Title: Comparison of conventional versus GPS-based data collection in the French National Travel Survey

Philippe Marchal (corresponding author)
Université Paris-Est, IFSTTAR, DEST, F-93166 Noisy-le-Grand, France
French institute of science and technology for transport, development and networks (IFSTTAR)
Department of Transport Economics and Sociology (DEST)
Tel: +33 (0)1 45 92 55 70
E-mail: philippe.marchal@ifsttar.fr

Thao Pham
Université Paris-Est, IFSTTAR, DEST, F-93166 Noisy-le-Grand, France
French institute of science and technology for transport, development and networks (IFSTTAR)
Department of Transport Economics and Sociology (DEST)
Tel: +33 (0)1 45 92 55 71
E-mail: thi-huong-thao.pham@ifsttar.fr

Abstract:

Mobility data collection methods based on new technologies have evolved in the past few decades, shifting from limited experiments to large-scale travel surveys. GPS-based data collection methods have become particularly popular in travel behavior research, mainly because of the worldwide coverage and the accuracy of the GPS system. Comparing with the traditional survey methods such as PAPI (Paper and Pencil Interview), CAPI (Computer-Assisted Personal Interview) or CATI (Computer Assisted Telephone Interview), the advantages are obvious:

- Firstly, GPS provides researchers and those conducting transport studies more specific information in terms of spatial and temporal resolution;
- Secondly, GPS limits forgotten trips, which is a known limitation in data collection with conventional surveys;
- Thirdly, the relatively low burden for the respondent allows substantially extended survey duration: at least one week with GPS, compared to one or two days with a conventional questionnaire.

But raw data are not directly usable:

- GPS traces are not segmented;
- There are missing segments;
- There is information neither on transport means nor on trip purposes.
Thus, for post-processing these data, more or less sophisticated software packages have to be elaborated depending on the accuracy needed by the users (e.g. much more spatial accuracy for the assessment of advertising by posters than for other users of travel survey results).

The introduction of a large scale sub-sample of follow-up by GPS in a conventional National Travel Survey prepares the transition toward more use of cheaper new technologies, which are less burdensome for the respondent and allow surveying on a longer period.

Our experimental design takes advantage of the general characteristics of a long survey (with two face-to-face interviews) to get through the CAPI-GPS additional information on the reliability of the device and on more detailed characteristics of a few trips, which will be useful for the calibration of the post-processing software.

1. General scheme of the GPS component in the French National Travel Survey (NTS)

For this first experimental attempt, it has been authorized by the French National Commission for Data protection and the Liberties (CNIL) under the condition, that the GPS component should concern only volunteers. When a respondent agrees with the GPS option:

- At the first face-to-face interview, the interviewer gives the "GPS Pack" to the respondent (older than 17) and explains how to use the equipment;
- Between the two visits, the respondent will travel and the unit will record trips;
- At the second face-to-face interview:
  - The respondent gives back the "GPS Pack" to the interviewer;
  - Immediately the interviewer downloads the GPS data on his laptop computer using a Bluetooth transfer, for a brief additional interview;
  - The interviewer checks the GPS unit, reloads it; the equipment is ready for a new interview.

1.1. The GPS units selected

For road safety reasons and to avoid an influence of GPS on travel behavior, it is a passive monitoring tool: the respondent has no graphical interface. It has only one button (on/off). The respondent has the possibility to skip some trips, if desired for confidentiality reasons.

No data are transmitted on a real-time basis: the device is only a datalogger. Data are downloaded to the interviewer's laptop during the second visit, and deleted inside the unit.

Two types of GPS unit are used for the survey: 100 "normal", with 15-17 hours autonomy and 70 with 10 hours autonomy, but modified, with movement detection and blinking mode. The movement detection makes it possible to automatically shut down the unit when the respondent forgot to do it, for example at the office. The blinking mode allows the recording of “partial traces” (10 min segments) in case of long distance trips, increasing the GPS “useful autonomy”.

1.2. Equipment and documents for respondents and interviewers

The "GPS pack" given to the respondent contains:

- The GPS unit;
- An A/C adapter;
- A one page user manual.
The interviewers are given the following equipment and additional documents:

- SD Card Memory placed in the laptop card reader from the beginning to the end of the survey: they are used for both software installation verification and for backup purpose. In case of problem detected with the GPS unit, the SD Card sent by the interviewer along with the GPS unit allows a quick diagnostic.
- A Bluetooth USB Adapter inserted before downloading the data (at each interview);
- A receipt for the GPS device;
- A 2 pages manual for the interviewer.

1.3. The CAPI-GPS design

If there are days without GPS record, the first step is to understand the reasons why. For each day without any record, the following question is asked:

"During this day, we have no data recorded, could you tell why?
- I have forgotten the GPS unit
- Unit off
- No trip
- Problem with the GPS unit (battery)

Based on the analysis of the downloaded GPS data, a day is selected, avoiding asking questions on days already described in the traditional CAPI interview (the day before and the last Saturday or Sunday). The GPS records are split into trips. First, memory is prompted with trip characteristics (day of the week, date, departure time, duration, municipalities of origin and of destination). Then the following questions are asked:

- "For what purpose?"
- "Number of accompanying persons?"
- "Means of transport used"

2. Some limits… and solutions: the importance of the post-processing tools

There are clear technical limits, with signal loss in tunnels and subways. It works better when it is positioned near the window in public transports and starts after a few minutes when getting out of a building. But we have noticed substantial improvements with new generation devices.

There are also human limits:

- The respondent may forget it at home or elsewhere, or forget to put it on;
- The battery might turn uncharged (it has to be plugged every night).

For checking and analyzing data, we can combine three measurement tools:

- The conventional questionnaire;
- GPS traces; and
- CAPI-GPS results (purpose, transport means, accompanying persons).

Based on the comparison between data produced by each of these instruments, methods for imputations (of purpose and mode) from the GPS traces only have been built and are currently improved.
A challenge in the post-processing of data for a passive GPS-based travel survey is to develop a series of methods to automatically restore continuous sequences, both in space and in time. This means that trips and activities occurring during the survey time should be programmatically identified in chronological order and should respect conventional definitions. A series of methods has been developed and applied:

1) Filtering invalid data,
2) Cutting the data chains into “moving” and “stopping” segments,
3) Estimating missing data,
4) Assuming the mode choice,
5) Grouping stop points into visited locations,
6) Identifying the type of visited points,
7) Map-matching,
8) Connecting segments to trips.

In order to use the maximum available information and guarantee the quality of the results, some feedback between the different methods is implemented within the algorithm.

Post-processing mobility data collected by GPS without any additional information yields:
• a good picture of the sequence of traces, depending on the frequency of measurement;
• a satisfactory imputation of transport means, especially for walking trips and road and rail trips;
• greater difficulty in identifying bicycle/car/bus among road modes;
• no information about drivers/passengers in vehicles, which would be important for the analysis of the occupancy rate (a major factor in fuel efficiency per passenger x km);
• poor information on purposes/activities, derived only from the identification of “Points of Interest” (POI) and from the frequency and duration of visits to these points.