A new methodology for a household travel survey by associating a GPS, a GIS and an expert system for Paris Region. First Part.

NGUYEN-LUONG Dany, COINDET Jean-Paul

Institute for Urban Planning and Development of Paris Region (IAU île-de-France)
15 rue Falguière, 75740 Paris Cedex 15, France.
E-mail : dany.nguyen-luong@iau-idf.fr , jean-paul.coindet@iau-idf.fr

---

Abstract

The Global Travel Survey ("Enquête Globale Transport" or EGT in french) in Paris Region is a very important tool of observation and measure of the persons' mobility. The methodology benefits of a background of more than thirty years. But there is a scientific consensus to consider that the EGT are not exempt from flaws : the deficiencies concerning short trips (especially walk trips), infrequent trips and stops during trips chain which are not always declared, the defiencies about routes, the inaccuracies about distances, the inaccuracies about durations, the complexity of the paper diary survey which makes it difficult to be managed, the increasing refusal to welcome at home pollsters and the problem of the cost. A review of the literature shows that the person-based GPS survey was not considered until 2007 as a mean to replace totally a traditional paper-based survey but just an aid to a paper survey. From 2008, the spirits seem to change a little. Thus you can cite the first experiment of a GPS-only survey in the US currently conducted in Cincinatti (4000 households).

So we study for Paris Region the feasibility of a new methodology by associating three technologies: a Global Positioning System, a Geographical Information System and an Expert System. It aims to remedy drawbacks of the current EGT by minimizing the human intervention and by homogenizing the data collect. We try to synthetize a complete household travel survey based *uniquely* on the collection of GPS data, their post-treatment by a GIS and the automatic generation by an expert system of a trips and segments database as complete as possible, similar in any points to a database resulting from a classical travel survey. This generated database would bring, on one hand, new information about routes and very short trips which are often omitted, on the other hand accuracy on fundamental attributes such as distances and durations. But the new methodology will encounter new difficulties. For example, the most difficult attributes to recognize are the trips purposes. They could be identified automatically by matching the GPS points cover and the detailed numerical land use cover of the Regional Geographical Information System. This cover contains on Paris Region (12,000 km2) roughly 400,000 polygons coded in 83 different uses. But it seems necessary to organize a light follow-up survey to precise the trip purposes. For the modes, they should be easier to be recognized. We will use the cinematic features of each model and the matching with the GIS transport networks covers.

After a quick review of a few GPS units types, we purchased two units, one called IGOTU and the other one Super Trackstick. Both are light and passive battery-powered devices. We think that a passive monitoring is the best suited method to minimize the burden on survey participant : the traveler has nothing to do except turn it on before his first trip of the day, carry it all day, turn it off in the end of the day, and reload the battery once in the middle of the week. It's even possible to give two loggers to the participant so that he has not to reload.

The GPS data are downloaded in @trip, an interacting mapping software and then connected automatically to Google Earth. Each route is displayed, with information on distance, time and speed. The GPS logger records raw data points as latitude/longitude coordinates which can be saved in a csv file. We have implemented a data transfer protocol in the ESRI ArcGis software to project the GPS points in two steps, at first into the World Geodetic System (WGS-1984), then into the Lambert 93 System to be consistent with the numerical land use layer.

We have tested the devices before launching the test on twenty volunteers. Our findings are the following ones :

- We confirm that the average margin of error is lower than 50 meters, as specified by the manufacturers. The highest gaps are due to the canyion effect, but we will be able to solve this problem by working with buffers in the GIS. One of the advantages about GPS is its spatial accuracy. For the classical EGT, the destination end point has to be located manually by the pollster in a grid of 300 meters squared cells. It's a tedious task. With a person-based GPS survey, the location is automatically recorded at the postal address.
- The GPS data provide very accurately the start time and start location, end time and end location of individual modal segments, as well as the route which can be displayed on a map. These information provide a high-level view of how people move around. You can't visualize the routes with just an alpha-numerical database coming from the classical household travel survey.
- For the modes, it seems feasible to identify the right mode. You can match on the GIS a buffer along GPS links and the transit infrastructure use to identify the mode "train". Given the complete loss of data during subway trip segments and the unique locations of subways entrances, the starts and ends of individual subway segments are also identifiable. To discriminate the surface modes, the variables should be the speed, the acceleration, the stops time and the cumulative distance traveled. Each mode has its own cinematic features, which could be described by period of time (peak hour and off-peak) and stored in a database that will be used by the expert system to recognize the mode of each segment. But we're aware that there may be various situations where these features will be wrong (for example people with physical disabilities). Many other geocoded databases are available in our GIS : a named-street cover, a subway network cover with all its stations entrances, a transit network cover with all its stops, a bus network cover (not yet with the stops).
- An important concern is the warm start. It's the time it takes for a GPS device to lock on to a satellite signal from the moment the traveler goes out a building or a subway station. This time can vary depending on the available satellite coverage in the sky and the height of buildings in the neighborhood. There is a systematic delay at the beginning of the trips. The data may miss up to a block of walking if people leave a building or a home in a hurry. So it will be necessary to put down the device fifteen minutes at least before leaving if possible somewhere exposed to the open sky.
- There may be parasite points recorded during time spent inside buildings (problem known as "multipath"), so we'll have to develop rules and algorithms to interpret these parasite points.
- The main problem to address concerns the purposes. At first sight, it seems difficult to infer in an automatic manner the purposes other than home purpose, because the trips origin or destination do not always match well with the underlying land use. But we'll try to build rules to discriminate as much as possible. For example, the length of stay could help to differentiate the workers from the visitors at a medical office. Another case where the spatial matching between trips origin or destination with the detailed land use may not work is for example a traveler who parks his car outside the building and forgets his device in the car. In that case, in order to obtain indications about trip purposes, we could set tolerance distances to

find suitable assignments of trip ends to particular lots or buildings. It will be also necessary to cross-check with other information collected in the household-individual survey and probably with a short follow-up call survey. And this additional survey could be personalized according to each trip and be generated automatically by the expert system.

The study will be carried out in two stages. The communication at the Tempe Conference will present the first stage : the context and motivation of the project, the problems of the classical travel survey, the objectives of the new methodology, and especially the results of the first test which will take place in January 2010 and consist in realizing **a one-week survey test on twenty volunteers** of the Institute or members of their family. For this test, the traditional paper survey will be conducted in the same time by another team, that will allow to compare the results of both surveys. There will be four waves of five persons surveyed over a seven-day period (one week). This survey will be held in January 2010. The method of the blind test will be used. Two teams will be involved. The first one will tackle the GPS-equipped survey. The other team will conduct the classical paper-based survey for the 20 person's trips and then build the trips and segments database in the data structure of EGT. The first team will have to build the trips and segments database in the same structure without having any information coming from the second team, except the basic information about households and individuals features (home and workplace addresses). That means that at that stage, there is not yet an expert system. We are the human experts who will work in front of the screen, post-process data "by hand" to recognize the modes, the purposes, and the hundred or so other attributes. Then the two databases will be compared. This stage will allow to analyze the advantages and drawbacks of a total GPS-equipped survey, and to determine what is necessary as new data (for example, the cinematic tables of each mode), to formalize the human process that conducts to identify or discriminate the ambiguous cases. Finally, we will develop algorithms to generate in an automatic manner all the attributes. So we'll have to develop a prototype of an expert system (in the Cincinatti case, a neural network is developed) and prove it is practical and computationally feasible. It would be interesting to begin to think in the end of this stage about new conditions of survey with a minimized human intervention, knowing that the households-individuals survey can't be automatized. We have to envisage for this part a phoning survey or a self-administered paper-based survey as in the French National Census, with additional questions about the main frequent trips destination location (lunch, shopping, visit).

For the second stage, we plan a real bigger survey pilot in 2012, with a random sample of about one hundred households in Paris region, in the conditions of a future full-scale regional travel GPS-equipped survey.

We have developed a French website to communicate on this research : www.iau-idf.fr/egtpargps

Acknowledgment

We gratefully acknowledge the financial support by the French Predit GO3 for the first stage of the research.