# Do Mobile Phone Data Allow Estimating Real Human Trajectory?

Sahar Hoteit, Stefano Secci, Stanislav Sobolevsky, Guy Pujolle, Carlo Ratti

# Do Mobile Phone Data Allow Estimating Real Human Trajectory?

Sahar Hoteit[†§], Stefano Secci[†], Stanislav Sobolevsky[§], Guy Pujolle[†], Carlo Ratti[§]
[†] LIP6/UPMC - University of Paris VI; 4 Place Jussieu, 75005 Paris, France
[§] MIT Senseable City Laboratory, 77 Massachusetts Avenue Cambridge, MA 02139 USA.
E-mail: sahar.hoteit@lip6.fr, stefano.secci@lip6.fr, stanly@mit.edu, guy.pujolle@lip6.fr, ratti@mit.edu

Nowadays, the huge worldwide mobile-phone penetration is increasingly turning the mobile network into a gigantic ubiquitous sensing platform, enabling large-scale analysis and applications. In recent years, mobile data-based research reaches important conclusions about various aspects of human mobility patterns and trajectories. But how accurately do these conclusions reflect the reality?

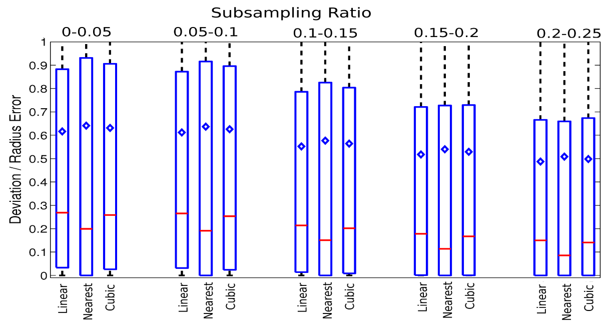
In order to evaluate the difference between the reality and the approximation methods, we study the error between real human trajectory and the one obtained through mobile phone data using different interpolation methods (linear, cubic, nearest, spline interpolations) and taking into consideration mobility parameters.

We use for this aim a dataset consisting of anonymous cellular phone signaling data, it consists of location estimations for about one million devices in the Boston metropolitan area.

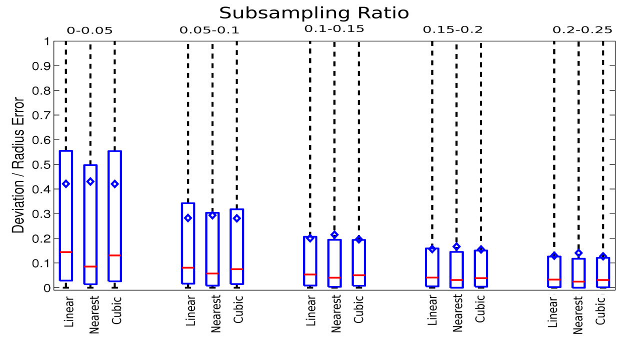
To evaluate the error between real human trajectories and the estimated ones, we fine-select data of those smartphones holders with a lot of samplings, typically those data-plan users with persistent Internet connectivity due to applications such as e-mail synch. Then, in order to reproduce *artificial* “normal user” sampling, we subsample *real* data-plan smartphone quasi-continuous traces according to an experimental inter-event statistical distribution. Therefore, we extract, from the real trajectory, a first random position then the corresponding next positions are extracted according to the inter-event time distribution values.

Hence, given a real trajectory with a high number of positions, and its subsampling that reproduces normal user’s activity, we apply an interpolation method to estimate the trajectory across the given points. Given the real trajectory points $P_i$, we estimate its corresponding position in time in the estimated trajectory: $P_i'$. Then we determine the deviation between the two points]as the distance separating the exact position $P_i$ to the estimated position $P_i'$ in the interpolating curve joining the samples. To take into account mobility habits, we categorise the users depending on their “radius of gyration” defined by the deviation of user positions from the user centroid position.

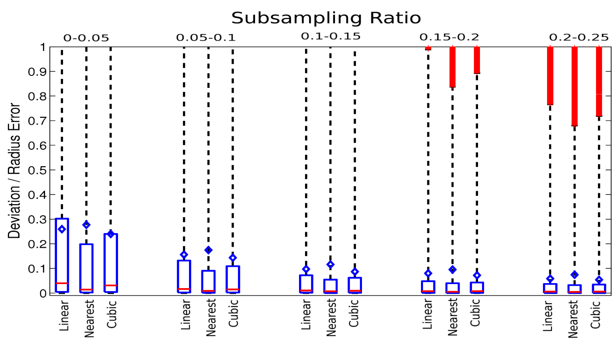
From extensive evaluations based on real cellular network data of the Boston metropolitan area, we show that the linear interpolation offers the best estimation for sedentary people (with a small radius of gyration) and the cubic one for commuters (having a big radius of gyration). Moreover, the nearest interpolation appears as the best



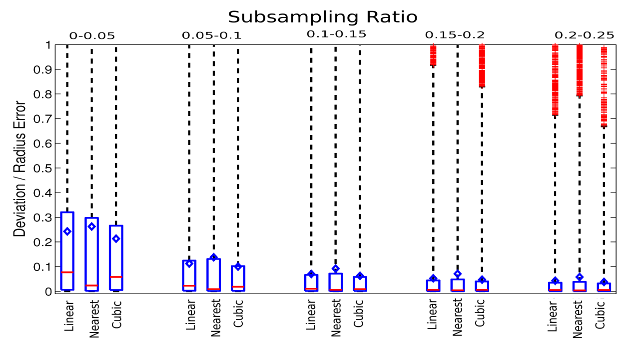

(a) rg<=3km




(b) 3km<rg<=10km




(c) 10km<rg<=32km




(d) rg>32km

Fig. 1: Boxplots of trajectory error

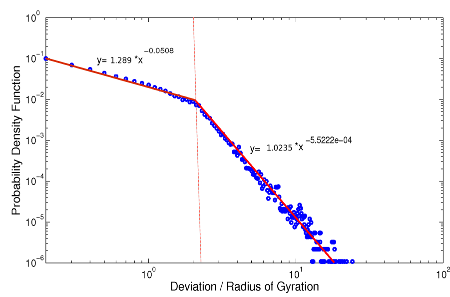



Fig. 2: Probability density function of error

one for "ordinary people" doing regular stops and standard displacements (Figure 1).

Another important experimental finding is that trajectory estimation methods show different error regimes whether used within or outside the "territory" of the user defined by the radius of gyration. The distribution of errors over all users' positions is approximated by a combination of two power law distributions joined by a breakpoint (approximately equal to 2.2) for the different interpolation methods (Figure 2).

As a future work we aim to estimate the positions of hotspots in a region knowing only the mobility characteristics of its residents.