Combining a lightened conventional questionnaire with a person-based GPS component for longer survey period and response burden reduction

Philippe Marchal\textsuperscript{1}, Jean-Loup Madre\textsuperscript{2}, Thanh-Tu Nguyen\textsuperscript{3}
\textsuperscript{1}IFSTTAR-DEST, e-mail: philippe.marchal@ifsttar.fr
\textsuperscript{2}IFSTTAR-DEST, e-mail: jean-loup.madre@ifsttar.fr
\textsuperscript{3}IFSTTAR-DEST, e-mail: than-tu.nguyen@ifsttar.fr

Abstract

The utilisation of new technologies for mobility surveys may reduce respondent burden and the survey cost which should have first-class impacts on data accuracy and quality. Moreover, the relatively low burden for the respondent allows substantially extended survey duration: at least one week with GPS, compared to one day with the conventional questionnaire. Anyway a few questions are needed for collecting the socio-demographic characteristics of the respondent. We have taken the opportunity of the French National Travel Survey for a first nationwide experience with embedding a “GPS package” in a traditional survey.

Keywords: Travel Survey, GPS, Post-Processing

1- Introduction

Mobile communication and positioning technologies demonstrate great potential as survey instruments for tracking individual mobility and travel behaviour by enabling to conduct surveys for longer periods and providing more accurate data on the spatial and temporal frameworks of travels. Beside these improvements, the utilisation of new technologies may reduce respondent burden and the survey cost which should have first-class impacts on data accuracy and quality. Moreover, the relatively low burden for the respondent allows substantially extended survey duration: at least one week with GPS, compared to one day with the conventional questionnaire. Anyway a few questions are needed for collecting the socio-demographic characteristics of the respondent. Could it be also used, without extending it too much, to improve the post-processing of GPS traces, especially for the imputation of trip purpose and transport mode?
We have taken the opportunity of the French National Travel Survey for a first nationwide experience with embedding such a “GPS package” in a traditional survey, with a sub-sample of approximately 600 voluntary interviewees.
2- The French National Travel Survey (FNTS)

Once per decade, the Ministry of Transport and the National Institute of Statistics use to conducting a National Household Travel Survey with the scientific support of INRETS. It is the data source providing the most transverse and consistent overview of mobility, whatever the modes and transport situations of people living in France may be. The objective of these surveys is the description of short and long distance trips made by households living in France, as well as their access to and use of public and private transport means.

The sample size for daily mobility is 18,632 respondents. Data collection is spread over six waves covering 12 months, in order to neutralize the seasonal variations which affect mobility. Taking into account the overall length of the interviews (approximately a total duration of 115 minutes), the data has been collected in two visits, which makes it possible to distribute the GPS receiver at the first visit, and to collect it at the second one. For road safety reasons and to avoid an influence of GPS on travel behaviour, 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. In the FNTS, data are recorded every 10 seconds. No data are transmitted in real-time: the device is only a datalogger. GPS data are downloaded to the interviewer's laptop during the second visit, and deleted inside the GPS unit. A prompted recall interview (CAPI-GPS) is then conducted: it contains questions about days with no trace and about purpose, transport means and accompanying persons for trips made during one day taken at random.

3- Elaborating a post-processing software

In Australia, Canada, Switzerland, the Netherlands, etc. several authors have addressed the difficulties encountered in analyzing GPS traces.

Basically, all approaches contain individual modules accounting for:
- Data filtering,
- Detection of trips and activities,
- Mode stage determination,
- Mode identification,
- Map-matching.

Based on the experience of our colleagues, we have designed our own procedures, adapted to the dataset recently collected in France. This software yields:
- A good picture of the sequence of traces, depending on the frequency of measurement (every second/10s, the optimum seems to be 3 to 5 seconds),
- A satisfactory imputation of transport means, especially for walking trips and road and rail trips,
- But greater difficulty in identifying two wheels/car/bus among road modes,
- No information about drivers/passengers, which would be important for the analysis of rate of occupancy; comparing the traces for all members of the household doesn't seem enough, since 30% of passenger trips are made using a vehicle which doesn't belong to the household;
- A poor information on purposes/activities derived only from:
4- What small amount of additional data could be collected for improving data processing?

4.1 Days with no record
From the 285 days taken at random for the CAPI-GPS, the reasons given are:

- Stayed at home: 43%
- GPS forgotten: 22%
- Made trips as usual: 17%
- Don’t remember: 18%
- Total: 100%
without clear correlations with respondent's characteristics.

4.2 For trip purpose
The address of the main activities can help, not only the location of the residence and work/study place, which can be easily detected through the analysis of the frequency and time/duration they are visited. But we have to take care of confidentiality issues.

4.3 For trip mode
- Driving license status (for car/motorcycle);
- Driving habits (always the driver, occasionally, never) e.g. only 10% of car trips made by persons driving frequently are made as passengers;
- Ability to ride a bicycle and habit of doing so, e.g. 78% of trips by bike are made by persons using frequently this mode;
- Season ticket status for Public Transport, e.g. only 2% of trips made by non holders are by PT.
### Logit Model for Modal Split

<table>
<thead>
<tr>
<th>Variables</th>
<th>Car Driver</th>
<th>Car Passenger</th>
<th>Walk</th>
<th>Public Transport</th>
<th>Bicycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>+</td>
<td>**</td>
<td>.</td>
<td>.</td>
<td>+++</td>
</tr>
<tr>
<td>Holds a Driving License</td>
<td>+ + +</td>
<td>**</td>
<td>*</td>
<td>**</td>
<td>.</td>
</tr>
<tr>
<td>Drives Frequently</td>
<td>+ +</td>
<td>*</td>
<td>*</td>
<td>**</td>
<td>.</td>
</tr>
<tr>
<td>Age&gt;15</td>
<td>.</td>
<td>**</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Number of Trips in the same Day</td>
<td>.</td>
<td>+</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Distance Travelled on the same Day</td>
<td>.</td>
<td>.</td>
<td>***</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>Holds a Season Ticket</td>
<td>*</td>
<td>.</td>
<td>.</td>
<td>+++</td>
<td>.</td>
</tr>
<tr>
<td>Uses a Bike or Motorcycle</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>+</td>
<td>+++</td>
</tr>
<tr>
<td>Live in a Municipality with a high jobs Density</td>
<td>**</td>
<td>+</td>
<td>+</td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

Sources: INSEE - SOeS– ENTD 2007-2008

(·): Not significantly different from 0 or <0.2 in absolute value
(+): significant and positive
(*): significant and negative

**Remark:** Car ownership, "enjoy driving" and obesity have no significant influence. Those living in:
- conurbations make more trips by bike (two wheels?)
- high population density communities make less trips as car passengers.

From these questions, probabilities could be derived for the use of each transport mean, when the algorithm gives an uncertain answer. In many cases, this probability can be almost 0 (e.g. driving without a license, unable to ride a bicycle, etc.).

### 5- Conclusions

The ideal solution is, of course, to let the respondent edit his own traces on the internet (e.g. Harry Timmermans on the Dutch experience; the validation sub-sample in Cincinnati), but the ability to use the internet still introduces a strong selection bias for a large scale survey.

Thus, a short set of easy questions could considerably help for post-processing traces collected by GPS:
- Reasons for days without traces;
- Addresses for determining activity and trip purpose;
- Season ticket or driving license status, habits to drive a car or a bicycle and zone of residence for improving the imputation of trip mode by probabilistic rules.

The following methodologies should be discussed:
- When (before/during/after the survey period) and
- How (mailback, phone call) to collect this brief information?

References

5. Flavigny P.-O., Hubert J.-P., Madre J.-L., Analyse de trafic routier observé par GPS et comparaison avec d'autres sources statistiques, Rapport pour l'ADEME.
6. Wendy Bohtel and Kees Maat, Deriving and Validating Trip Destinations and Modes for Multi-day GPS based Travel Surveys: an Application in the Netherlands, Transportation Research Record: Journal of the Transportation Research Board, No. 2268, Transportation Research Board of the National Academies, Washington D.C., 2008


18. Zmud, Johanna and Jean Wolf, Identifying the Correlates of Trip Misreporting – Results from the California Statewide Household Travel Survey GPS Study. Paper presented at the 10th International Conference on Travel Behavior Research, Lucerne, August 2003


