ and ‘taxis’ and ‘public transport vehicles’ with the vehicle types cars, truck and busses and the ownership status ‘private owned’ and ‘company owned’. The log intensity differed by the installed equipment: Most of the vehicles were provided with computers which logged the drive’s detail every second, a smaller number of vehicles were equipped with devices which differentiate by the fact of whether the applicable speed limit was kept or broken. In the latter case the log intensity was 1 second, in case of keeping the speed limit the log intensity decreased to 10-second intervals.
**Tab. 2:** Available data: GPS data and additionally collected information

<table>
<thead>
<tr>
<th>Category</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trips (GPS)</td>
<td>Vehicle number, trip number, start time of trip, end time of trip, total length of trip, total duration of trip, total duration of trip within monitoring area</td>
</tr>
<tr>
<td>(Source: Rätt Fart project team)</td>
<td></td>
</tr>
<tr>
<td>Logs</td>
<td>Vehicle number, trip number, sequential number of log, log date and time, used road network link at time of logging, direction of travelling on link, exact position on vehicle (given as sector of the used link), speed at time of logging, speed limit at position of log, acceleration at time of log</td>
</tr>
<tr>
<td>(Source: Rätt Fart project team)</td>
<td></td>
</tr>
<tr>
<td>Vehicle data</td>
<td>Vehicle number, type of vehicle, traffic group type of owner, type of ISA device, ISA device activation date</td>
</tr>
<tr>
<td>(Source: Rätt Fart project team)</td>
<td></td>
</tr>
<tr>
<td>GIS Road network</td>
<td>Link number, length of link, permanent speed limits on link (for both directions), temporary speed limit(s) on link and details, log intensity (every second/every tenth second), type of warning at speed limit exceeding, permissible turnings, direction of flows</td>
</tr>
<tr>
<td>(Source: City of Borlänge / Swedish National Road Administration)</td>
<td></td>
</tr>
<tr>
<td>Other data</td>
<td>Selected socio-demographic characteristics of the test drivers (Type of driver (actual test driver or additional driver of the car), sex, age, home address, kilometrage of the last year, number of cars in household, main obligatory (work or education) and leisure activities: Type, time of day, weekly frequency and locality, main shopping locations, share of non-car travel (public transport and bicycle), access to bicycle, most used mode of getting to: Work/school, groceries, leisure activity locations, ownership of season ticket / public transport discount card, places the test drivers would never go by car)</td>
</tr>
<tr>
<td>(Different sources)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Digital land use data in GIS; scale of base maps: 1:1000: Selection of land use layers, including residential and commercial areas, leisure and cultural sites, nature and agricultural areas</td>
</tr>
<tr>
<td></td>
<td>Swedish national travel survey data 1994-2001 (RIKS RVU &amp; RES)</td>
</tr>
<tr>
<td></td>
<td>Comprehensive national one-day mobility data</td>
</tr>
</tbody>
</table>

**Tab. 3:** Borlänge GPS data: Availability (first part)

<p>| Data detail | Availability |
|-------------|--------------|-------------|
|             |              |             |</p>
<table>
<thead>
<tr>
<th>Study period</th>
<th>22 June 2000 - 17 August 2001</th>
</tr>
</thead>
</table>
| Vehicle types and numbers | Private cars: 261  
Commercial cars: 17  
Taxis: 32 |
| Movements (trips) | 161.029 all vehicles / 103.105 private cars only  
Minimum observation period for private cars: 15 days of monitoring / 5 days  
Maximum: 243 days of monitoring / 223 days actually driven  
Mean: 146 days of monitoring / 109 days actually driven |

4. DATA PROCESSING ISSUES

It is clear that car-based tracking has several problems which need to be addressed: the identification of the driver, the omission of non-car-based travel or the missing knowledge about the reasons for travelling (i.e. trip or activity purposes). Besides, the availability of individual socio-demographic data is limited which has strong implications for data processing and later analysis.

One of the core questions which is often raised when discussing the use of automatically collected GPS data for travel behaviour analysis is whether it is possible to post-process the data to obtain missing trip attributes such as trip purposes (see Wolf et al., 2001). The data-processing and analysis framework for the Borlänge GPS data includes those imputation steps, but also an approach to reconstruct a picture of the daily activity chains with non-car movements. The challenge in this particular research setting is the total absence of comparative match-travel-data. As the original Borlänge Rätt Fart study is uniquely designed for transport safety purposes, it was not possible to ask the test drivers to keep ordinary travel diaries parallely nor to collect comprehensive socio-demographic characteristics – as in most other studies mentioned above. Hence, the quality of all data imputation results mainly will rely on revealing the inherent structure of the longitudinal data. As the period of monitoring for any of the test drivers by far exceeds the usual observation period of travel surveys, it is believed that this goal will be achieved sufficiently. Mobidrive showed that daily life is still dominated by fixed temporal and especially spatial structures of travel demand.

4.1 Post-data-processing steps

In the following, the imputation steps leading to a more complete picture of daily travel of the test drivers are described in detail. As the analysis
direction is towards person and household related mobility patterns, all strategies involve only private car travel.

Due to the mentioned data limitations, the final mobility data set will likely not offer the full range of variables which activity-based travel analysis typically works with. Nevertheless, it needs to define minimum requirements for the quality of the final data set in order to set a framework for the actual analysis steps. The planned structure of the data set and its quality of the data is described in Table 4.

**Tab. 4:** Structure and expected *quality* of the Final Borlänge data set

<table>
<thead>
<tr>
<th>Expected level of availability and precision</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exact values / high accuracy</td>
<td>Route choice, departure and arrival times, travel times, travel distances, speeds</td>
</tr>
<tr>
<td>Acceptable precision</td>
<td>Trip purposes, activity locations (addresses)</td>
</tr>
<tr>
<td>Approximations / assumptions</td>
<td>Overall mobility patterns / trip chains including non-car travel</td>
</tr>
<tr>
<td>Missing</td>
<td>Activity expenditures, size of company</td>
</tr>
</tbody>
</table>

**4.1.1 Identifying the driver**

The system used in Borlänge did not require the driver to identify himself/herself. With only few exceptions, the private Borlänge vehicles were used by one driver in general, as only households with the same number of drivers and vehicles were recruited. In any case, the system will provide a detailed record of the driving style of each driver through measurements of average speed and acceleration behaviour. In the case, that the analysis of the travel behaviour were to suggest different drivers, these low-level behaviour could be used to construct tools to discriminate the drivers. This includes the regularities in daily life travel and especially the time budgets of the test drivers which show a high stability over the month-long monitoring period.

---

2 Note: The Atlanta-based SMARTRAQ/Drive Atlanta study will also not be able to provide positive driver identification; again vehicles will be recruited into the study based on a household reported one-person to one-vehicle relationship; i.e., shared vehicles will not be included in the study.
4.1.2 Identifying unique origins and destinations of travel

Trip starts and ends are predefined in the Rätt Fart approach by switching on respectively turning off the car (which includes the ISA device). This automatically starts or interrupts the transmission of the drive’s characteristics for data storage. All positions of the drive including the start and end locations (i.e. first and last positions) are given by

- the network link (road) used
- the exact position on the link (measured in meters from one the start node of link)
- the direction of travelling on the link.

This information is re-calculated into real-world coordinates in order to better represent locations in a GIS. Figure 4 shows an example for trip end positions of one vehicle: The bold numbers represent the network link ID, the small dots and numbers give the last position of various drives.

It can be seen that the final positions of the drive vary by e.g. the choice of certain parking spaces varying over the period of monitoring. Although one can assume that due to the proximity of positions only one activity location is targeted by the test driver (such as “home”), its identification requires clustering techniques. A suitable approach is to set a tolerance distance or zones in which different final positions are per definition considered as only one destination. For the different activity purposes, it seems reasonable to vary the value of the tolerance, with e.g. smaller tolerances for ‘home’ and ‘work’ and bigger ones for e.g. ‘shopping’ with probably greater flexibility of destination and parking space choice.
Due to the fact that data logs are only available for movements which are made on the official road network, private access roads to the premises and parking spaces have to be considered separately. As one can see from Figure 5 which shows the last trips of the day of one particular vehicle over the monitoring period, the data does unfortunately not cover the access to the premises situated further away from the road. For all purposes apart from home and work where the addresses of the locations are known from the framework data, the exact destination will be difficult to identify in those cases. In those cases, it is necessary to make locality assumptions based on land-use - trip end matches (see below).
One further difficulty which has to be faced is the problem of systematic inaccurate transmission at the beginning of the drives. In contrary to the reliable information on the trip ends with arrival time and exact position, there is a systematic delay in data transmission at the beginning of the trips for technical reasons. An initial investigation of the logs of only six private test drivers showed that in about 50% of the drives the origin links do not match with the final link of the preceding trip. As – with high probability – it can be assumed that the trip end position of the preceding trip is identical with the trip start position of the actual one, for most of the analysis work the start positions can be simply replaced. For the start times of the trips, only negligible differences have to be considered according to the small spatial differences in the locations.

Nevertheless, the route choice analysis and modelling will be concerned since a first short period of monitoring the route choice is missing. As three quarter of these differences do not exceed 200 meters, the missing link(s) may be found without big difficulties in most of the cases.

### 4.1.3 Detecting additional trip ends

As already noted, the trip ends are predefined in the Rätt Fart approach by powering on the ISA device. The question nevertheless is, if there are short stops to regard which include an activity such as dropping off passengers, buying something quickly or getting cash at an automated teller machine. Considering short stops of only about one to five minutes, it is difficult to distinguish between stops due to the experienced traffic condition (e.g. congestion, waiting at traffic lights) or transmission gaps and stops for performing an activity. Whereas in common paper based travel diary data, the trip end is explicitly defined by the travellers’ specification of arrival time and destination, the automatic collection of movements and stops eventually yields ambiguous results.

A second case for which tests for additional trip ends is necessary, is the delay or the obstruction of the signal transmission. This eventually indicates the execution of an activity in but especially outside the car while the ISA device is still turned on and transmitting logs to the computer. As mentioned earlier, the computer stored signals received from the GPS equipped car every second resp. every tenth second, depending on the particular link used at the log time. At some points of time, though, these periodic transmission failed which lead to larger intervals between the data storage. The reason for the interruption or the delay in the transmission could be manifold, such as the general but temporal failure of the sender / receiver...
system or the obstruction of the signal due to physical barriers such as high buildings.

A starting point for the detection is the investigation of the short durations during which the vehicle does not move – or in other words, during which zero speed is observed. Fortunately, this seems to be only seldom the case. Again, for the six initially selected vehicles with a total of about 700,000 logs and more than 2,000 trips, the number of problematic cases such as lost signals or suspiciously long stops during the course of a drive is infinitesimal. If this is true for the majority of the vehicles, the question of unobserved stops needs not be prioritised given the large amount of data available.

**Tab. 5:** Transmission problems of six arbitrarily selected vehicles; \( N = 693,685 \) logs

<table>
<thead>
<tr>
<th>Problem</th>
<th>Cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signal “lost”</td>
<td></td>
</tr>
<tr>
<td>more than 30 seconds</td>
<td>1018</td>
</tr>
<tr>
<td>more than 60 seconds</td>
<td>566</td>
</tr>
<tr>
<td>more than 120 seconds</td>
<td>420</td>
</tr>
<tr>
<td>Periods of “zero speed”</td>
<td></td>
</tr>
<tr>
<td>more than 30 seconds</td>
<td>86</td>
</tr>
<tr>
<td>more than 60 seconds</td>
<td>58</td>
</tr>
<tr>
<td>more than 120 seconds</td>
<td>18</td>
</tr>
</tbody>
</table>

Nevertheless, there is already expertise to handle the eventual problem with intervals exceeding the usual transmission intensity (see Wolf, 2000; Pearson, 2001; Wolf *et al.*, 2001). A possible trip end detection rule for the Borlänge data set could be as follows:

1. Intervals between logs, i.e. potential stops shorter than 120 seconds are not considered as trip ends.

2. Stops equal or longer than 120 seconds are considered as trip ends (with a following activity) if interval is not at the beginning or the end of the particular trip. In these cases it may be assumed that the stop is caused by preparing the car for parking / starting from a parking places.
### 4.1.4 Identification of trip purposes

The identification of trip purposes will be the core post-processing issue in the GPS data adaptation experiment. The planned strategy is based on three pillars:

- The knowledge about the daily life’s structures and the regularities of travel
- Comparing trip ends with land use and address information
- Exploiting the post-in-depth interviews on occupation and fixed commitments

The longitudinal analysis in Mobdrive showed that for most of the travellers, the daily travel is structured by clear regularities which yield credible assumptions on time usage and activity performance. In many cases, departure, arrival and travel times as well as activity durations indicate purposes of travel – at least with great probability.

Figure 6 shows an example of the long-term regularity in departure times and activity performance of one test driver in Borlänge. The figures show selected characteristics of GPS track data for one arbitrary chosen vehicle over time. On the left, one can see the departure times of the first movements on weekdays, whereas the right give the sum of daily activity durations at the four most often frequented (non-home) destinations. The different shadings in the right figure indicate the different (activity) locations which were found by clustering final positions (trip ends) according to the concept described above. Together with the knowledge about the average working hours of the test drivers reported in the post-in-depth interview and his/her fixed commitments such as hobbies, it is possible to make relatively clear assumptions about the trip purposes performed at the given locations. In this case, it is reasonable to identify the working hours and durations of the particular test driver. As a majority of all private test drivers in Borlänge are full-time workers with a presumably high regularity in daily life travel, the data processing will use the knowledge about the stability of the work activity performance as starting point for detecting comprehensive activity patterns.
Exploring the potentials of automatically collected GPS data for travel behaviour analysis

One other obvious approach to identify trip purposes is to match trip end destinations with the underlying land use. There are two drawbacks which restrict the success of this data processing step: First, as in many other municipalities worldwide, the available accuracy of the digital land-use and property data base for Borlänge is limited. In addition to that, the Borlänge GPS data is only accurate for movements on the main road network and may not be matched easily with the land use GIS layers. In most of the cases as in Figure 7, there is no direct matching possible (as for example planned to perform in SMARTRAQ study; Wolf et al., 2001). The drives’ ends are close to lots and buildings but do not spatially match with them at one go. In order to obtain indications about trip purposes, it requires to again set tolerance distances to find suitable assignments of trip ends to particular lots or buildings.

Fig. 6: Stability in departure times and activity performance at four activity locations
– Example from the Borlänge data base; weekdays only

Notes: The order of the colours does not represent any temporal order of activities on the days shown. It rather shows the hierarchy in the overall numbers of activities performed at the particular locations. – The week 16-22 April in 2001 was the week after Easter.
4.1.5 Constructing “full” activity patterns

One of the challenges of the Borlänge experiment is the revealment of full activity patterns exceeding the car trips given by the analysis of the Borlänge GPS data set. The idea is to imbed the car based travel into a fictitious but close-to-reality picture of non-motorised travel in order to perform common travel behaviour analysis known from most studies performed in the framework of activity-based research. This step will be centred on two strategies: (1) to use the knowledge of the Mobidrive survey about frequent drivers patterns of mobility and (2) to impute those full patterns by using one-day national travel survey data and its information about average travel times, distances, activity chain frequencies etc.

In the Mobidrive data set for example, regular car drivers visited only about 15% of their destinations never by car, mostly by walking (2/3 of all non-car trips). Based on the Mobidrive data it is reasonable to expect that most of these destinations will be clustered within a 1 km radius around the home, the work place and other locations reached by the motorised modes. This implies that our estimates of the size of the activity spaces will be affected only modestly by the omission of these trips.

Besides, for Borlänge it is possible to test the importance of the omission of non-car-based trips using the Swedish National Travel Survey (RVU). The Swedish RVU provides a rich source of information about travel behaviour of Swedish residents based on a one day recall diary. Using the sample as a whole, as well as matched samples (private drivers) it is possible to assess the number of locations omitted, their type and further attributes of travel.
4.2 Automating the data processing?

Due to the size of the data set, the data imputation strategy will only succeed if it is possible to automate the data processing steps to a large extent. So far, there is only little experience with automating this work as in most recent studies the derivation of trip purposes or trip end addresses was done manually and in a piecemeal way. Considering the final size of the data set with more than 200,000 private trips over two years, it is illusionary to get into detail for each and every single trip. Hence, an automating strategy will have to focus on the following steps:

- Initially reveal representative daily mobility profiles for the individuals in order to make out activity purposes and trip destinations by the regularity of daily life as well as to obtain high certainty about the driver identification
- Identifying the main activity locations of the Rätt Fart test drivers by clustering trip end positions
- Assigning trip purposes by combining observed activity durations, assumptions on average duration and information of the post-in-depth interviews (mainly on obligations and fixed commitments)
- Detecting non-relevant short drives and technical bugs in the log data.

5. Conclusions and outlook

The description of the data processing steps has already shown that the usage of the Borlänge GPS data set for travel behaviour analysis is connected with a range of inherent data difficulties. Summarising the challenges, it can be noted that

- there exists only limited digital land-use information and data on the socio-demographics of the test drivers is restricted
- the selective inaccuracy of the available GPS data requires additional data processing work
- the Borlänge experiment cannot rely on supplemental follow-up answers of the test drivers in order to clear ambiguous findings and mismatches.

Why is it nevertheless worthwhile to undertake the efforts with the data processing and the reconstruction of the full activity patterns of the test drivers? On the one hand, the project adds to the expertise of researchers
on the implementation of passive monitoring for travel behaviour surveys. The perspective for the future data collection and availability may be described by less burden for the respondents, higher accuracy of several data elements and the addition of new travel data attributes such as route choice and speed behaviour. Furthermore, there is added value for travel behaviour research which exceeds the methodological component of survey design and implementation: The Borlänge data set will provide us with unique findings on the variability of individual travel behaviour over prolonged periods. So far, travel behaviour and time use research has never had the opportunity to track individuals longer than some weeks in order to find out about the systematic and the spontaneous part of behaviour. Whereas it may be argued that everything is known about daily life’s travel, e.g. the average daily travel time, mode choice, average distances per day etc., this is only true for one putatively representative day of the travellers’ overall mobility patterns. Alike Mobidrive, Borlänge will reveal a big portion of regularity in daily life travel which supports the common one-day survey approach. But it will also show the complexity of daily life with the interaction between periodicity and variability which eventually has contributed to the difficulties which transport planning has to face and the partial failure of several transport policy strategies in the past.

**An outlook on future analysis directions: Route choice**

One example for future analysis approaches based on the Borlänge data is route choice analysis and modelling. This model direction is integrative part of most transport models which have been the key decision making support tools for transport planning since the 1960s (Bovy and Stern, 1990 for an introductory overview of route choice model techniques).

By its longitudinal character, the Borlänge data allows to analyse the actual path choice of the test drivers, variability and stability issues and comparisons between drivers taking identical trips (similar analysis was made for Lexington by Jan, Horowitz and Peng, 2000). In addition, the investigation of the long-term Borlänge data will reveal the dynamics as well as the seasonalities in path choice given different departure times, seasons, working and holiday periods, the occurrence of special events etc.

The project partners’ future research emphasis is to test new discrete choice approaches which represent the behavioural reality of spatial decision making better. The application of discrete choice methodology to route choice analysis is a complicated process, mainly due to inconsistency with classical modelling assumptions and data unavailability (see Ben-Akiva and Bierlaire, 1999). The Borlänge data set is a great opportunity to deal with the
second problem. The first problem can be summarised by the following questions:

- Is it possible to identify the decision-maker’s choice set which includes only the actually considered alternatives for route choice? The universal choice set, containing all potentially chosen alternatives, is very large for realistic networks, and behaviourally unrealistic.

- How do we identify topological overlap or subjective aggregation of choice alternatives which are assumed to be independent? And: Which modelling tools, model estimation algorithms and large samples of relevant data are suitable for revealing the high-level of correlation among different alternatives?

One possible framework to be tested is the introduction of an intermediate destination approach and sub-paths models (Figure 5; see Ben-Akiva and Bierlaire, 1999; Antille, 2002). Intermediate destinations which induce obligatory sub-paths may be of different types: locations at which short activities such as dropping down passengers are performed (e.g. kindergarten, shops etc.), certain landmarks in the cityscape which act as points of spatial orientation, important junctions such as motorway exits etc. The introduction of an intermediate destinations / sub-path concept to route choice modelling would expand the utility maximisation led route choice theory by a stronger behavioural basis via spatial navigation.

![Illustration of a possible sub-path model](image)

**Fig. 8:** Illustration of a possible sub-path model

Initial modelling experiments with a small sub-sample of the Borlänge data set and by integrating a mathematical representation of the intermediate destination / sub-path phenomenon showed that the model goodness of fit improved significantly (Antille, 2002, 86ff.). Without drawing final
conclusions this shows both, a better way of representing of behavioural reality and the advantage of accurate track data for route choice modelling.

**REFERENCES**


Vägverket (2000a) ISA Intelligent Speed Adaptation, Vägverket, unpublished, Vägverket, Borlänge.


Wolf, J., R. Guensler and W. Bachman (2001) Elimination of the travel diary: An experiment to derive trip purpose from GPS travel data,
Presentation at TRB 80th Annual Meeting of Transportation Research Board, Washington.