Using Surrogate Road Network for Map-Matching A Sensitivity Analysis of Positional Accuracy

Y. Méneroux*[1], A. Le Guilcher[1], O. Orfila[2], B. Lusetti[2], G. Saint Pierre[2], and S. Mustière[1]

[1]Univ Paris-Est, LASTIG COGIT, IGN, ENSG, 73, avenue de Paris, F-94160 Saint-Mandé, France

[2]French Institute of Science and Technology for Transport, Development and Networks (IFSTTAR/LIVIC), 77, rue des Chantiers, 78000 Versailles, France

*Email: yann.meneroux@ign.fr

Abstract

Map-matching GPS traces on a reference road network often increases point positional accuracy, especially in urban environment where satellite signal is frequently obstructed by buildings. However, with receivers improvement, road data errors may no longer be ipso facto considered as negligible in front of GPS trace errors. This raises the question of sensitivity of map-matching output precision to input network geometric quality. We address the problem by attempting to relate gain in precision with a network quality index, based on two commonly used measures : averaged Hausdorff distance and areal difference. Analysis has been conducted by performing both GPS traces simulation and network perturbation. Results highlighted that the areal difference (for which we demonstrate that it can be considered as a distance) seems to be better related to map-matching output errors. We also provide an upper bound on the impact of using a surrogate reference network and we apply it to a practical case study.

Keywords: Spatial Data Quality, Map-Matching, Road Networks, Sensitivity Analysis.

1 Introduction

With the spread of terminal devices and applications relying on accurate location, Global Positioning System (GPS) traces are being increasingly used by studies focusing on individual mobilities, especially when massive, volunteered and inexpensive data are required. However, most of the time, an important assumption is that receivers are moving in a constrained space (e.g. road network) and map-matching becomes a prerequisite operation. More formally, given a sequence of points corresponding to noisy observations of a vehicle trajectory, map-matching consists in finding a trajectory that both explains the observations and is a realistic displacement on the network. This pre-processing step is particularly important in *map inference*, where one of the objectives is to provide attributes (e.g. road width, number of lanes...) to a given road network, based on the analysis of traces stemming from in-vehicle embedded GPS devices (van Winden et al. (2016)).

Operating map-matching before analyzing traces has two, possibly combined, main advantages:

*

providing a mapping function between GPS positions and network edges (which is necessary for updating the network with traces-based inference) and enhancing position accuracy (since the location uncertainty is completely reduced along at least one direction). The latter is particularly important in urban environment, where GPS satellite signal is likely to be partially covered by buildings. Of course, all this holds provided that the reference network has *absolute* precision. This is obviously unrealistic in practice, but it is usually enough to assume that its errors are one order of magnitude lower than GPS error to be confident in the benefits of map-matching. However, what happens when it is not the case ?

In this paper we investigate the impact of using a low-quality reference network for map-matching an in-vehicle embedded GPS trace. This is particularly important for example, if two networks of different accuracy levels are available, but the less accurate of them is also the most complete regarding attributes. It may also be the case, that getting a more accurate reference network is costly (in terms of money or computing time). Hence we need some criteria to decide whether the loss in positional accuracy is acceptable. We shall note that some algorithms have been specifically designed to merge two networks, when one of them has better geometric quality, while the other one is more complete (Mustière and Devogele (2008)). Yet, these algorithms are seldom straightforward to implement, while a simple decision criteria may sometimes be enough to save a lot of development time.

We shall notice that many different map-matching approaches have been proposed so far (Quddus et al. (2007)). In most algorithms, the final stage consists in orthogonal projections of GPS points onto road segments. As in our study, we are dealing with network geometric imprecision, we will restrain to investigate the impact of positional quality on this common processing step.

2 Methodology

Map-matching operation can be thought of as a mathematical function f, whose inputs are GPS sequence of points $X = \{x_1, x_2, ..., x_n,\ x_i \in \mathbb{R}^2\}$ along with a reference network $\mathcal{R}$ (*i.e.* a topological graph, where each edge is associated with a polyline describing the geometric modeling of a road portion) and output is the *projection* of X on $\mathcal{R}$. It is then possible to formulate our problem as a sensitivity analysis of function f, i.e. how the variation in the output of a numerical model is dependent on the input uncertainties. Our methodology is described hereafter, according to figure 1 (inspired from Saltelli et al. (2000)).

BDTOPO© [1] road network of Paris has been used for the experiment and is arbitrarily considered as a ground truth ideal network $\mathcal{I}$. Let us denote E the space of all representations of $\mathcal{I}$. We only focus on sensitivity due to the network geometric accuracy, then a collection of perturbated networks $\mathcal{R}_1, \mathcal{R}_2, ...\mathcal{R}_k$ is generated with a method inspired by Vauglin (1997), i.e. by simulating two stochastic processes ε_x and ε_y along each direction, with LU-factorization of covariance matrix. We used an exponential model of variogram, with random sill and range, and a 2% nugget effect.

[1] French national mapping agency topographic database

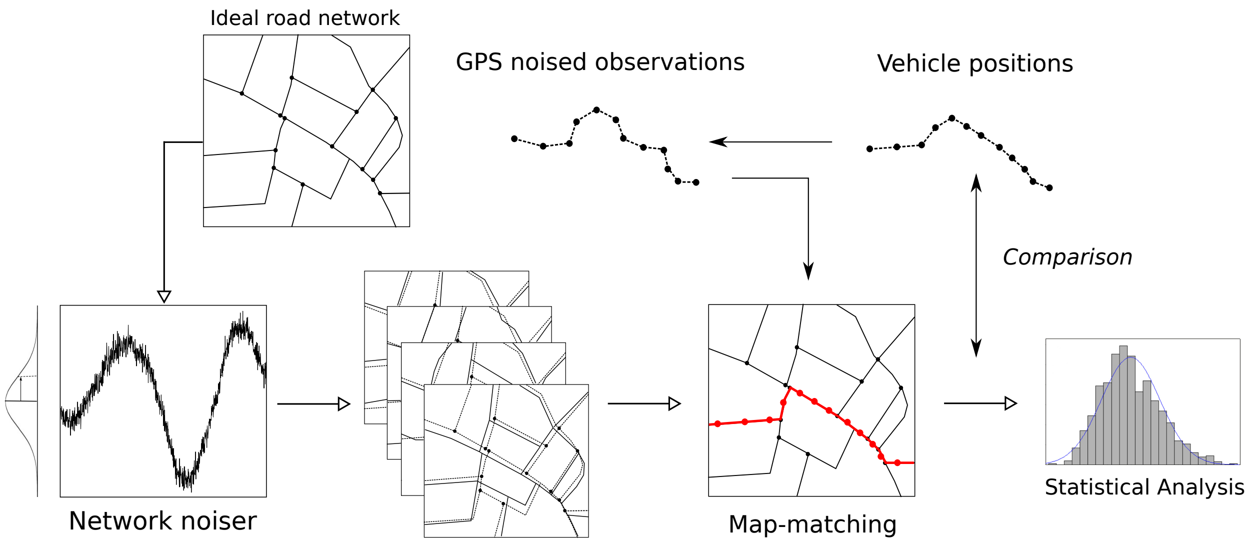

Figure 1: Experimentation framework

This step enabled to introduce realistic noise in the ground truth network, while avoiding most of topological inconsistencies (Bonin (2002)).

It is then required to define a measure of geometric quality for each representation. For some reasons that will be detailed further, we want this measure to behave as a distance on space E. As a direct implication $||\mathcal{R}|| = d(\mathcal{R}, \mathcal{I})$ may be considered as a quality index of network $\mathcal{R}$. It is also required that this measure is often used in the geospatial sciences, to provide our results with potential for application in practical situations. Accordingly, we used averaged Hausdorff distance, which is classically used to measure the difference between networks (Girres and Touya (2010)) and which is computed as a mean of all Hausdorff distances between couples of corresponding polylines. We also considered areal difference (McMaster (1986)):

$$d(\mathcal{R}_1, \mathcal{R}_2) = \frac{1}{N} \sum_{i=1}^{N} \frac{\mathcal{A}(l_1^i, l_2^i)}{\min(|l_1^i|, |l_2^i|)} \tag{1}$$

where l_1^i denotes de i^{th} polyline of network j, $|.|$ is its length and $\mathcal{A}$ is the area of the polygone defined by the closed linestring $l_1 l_2$.

Note that this is a modified version of the classical definition, which is normalized by the maximal or the average length of polylines. Here we propose an alternative definition which, in turn, enables to demonstrate under some hypotheses that $d(.,.)$ is a distance. In particular, triangular inequality guarantees that for any three networks $x, y, z \in E$: $d(x, z) \leqslant d(x, y) + d(y, z)$.

Eventually, for each generated instance of network representation, we simulated a vehicle path along with its GPS observation trace, and performed map-matching on both the ideal ground truth $\mathcal{I}$ and its perturbated version $\mathcal{R}$. Root mean square of errors (rmse) in positions is then computed and we define the gain in precision as $g : t \to \mathbb{E}\big[\sigma/\sigma(\mathcal{R}) \mid ||\mathcal{R}|| = t\big]$ where σ is the GPS rmse and $\sigma(\mathcal{R})$ is the rmse of GPS points map-matched on network $\mathcal{R}$. Simply put, g quantifies the conditional expectation of gain in accuracy (compared to GPS accuracy) after map-matching points on a network whose quality is known.

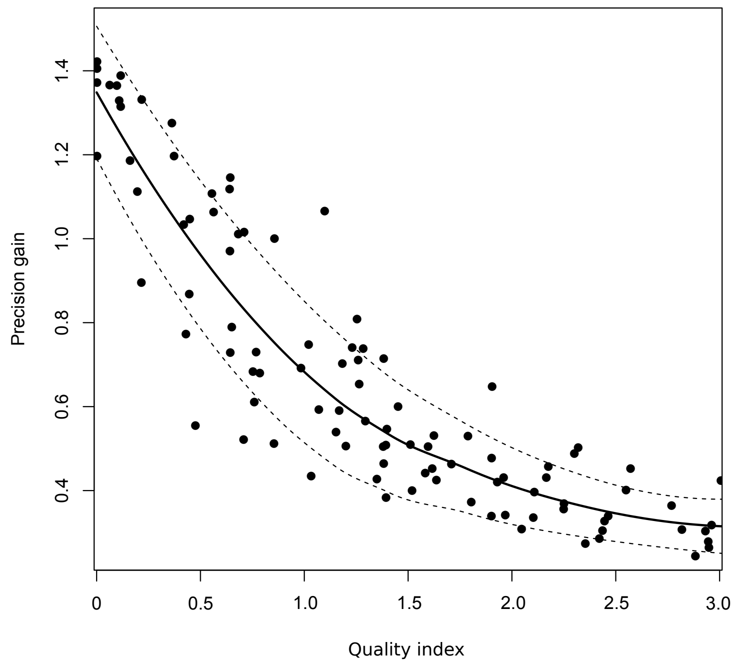

Figure 2: Gain in accuracy for different levels of quality (measured as areal distance to ideal network), loess regression and 95% fluctuation interval. Horizontal unit normalized by gps rmse.

Each random realization is a point in the space [gain $\times$ network quality] and we used local least squares regression to get a nonparametric estimate of g (figure 2). Then results (for each distance definition) are compared.

3 Results and Discussion

We conducted two experiments: in the first one both the path in the network (hence its associated GPS simulation) as well as the network representation have been generated at the same time for each simulation. In the second experiment, we assessed the impact of the network quality *ceteris paribus*, and all points are simulated from the same network realization. As a consequence, graphics depicted on figure 3 illustrate two different phenomena. The left one (first experimentation) depicts the regressions of g for each distance definition, with variable path. The right graphics depicts the evolution of an approximation of g, for a single path, and then reflects how smooth g is likely to be.

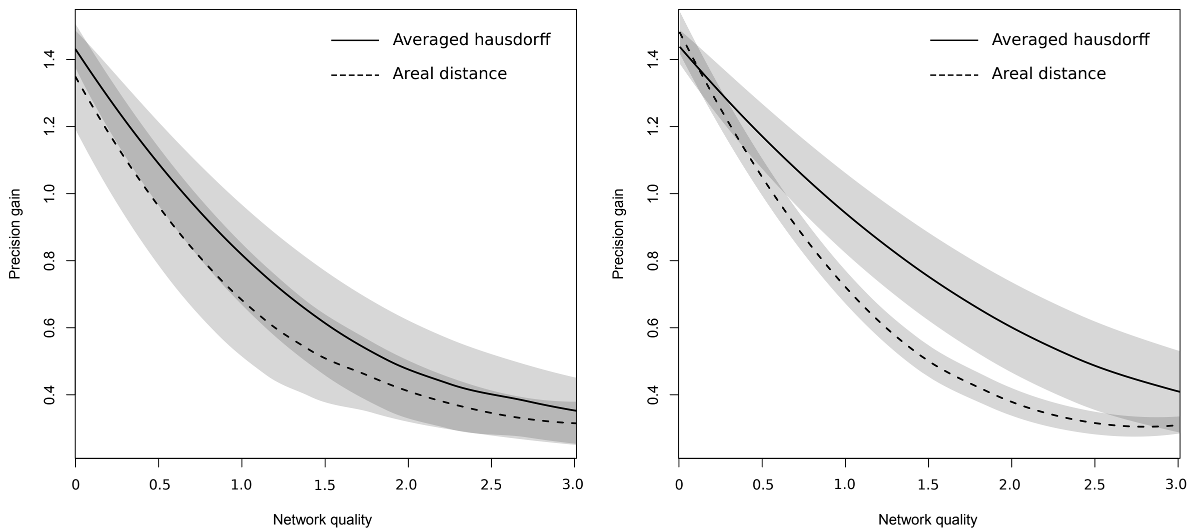

Figure 3: Gain in accuracy for different levels of quality (measured as averaged Hausdorff and areal distances to ideal network). Left graphics depicts experimentation results on 1000 different paths. Right side is average result on a single path. Horizontal unit normalized by gps rmse.

From these results, it seems that the areal distance based quality measure has some higher degree of proficiency to *capture* the underlying phenomenon. Indeed, its fluctuation interval seems much thinner than its averaged Hausdorff distance counterpart for the case of a single path, while both measures seem to have same stability when no prior knowledge is available on the vehicle trajectory. At this step, using the distance property of the quality index and mean value inequality, it is possible to get an upper bound for the impact on map-matching accuracy gain of using a network $\mathcal{R}_1$ instead of $\mathcal{R}_2$, assuming we have a rough estimate of the distance between these two representations.

$$\frac{|\Delta\sigma|}{\sigma} \leqslant \max_{t\in\mathbb{R}^+} \left|\frac{g'(t)}{g(t)^2}\right| \times d(\mathcal{R}_1, \mathcal{R}_2) \tag{2}$$

Using numerical derivation and assuming both networks are at least in the quality range [0-10] m, an estimate of the factor is computed at 1.08 for areal distance and 0.81 for averaged Hausdorff distance.

As an example of practical application, for a map inference research work, it has been decided to map-match GPS traces (σ = 10 m) on Open Street Map road network, whose geometric quality is believed to be lower than national cartographic reference. The motivation behind this choice, is that OSM network contains additional information that may be of interest for the study. To confirm that this choice would not have too much impact, we measured the areal distance between these two reference networks ($d = 0.13\ \sigma$) and by applying equation (2) we found that the difference in map-matched accuracy on each network is expected to be at most $1.08 \times 0.13\ \sigma = 1.40$ m.

4 Conclusion

In this paper, we have conducted GPS points simulation and road network perturbation to assess the impact of using a low-quality network as input in a map-matching algorithm. Network accuracy has been quantified with averaged Hausdorff distance and we introduced a modified version of areal displacement measure for which we demonstrate that it behaves as a mathematical distance. Results highlighted that this second measure seems to have better properties as far as map-matching sensitivity analysis is concerned. Eventually, we derived an upper bound on the impact, which practitioners may utilize as a guideline to decide whether map-matching on a given surrogate network is acceptable. In the near future we would like to confirm these preliminary results with a field experimentation, by map-matching actual GPS traces on low-quality networks. It would be interesting as well to conduct similar analysis with different quality indexes such as polyline sinuosity, which may provide a sharper bound on the impact upon map-matching accuracy enhancement.

5 Acknowledgements

We would like to express our gratitude to Xavier Collilieux, Clémentine Mathey and Guillaume Touya from IGN, for precious assistance and advices on GPS techniques and road data quality.

6 References

Bonin, O. (2002), 'Modèle d'erreurs dans une base de données géographiques et grandes déviations pour des sommes pondérées ; application à l'estimation d'erreurs sur un temps de parcours'.

Girres, J.-F. and Touya, G. (2010), 'Quality assessment of the french openstreetmap dataset', *Transactions in GIS* **14**(4), 435–460.

McMaster, R. B. (1986), 'A statistical analysis of mathematical measures for linear simplification', *The American Cartographer* **13**(2), 103–116.

Mustière, S. and Devogele, T. (2008), 'Matching networks with different levels of detail', *GeoInformatica* **12**(4), 435–453.

Quddus, M. A., Ochieng, W. Y. and Noland, R. B. (2007), 'Current map-matching algorithms for transport applications: State-of-the art and future research directions', *Transportation Research Part C: Emerging Technologies* **15**(5), 312 – 328.

Saltelli, A., Chan, K. and Scott, E. M., eds (2000), *Sensitivity analysis*, Wiley series in probability and statistics, J. Wiley & sons, New York, Chichester, Weinheim.

van Winden, K., Biljecki, F. and van der Spek, S. (2016), 'Automatic update of road attributes by mining gps tracks', *Transactions in GIS* **20**(5), 664–683.

Vauglin, F. (1997), 'Modèles statistiques des imprécisions géométriques des objets géographiques linéaires'.