



HAL
open science

An extensive analysis and calibration of the Modular Aggregation Algorithm across three categories of for GNSS trajectories data sources

Marie-Dominique van Damme, Yann Méneroux, Ana-Maria Olteanu-Raimond

► To cite this version:

Marie-Dominique van Damme, Yann Méneroux, Ana-Maria Olteanu-Raimond. An extensive analysis and calibration of the Modular Aggregation Algorithm across three categories of for GNSS trajectories data sources. Laboratoire sciences et technologies de l'information géographique. 2024. hal-04697576

HAL Id: hal-04697576

<https://hal.science/hal-04697576v1>

Submitted on 13 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

An extensive analysis and calibration of the Modular Aggregation Algorithm across three categories of for GNSS trajectories data sources

Marie-Dominique Van Damme*, Yann Méneroux, Ana-Maria Olteanu-Raimond

Univ Gustave Eiffel, IGN-ENSG, LASTIG

73 Avenue de Paris

Saint-Mandé

France

`marie-dominique.van-damme@ensg.eu`, `yann.meneroux@ign.fr`, `ana-maria.raimond@ign.fr`

13 septembre 2024

* All authors contributed equally to this research.

1 Data and materials

Since our goal is to reconstruct the geometry of the path from a given subset of GPS trajectories, disregarding the time dimension, each dataset used in this study contains trajectories following the same route, travelling in the same direction, and having approximately the same start and end points. In this study, the data used consist of a collection of datasets that respect these conditions, aligning with those outlined in [Etienne and Devogele, 2014]. The scenarios of greatest interest for studying the aggregation method are paths under forest cover, due to the noise effect, and those containing a series of sharp handling challenges in aggregation. Moreover, our analysis will be carried out by using both synthetic and real trajectories. Thus three categories of data are considered : 1) Realistic Synthetic GNSS Trajectories : These simulated trajectories provide a controlled environment to test the algorithm’s performance under various predefined conditions, allowing us to assess its accuracy and robustness; 2) Multi-Sensor Trajectories : data collected with different sensors (professional, watched and smartphone applications) to evaluate how well the algorithm handles the heterogeneity of sensor data ; and 3) Multi-Canopy Trajectories Acquisition : this involves collecting trajectories in environments with varying canopy cover (*i.e.* open, medium, and dense forest). This category helps in understanding the algorithm’s performance in dealing with signal obstructions and varying environmental conditions. Both multi-sensors and multi-canopy trajectories are acquired by following the same repeatable data collection protocol, we defined.

1.1 Synthetic GNSS trajectories

In order to assess the metrological performances of the algorithm (*i.e.* its ability to reconstruct accurately the common path followed by all the individual sample trajectories) it is required to test it on a large array of configurations, with a substantial amount of GPS trajectories. This raises two operational problems :

First, to compare the estimated trajectory with the real route actually followed by the individual samples, it is required to perform a costly and time-consuming on-the-field survey to measure the ground truth. This is especially problematic since moderately to densely covered forest areas are of interest, where the direct use of precision carrier-phase GNSS with Real-Time Kinematic fast and convenient procedure is often impossible. Therefore, ground truth has to be measured with classical surveying traverse, tied to an absolute reference point, possibly located hundreds of meters away ; spending a full day of topometric survey for each case study is not realistic.

Secondly, for each case study, individual GNSS trajectories must be collected by walking several times along the ground truth route. Because GNSS error is known to be auto-correlated in time [Roberts, 1993], for a realistic acquisition of data, trajectories must ideally be sampled at different times of day, which increases again drastically the effort needed to collect real data.

To overcome these limitations, it was first decided to proceed to extensive experimentation of the algorithms on simulated GNSS trajectories. This enables to simulate as many case studies as needed, with potentially unlimited number of GNSS trajectories, and a readily available ground truth track to compare the results with. The used methodology is the following : for each case study, a reference track is simulated (or extracted from an existing topographic database) and is considered as the ground truth track from which all GNSS trajectories are simulated. The error between the estimated and ground truth track is then evaluated, which in turn, enables to assess the sensitivity of the algorithm to all its parameters. Similar methodologies have been used for example by [Biljecki et al., 2015] and [Zhang and Yang, 2015].

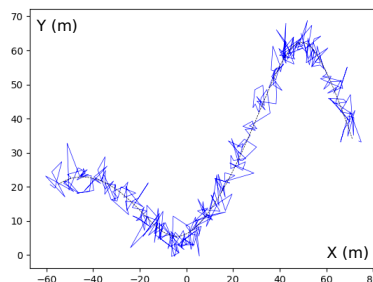


FIGURE 1 – Example of a non-realistic simulation of a GNSS trajectory with a white noise process, completely missing out the true correlation pattern of GNSS measurements.

This methodology however, requires an accurate modeling of auto-correlation error of GNSS trajectories, to avoid

non-realistic simulations, as depicted for example in Fig. 1, which can also result in topological errors [Bonin, 2002, Vauglin, 1997].

GNSS errors were then modeled through their covariance function $\gamma(s_1, s_2) = \text{Cov}(X(s_1), X(s_2))$ describing the statistical covariance between positioning errors $X(s_1)$ and $X(s_2)$ at two locations $s_1, s_2 \in \mathbb{R}_+$ (described through their curvilinear abscissa along the ground truth trajectory). Further, the error X is supposed to be a second-order stationary process (hence described only by the difference $s_2 - s_1$), and is modelled to take into account different error components in the GNSS trajectory measurement process, as illustrated on Fig 2.

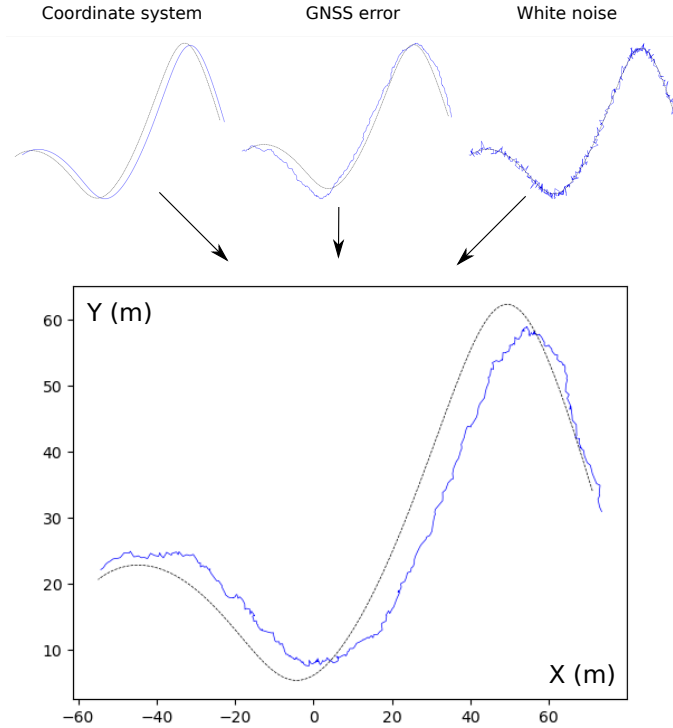


FIGURE 2 – Illustration of the three error components in the GNSS trajectories. From left to right : (1) a long wave-length process describing coordinate system errors, (2) an intermediate wave-length process describing GNSS observation errors (auto-correlated in space and time) and (3) a white noise process (e.g. heat, vibrations, electronic noise).

Generation of GNSS trajectories was done independently on each of the two planimetric components, with a methodology described in [Ripley, 2009] : with a random generator, we sampled n i.i.d. unit-variance and zero-mean gaussian values, compiled in a vector \mathbf{x} . It can easily be shown that, for any positive-definite matrix $\Sigma \in \mathbb{R}^{n \times n}$, the random vector $\mathbf{y} = \mathbf{A}\mathbf{x}$ where \mathbf{A} is a Cholesky factor of Σ , is a realization of a correlated random vector \mathbf{Y} having covariance matrix Σ . The covariance matrix Σ is formed with $\Sigma_{ij} = \gamma(s_j - s_i)$.

In our experimentation, trajectories have been generated with a 5 m-amplitude exponential covariance process [Grejner-Brzezinski, 2010] completed by a 1 m white noise process, a 50 cm range Gaussian Process for referencement error.

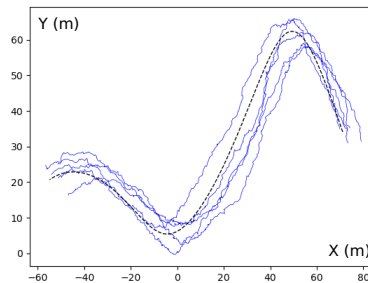


FIGURE 3 – Example of 5 synthetic GNSS trajectories (blue) generated on a common ground truth track (dashed line).

1.2 Multi-sensors and multi-canopy traces acquisition

To assess the impact of canopy and the sensors on precision accuracy, we defined and implemented a data collection protocol.

First, based on the literature, we identified three types of canopy (i.e. open area, moderate coverage, and heavy coverage). Additionally, we delineated five types of sensors (i.e. mobile phone equipped with VisioRando application, Polar GPS device, Garmin GPS device, Keymaze device and professional Ublox GPS sensor chip).

Second, for each type of canopy, the following data collection protocol was defined :

- Identification of areas without spatial constraints (e.g. bridge, stream, unobstructed)
- Identification within a rectangle of an Origin/Destination route with moderate winding and approximate length of 300 m.
- Collect five round trips following the route exactly.

Third, the field work was done by two of the authors of this paper. The placement of the sensors is also relevant according to the literature [Blunck et al., 2011]. To limit this effect, the GPS watches were worn on the wrist, the professional GPS devices were carried in a bag with the antenna positioned externally, and the mobile phones were held in the hand. Data collection has been carried out during the summer season. In total, for the three types of canopy and five sensors, 150 trajectories are collected.

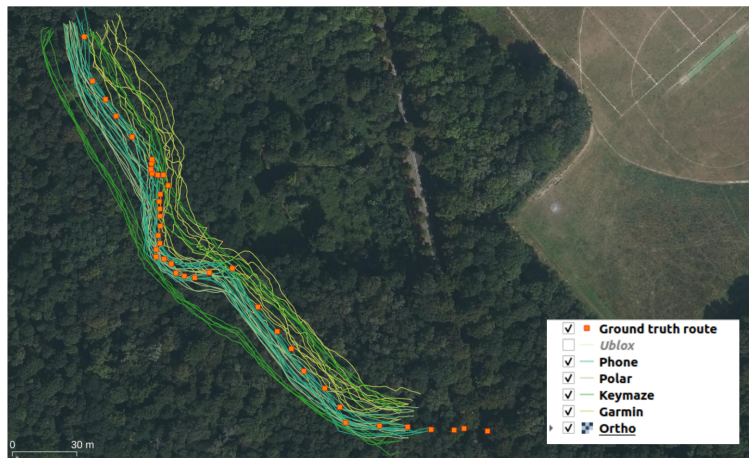


FIGURE 4 – Set of 50 trajectories collected with five sensors in dense forest and ground truth route.

The ground truth route is obtained through a topometric survey conducted by a group of students as part of a topometric project, under the supervision of the authors of this paper and their teachers [Calloch et al., 2024]. Absolute positioning is conducted through GNSS differential static positioning of a set of reference points located in open-sky conditions, between 300 and 500 meters from the surveyed route. Topometric determination of the route geometry is performed with surveying traverse. The output ground truth is sampled with about 42 points (i.e. about 1 point every 7 meters) with an absolute positioning accuracy of 5 mm in each 3D axis (1σ).

1.3 Crowdsourced traces

To validate the results obtained with synthetic GNSS tracks and data acquired according to the proposed protocol, we have chosen to test the algorithms on crowdsourced trajectories, as these will be used to derive pressure and route frequency indicators for the end-users.

To this end, our third experiment was carried out considering tracks downloaded from Visorando¹ and from Wikiloc², websites offering downloading tracks published online by contributors. The GPS sensors used are therefore unknown, as well as data changes, context conditions.

1. www.visorando.com

2. www.wikiloc.com

In this experiment, we focused on hiking activity in the mountain area of the Pralognan Valley in the French Alps. Specifically, from all the trajectories, we selected some of them having specific and challenging configurations and spatial constraints : terrain constraints (e.g. ridge, river), infrastructure support (e.g. switchback trail) and variations of the shape of the path (e.g. a straight line followed by a series of twists and turns with varying distances between them). Note that, in a dataset, all the selected trajectories have the same origin/destination (see Figure 5).



FIGURE 5 – Trajectories in three different contexts with spatial constraint : (a) ridge (terrain constraint), (b) series of sharp (infrastructure constraint), (c) heterogeneous trail shape (scale variation)

Note that in Figure 5, traces may have been previously filtered via a simplification algorithm like Douglas-Peucker. Indeed, all geometries offer the same characteristics like vertex of bends, turns in route. If this was not the case, traces would be more dissimilar in shape.

2 Bibliography

Références

- [Biljecki et al., 2015] Biljecki, F., Heuvelink, G. B., Ledoux, H., and Stoter, J. (2015). Propagation of positional error in 3d gis : estimation of the solar irradiation of building roofs. *International Journal of Geographical Information Science*, 29(12) :2269–2294.
- [Blunck et al., 2011] Blunck, H., Kjærgaard, M. B., and Toftegaard, T. S. (2011). Sensing and classifying impairments of gps reception on mobile devices. In Lyons, K., Hightower, J., and Huang, E. M., editors, *Pervasive Computing*, pages 350–367, Berlin, Heidelberg. Springer Berlin Heidelberg.
- [Bonin, 2002] 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*. PhD thesis, Paris 6.
- [Calloch et al., 2024] Calloch, P., Labbé, B., and Lorine, P. (2024). *Établissement d’une méthode low-cost de vérité terrain pour l’étude de la précision de traces GNSS de randonneurs en forêt*. PhD thesis.
- [Etienne and Devogele, 2014] Etienne, L. and Devogele, T. (2014). Trajectoires médianes. In *14 ème conférence Extraction et Gestion des Connaissances, Ateliers fouille de données spatiales et temporelles & construction, enrichissement et exploitation de ressources géographiques pour l’analyse de données*, Rennes, France.
- [Grejner-Brzezinska et al., 2005] Grejner-Brzezinska, D., Toth, C., and Yi, Y. (2005). On improving navigation accuracy of gps/ins systems. *Photogrammetric engineering & remote sensing*, 71(4) :377–389.
- [Ripley, 2009] Ripley, B. D. (2009). *Stochastic simulation*. John Wiley & Sons.
- [Roberts, 1993] Roberts, W. D. S. (1993). *GPS time correlation and its implication for precise navigation*. PhD thesis, Newcastle University.
- [Vauglin, 1997] Vauglin, F. (1997). *Modèles statistiques des imprécisions géométriques des objets géographiques linéaires*. PhD thesis, Université de Marne-la-Vallée (1991-2019).
- [Zhang and Yang, 2015] Zhang, Y. and Yang, Y. (2015). Cross-validation for selecting a model selection procedure. *Journal of Econometrics*, 187(1) :95–112.