Measure Door-to-Door Mobility Through GPS Tracks & Big Data Methodologies

Illustrated by 2 Case Studies:
City of Caen: Mobility in Public Transportation Network
Working Area in Ile de France: In and Out flows (modes & frequentation)

by

Thibaud Arnoux, Sylvain Coppéré, Stéphane Mastalerz, Félix Motot, Vincent Loucel

Kisio Etudes & Conseil - 12 rue Hector Malot, 75012 Paris - France

Abstract

Before dreaming of shared, autonomous and virtuous mobility, we should be able to answer these simple questions: where do people come from, where do they go, and which mode(s) do they use?

It is clear that the decision-makers are still severely lacking comprehensive data on the understanding of the movements of travelers. The knowledge of the movements on the territories is partial and punctual. It can only be based on samples or models, which are by nature limited.

The interest of GPS tracks

The global transport surveys, cell phones tracks collected by cell towers, household surveys or counting fieldworks allow to apprehend only part of the trips over limited periods or geographical scope because of their cost and their methodological limits.

Without all the virtues, GPS data from smartphones or cars are a promise to tackle our main issues: observed data, anonymous data, updatable data, cheaper data, massive data. These GPS data are the XY coordinates (longitude and latitude) left by smartphones that generate between 200 and 400 points per day depending on the use of the device and the frequency of rides.

A paradigm shift for the study of mobility

We go from a “deductive” model analysis of a typical day or timeslot to an “inductive” analysis of evolution and variability, much closer to real behaviors.

Once well processed, GPS tracks allow impressive results:

- A true multimodal analysis door to door: end of analysis by combining isolated heterogeneous sources (public transport, road, bike loops...)

- An observed mobility: reality versus feelings (rides’ duration, lines saturation…)

- A detailed geographical analysis adapted to the new mobilities (intelligent transport on demand, carpooling, autonomous vehicles ...)

- A regular thermometer of the evolution of the mobilities (the data being less expensive and updatable) which makes possible the analysis of the variability of the behaviors

Technical issues to master

It is easy enough to drown in this flood of raw and massive data. The challenge is to extract value by enriching these data and deriving key indicators for a complete and robust mobility diagnosis. We investigate in this paper the different steps of data cleansing and analysis.

1/ Data cleansing: like in any data source, there are artifacts or noises in GPS signals that can be reassembled. It is then necessary to set up a succession of controls to qualify and keep robust data (smoothed data, sensors diagnosis…)

2/ Detection of rides: it is a question of extracting from several hundreds of time stamped GPS points a succession of paths. Using spatial libraries from PostGIS, we work on clustering to densify GPS points and detect origin & destination.

3/ Modal allocation: each ride track can then be confronted with an infrastructure network and transport offers (often materialized in GTFS format) to assign a travelling mode: car, public transportation, walk or bike.... We investigate several approaches as adapted map matching algorithms or correlation with searches of theorical optimal itineraries (PgRouting with OpenStreetMap data for example).

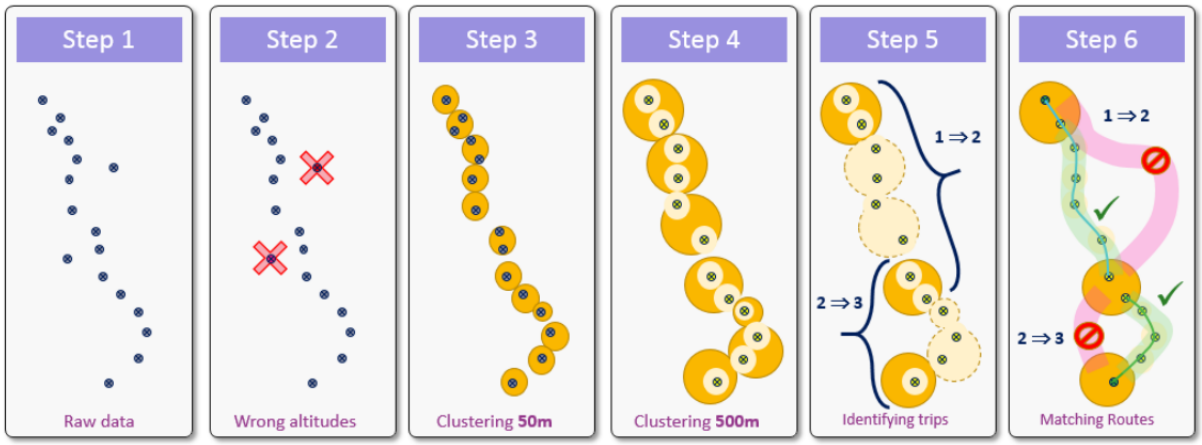
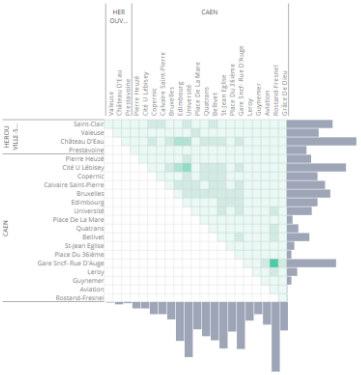
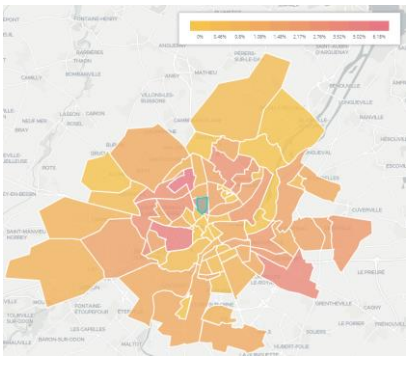


Fig1. First processing steps illustration

4/ Calibration: the data are collected on a "panel" of the population (users holding applications and agreeing to trace their locations). Then there is a matter of weighting these data to access information representative of the global studied population. We put in place a calibration methodology including socio-demographic data to de-bias our panel. From 4 to 6% of the population in our sample, we can offer robust results.

5/ Data visualization: the calibrated data are transformed into a number of different indicators: hour-by-hour line usage, origin-destination matrix, modal share by OD, trip duration… These indicators are then valorized through various dynamic and original outputs allowing a complete diagnosis of the mobility of the studied territory.





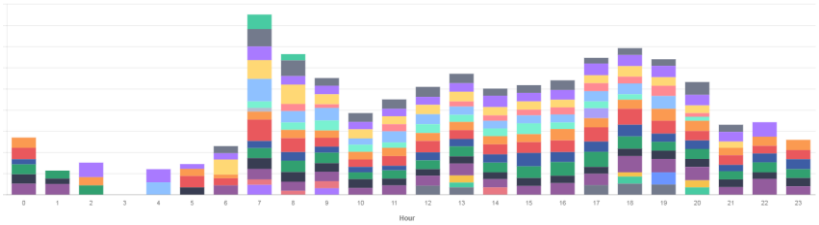


Fig2. Examples of interactive dataviz

Case studies

With our partners Moovit & Kisio Digital, those developments allow us to expand our vision of mobility. We illustrate in this paper some of our outcomes by exposing two case studies.

1/ City of Caen: from the analysis of GPS tracks from 14 000 citizens, we define an observatory of the mobility at Caen.

- Modal market shares by Origin-Destination combination
- Frequentation of bus lines at different timeslots (peak hours, workdays, vacations…), boardings, alightings, loading factors…
- A full understanding of intermodalities connections between lines, stations, modes…
- A measurement of the efficiency of the network district by district
- The understanding of the full rides: duration, walking distances, number of connections...

2/ In & Out flows of a working area at Velizy: Focusing on the rides of people in destination of a targeted area, the purpose is to evaluate the relevancy of the transportation offer and propose adjustments if necessary.

- Arrivals & departures hours on the area
- Duration and complexity of employees' rides
- Congested modes of transportation (analysis in peak hours, adequation between offer & demand…)

References

- Andrienko, Gennady & Andrienko, N & Kopanakis, Ioannis & Ligtenberg, Arend & Wrobel, S. (2008). Mobility, Data Mining and Privacy. "Mobility, Data Mining and Privacy - Geographic Knowledge Discovery", F.Giannoni and D. Pedreschi (eds), Springer. 375-409. 10.1007/978-3-540-75177-9_14.

- Patrick Bonnel, Mariem Fekih, Zbigniew Smoreda (2018). Origin-Destination estimation using mobile network probe data

- Patrick Bonnel, Marcela A. Munizaga (2018). Transport survey methods - in the era of big data facing new and old challenges

- Danya Bachir. Estimating urban mobility with mobile network geolocation data mining. Networking and Internet Architecture [cs.NI]. Université Paris-Saclay, 2019. English. .

- Thi Huong Thao Pham. Apports et difficultés d'une collecte de données à l'aide de récepteurs GPS pour réaliser une enquête sur la mobilité. Economies et finances. Université Paris-Est, 2016. Français. 〈NNT : 2016PESC0060〉. 〈tel-01429677〉

- Michel Bierlaire, Jingmin Chen,Jeffrey Newman (2013). A probabilistic map matching method for smartphone GPS data - Transportation Research Part C: Emerging Technologies

- Florian Masse, Sébastien Oliveau, Samuel Carpentier, Frédéric Audard, Lionel Kieffer. Des Big Data à la place des enquêtes? Enjeux scientifiques et éthiques pour l'étude des mobilités quotidiennes. Communication à la journée d'étude interdisciplinaire « Enjeux, usages, éthiques et droit du Big .. 2016. 〈halshs-01296345〉

- Maguelonne Chandesris, Fabrice Ganansia, Anaïs Rémy (2017) / Chapitre 6 – Les données massives au service des mobilités de demain / Big Data et politiques publiques dans les transports / André DE PALMA, Sophie DANTAN / Economica

- Etienne Côme, Anaïs Rémy (2017) / Multiscale spatio-temporal data aggregation and mapping for urban data exploration / Processing, mining and visualizing massive urban data ESANN / Brugges, Belgique

- Anaïs Rémy, Maguelonne Chandesris, Stephane Mastalerz, Anaïs Hyenne, Etienne Bousquie (2016) / Multimodal travel demand based on itinerary requests / World Conference on Transport Research / Shanghai, China