



**HAL**  
open science

## A metrological analysis of a modular and iterative aggregation algorithm of GNSS trajectories

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

► **To cite this version:**

Marie-Dominique van Damme, Yann Méneroux, Ana-Maria Olteanu-Raimond. A metrological analysis of a modular and iterative aggregation algorithm of GNSS trajectories. SIGSPATIAL '24: The 32nd ACM International Conference on Advances in Geographic Information Systems, ACM SIGSPATIAL, Oct 2024, Atlanta (GA), United States. pp.633 - 636, 10.1145/3678717.3691325 . hal-04801943

**HAL Id: hal-04801943**

**<https://hal.science/hal-04801943v1>**

Submitted on 26 Nov 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.



# A metrological analysis of a modular and iterative aggregation algorithm of GNSS trajectories

Marie-Dominique Van Damme\*  
Univ Gustave Eiffel, IGN-ENSG,  
LASTIG  
Saint-Mandé, France  
marie-dominique.van-  
damme@ensg.eu

Yann Méneroux  
Univ Gustave Eiffel, IGN-ENSG,  
LASTIG  
Saint-Mandé, France  
yann.meneroux@ign.fr

Ana-Maria Olteanu-Raimond  
Univ Gustave Eiffel, IGN-ENSG,  
LASTIG  
Saint-Mandé, France  
ana-maria.raimond@ign.fr

## Abstract

This paper focuses on GNSS trajectory aggregation, building upon the work of [4], a modular and iterative aggregation algorithm. The last allows to compute aggregated trajectories with high geometric accuracy for a set of trajectories having the same origin/destinations and following the same path. To analyze the performances of the algorithm, we use a metrological perspective, *i.e.* its ability to reconstruct the accurate aggregated trajectory, from a minimum number of input GNSS trajectory samples to determine an optimal parameterization of the algorithm. We validate our modular framework on two types of data: synthetic and experimental GNSS trajectories collected with multiple sensors under varying canopy conditions.

## CCS Concepts

• Information systems → Spatial-temporal systems.

## Keywords

Trajectory aggregation, synthetic and real world trajectories

### ACM Reference Format:

Marie-Dominique Van Damme, Yann Méneroux, and Ana-Maria Olteanu-Raimond. 2024. A metrological analysis of a modular and iterative aggregation algorithm of GNSS trajectories. In *The 32nd ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '24)*, October 29–November 1, 2024, Atlanta, GA, USA. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3678717.3691325>

## 1 Introduction

Certain environments, such as mountains and coastal areas [6], are witnessing a rise in recreational visits for organized, individual, or family activities. The expansion of outdoor activities could have negative consequences on the preservation of biodiversity. Thus, it becomes crucial to better estimate the human pressure in time and space to guide informed management and conservation actions.

Recent research has demonstrated the potential of crowdsourcing data to enhance the representation of human pressure on ecosystems such as GNSS trajectories collected and shared openly by

\*All authors contributed equally to this research.

Publication rights licensed to ACM. ACM acknowledges that this contribution was authored or co-authored by an employee, contractor or affiliate of a national government. As such, the Government retains a nonexclusive, royalty-free right to publish or reproduce this article, or to allow others to do so, for Government purposes only. Request permissions from owner/author(s).

SIGSPATIAL '24, October 29–November 1, 2024, Atlanta, GA, USA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-1107-7/24/10

<https://doi.org/10.1145/3678717.3691325>

practitioners on different platforms (OpenStreetMap) or data provided by citizens through Citizen Science (CS) initiatives [10].

The global aim of this research is to study the potential of using crowdsourced GNSS trajectories to quantify human presence in outdoor environments and to create an accurate aggregated route network. This paper represents an initial step toward our goal, focusing on **trajectory aggregation**.

Among the methods used to compute accurate geometry of aggregated trajectories [2], [8], [1], the method proposed by [4] fits our purpose: it is iterative, making it robust to outliers; it relies on data matching points, ensuring high accuracy for the aggregated points in the final aggregated trajectory; and it has demonstrated satisfactory results on various data. Despite its advantages, the algorithm has drawbacks, including a lack of formalization, limited sensitivity analysis, convergence issues, impact of the choice of the reference trajectory, explicit validation of outlier robustness, and lack of open source code availability.

Thus, the goal of this research article is to address these shortcomings. We propose a modular iterative aggregation algorithm able to reconstruct accurately the common path followed by all the individual sample trajectories (*i.e.* accurate aggregate trajectories). Our work expands the research of [4] and introduces the contributions: (1) we formalize the properties of the existing algorithm and propose a modular aggregation algorithm based on different components. This modularity extension allows to enhance the efficiency, scalability, and usability of the algorithm according to user requirements; (2) we propose a comprehensive framework to study the sensibility and metrological performance of the algorithm to provide valuable insights into the performance and behavior of the aggregation algorithm under various conditions including noise; (3) we implement the algorithm into an open source Python library and share the used data, thereby enhancing accessibility and facilitating reproducibility of the research findings.

## 2 Data and materials

### 2.1 Synthetic GNSS trajectories

To overcome to the lack of ground truth data to compare the estimated aggregated trajectory with the real route actually followed by the individual, it was first decided to proceed to extensive experimentation of the algorithm on simulated GNSS trajectories. These simulated trajectories provide a controlled environment to test the algorithm's performance under various predefined conditions.

The methodology used is as follows: for each case study, a reference track is synthetically created and 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 used [13] require an accurate modeling of auto-correlation error of GNSS trajectories, to avoid unrealistic simulations and topological errors [12].

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 (Cf. Figure 1(a)) such as long wave-length process describing coordinate system errors, an intermediate wave-length process describing GNSS observation errors (auto-correlated in space and time), 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 [9]: 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  $\Sigma$ , is a realization of a correlated random vector  $\mathbf{Y}$  having covariance matrix  $\Sigma$ . The covariance matrix  $\Sigma$  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 [5], completed by a 1 m white noise process, a 50 cm range Gaussian Process for referencement error. We also analyze trajectory aggregation for three commonly mountain path shapes: nearly straight, moderately sinuous, and switchbacks. These paths were randomly generated each with a consistent length of about 300 meters.

## 2.2 Multi-sensors traces acquisition

To assess the impact of the sensors on precision accuracy in dense forest, we delineated five types of sensors (i.e. mobile phone equipped with VisoRando application, Polar GPS device, Garmin GPS device, Keymaze device and professional Ublox GPS sensor chip). Second, 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, and collect five round trips following the route exactly. Third, the field work was done by two of the authors of this paper. Data collection has been carried out during the summer season. In total, for the five sensors, 50 trajectories are collected (Cf. Figure 1(b)).

We identified a potential bias in the data acquisition process. All GNSS traces were collected within the same time slot, which implies that in addition to their individual noise, they share a common positional bias due to the relatively unchanged satellite configurations during the experiment.

The ground truth route was obtained via a topometric survey conducted by students under the supervision of the authors and teachers [3] (Cf. Figure 1(b)). Absolute positioning used GNSS differential static positioning of reference points 300-500 meters from the

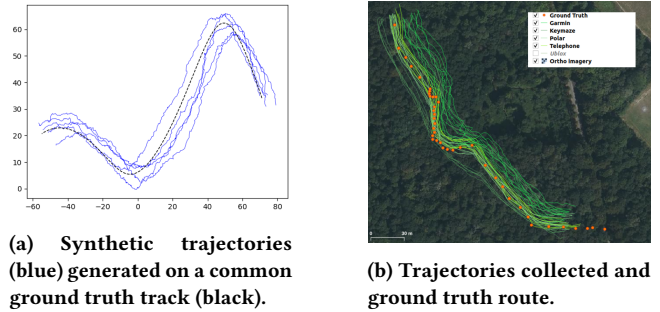


Figure 1: Types of data taken into account

route, while the route geometry was determined using surveying traverse. The output ground truth is sampled with about 42 points (i.e. 1 point every 7 meters) with an absolute positioning accuracy of 5 mm in each 3D axis ( $1\sigma$ ).

## 3 Modular and Iterative Aggregation Algorithm

### 3.1 Formalization of the original algorithm

The algorithm proposed in [4] uses an iterative refinement approach that improve the existing solution at each step, continuously performing multiple matches.

Considering a set of trajectories, where each trajectory is defined as an  $n$  ordered points:  $\mathcal{X} = (\mathbf{x}_i)_{i=1..n}$ , with  $\mathbf{x}_i \in \mathbb{R}^2$  and GPS records in a 2D space. Let  $d$  be a distance (Euclidian, Manhattan, etc.) between those points:  $d : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}_+$ . Note that heights and timestamps for example, could also be considered with  $\mathbf{x}_i, \mathbf{y}_i \in \mathbb{R}^3$  or  $\mathbb{R}^4$ , and defining  $d$  accordingly to measure distance between three-dimensional timestamped records.

An aggregated trajectory, noted  $\mathcal{AX}$ , is defined as the best geometric representation of a set of trajectories  $\mathcal{X}$  following exactly the same route defined from an origin to a destination:  $\mathcal{AX} = (\overline{\mathbf{x}}_j)$ ,  $j=1..m$ , where  $\overline{\mathbf{x}}_j$  represents the aggregated points of matched points for the  $j$ -th point in the master trajectory.

We defined an accurate aggregated trajectory, noted  $\mathcal{AAX}$ , an aggregated trajectory that optimizes a quality criterion  $Q$ , with respect to the (unknown) ground truth  $G$ . More formally,  $\mathcal{AAX}$  optimizes  $E[Q]$ , where  $E$  is the expected value of the quantity  $Q$ , which is our focus. A trace  $T_1$  is a good partial representation of  $T_2$  if  $Q_{T_2 \rightarrow T_1}$  is minimal with  $Q$  being the square root of the mean of the squared distances between each point in  $T_1$  and its closest neighbor on  $T_2$ . Thus, the quality of the trace  $\mathcal{AAX}$  can then be evaluated from the average of the partial qualities:

$\mathcal{AAX} = (Q_{AccAX \rightarrow G} + Q_{G \rightarrow AccAX}) / 2$ , with  $G$  the ground truth.

Algorithm **initialization** consists of choosing a first trajectory, called the master trajectory,  $\mathcal{R} = (\mathbf{r}_j)_{j=1..m}$ . This initialization avoids to match all the trajectories in pairs.

**First step: trajectory matching.** The points of each trajectory are ordered matching with the master trajectory (each point in  $\mathcal{X}$  is linked to at least one points in  $\mathcal{R}$ , and reciprocally and there should be no pair of *crossing* links). Matching algorithm uses  $L_p$ -norm optimal dynamic time warping matching (DTW) is used to calculate ordered matching. The discrete Fréchet distance used by

[4] is not identical to discrete DTW: the last given in the formula above degenerates to Fréchet distance when  $p$  grows to the infinity.

**Second step: representative selection on trajectory sections.**

At the end of matching process, positions of trajectory  $\mathcal{X}$  can be linked to many points of  $\mathcal{R}$  and reciprocally and then create an influence. Thus, the second step of algorithm iteration is to choose for each multiple matching links previously established, a representative position of each group of vertices from the trajectory  $\mathcal{X}$ :  $\overline{x_j}$ . [4], take the center of gravity of the segment.

**Third step: aggregation of representative points.** At this point, all positions of the master trajectory  $(r_j)_{j=1..m}$  have a unique homologous point  $\overline{x_j}$  from each trajectory  $\mathcal{X}$ . The positions of the new merged trajectory are calculated from the median (which is better suited to handle outliers) of matching points. [4] add a constraint in this step of the algorithm: each position aggregated must also be part of existing trajectory positions, so as not to be located in an unlikely place. At last, the aggregated trajectory is used as the new master trajectory. The algorithm **stops** when the distance between two subsequent estimation of the aggregated trajectory is below a predefined threshold. The distance used is the pointwise  $L_2$  distance.

### 3.2 Enhancing Algorithm Modularity

Understanding the variety of similarity, aggregation, and selection methods, we aim to transform the algorithm proposed in [4] into a new modular that we name Modular and Iterative Aggregation Algorithm (MIAA). For that, we transform each step into a component with multiple options. Modularity has substantial advantages such as flexibility scalability, and more relevant to study the behavior and the influence of measures and parameters in different contexts and with different data.

[C1]: *choosing the master trajectory.* The option, proposed in [4], to get the master trajectory, selects the trajectory whose length is closest to the median of the lengths of all trajectories to be aggregated. We add, two new options: (1) trajectory that minimizes the sum of distances to other trajectories, and (2) a random trajectory from the set of trajectories to be aggregated.

[C2]: *matching trajectories with the master trajectory.* This component contains measures to compute the distance between two trajectories. We considered four measures: matching parameterized with  $L_p$ -norm ( $p \in 1, 2, \infty$ ), and the nearest neighbour matching.

[C3]: *choosing the representative position of each homologous points.* It allows to choose the method to aggregate matched points. We implemented three methods: center of gravity, position with median time, position furthest from the master trajectory.

[C4]: *aggregating the representative position:* We propose four aggregate functions: marginal median ([4]), geometric median,  $L_2$  mean and the  $L_\infty$  which is the center of the minimum covering circle.

## 4 Metrological performance analysis

All implementations have been integrated into the *Tracklib* library enabling GNSS trajectory computation [7].

### 4.1 Algorithm Calibration

**First baseline: with anchor constraint**

To compare the MIAA variants, we analyze the aggregated trajectories from the baseline algorithm of [4] along each variant of the component individually. Our baseline is tested across three path shapes (almost straight, moderate sinuosity, and switchbacks) by generating  $N$  random noisy traces to be aggregated with  $N \in 2, 3, 4, 5, 6, 7, 10, 25, 50$ . Each scenario is replicated 10 times, and the results are averaged. For validation, we compute the distance pointwise  $L_2$  (RMSE) by finely resampling both the aggregated and ground truth trajectories from 120 to 1,000 points, ensuring one point every 30 cm over 300 meters. The calculations were executed on seven computers (CPU Intel® Xeon(R) W-2223 CPU@3.60GHz x 8 and 32Go memory). Figure 2(a) illustrates the variation in RMSE based on the sample size of trajectories, with one graph per shapes of path, one curve per options of component C4 (aggregation of the representative positions). Notably, anchoring the new aggregated position to an existing trajectory position has a quite unfavorable effect compared to not anchoring it. The curves with no anchor constraint (NA) show a lower RMSE (the purple, light green, and cyan curves), while the curves with anchor constraint (WA) show a higher RMSE. Based on these results, we propose a second baseline by adjusting the parameters to remove the anchoring constraint.

**Second baseline: without anchor constraint**

We still compare the aggregated trajectories obtained with the algorithm [4], but without the constraint of anchoring the aggregated points to trajectory existing points. Otherwise, as before, only one other variant differs in the set of approaches for each MIAA algorithm component. It should be noted that in Figures 2(b) and 2(c), the curves are quite close together and their profiles are quite similar; so, the conclusions will not be definitive. First, as shown in Figure 2(b), we notice that the choice of master trajectory has no influence on the quality of the final aggregated trajectory whatever the sample size of trajectories. Second, trajectory matching using the discrete Fréchet distance with a large sample size of trajectories (greater than 10) converges faster (see Figure 2(c)). A precision of 2 meters is achieved between the aggregated and the ground truth trajectory with at least 20 trajectories in the sample. The curve follows a square root shape, indicating that 50 trajectories are needed for 1 meter precision. To conclude, for each scenario and each types of route, options of the algorithm by [4] performs well.

### 4.2 Speed of convergence

Besides, regarding the use of real data (Multi-Sensors trajectories), we have analyzed the algorithm speed of convergence to find out how many sample traces are required for the aggregated trajectory to converge to the ground truth. We compute, on the experimental trajectories, the aggregated trajectory with the DTW- $L_2$  matching distance. We choose for sample size  $N \in 2, 3, 4, 5, 8, 10, 15$ . Due to the bias of data (mentioned in Section 2.2), the geometry is preceded by a translation of 5m in x and 3m in y, which provides insight into the algorithm's speed of convergence (Cf. Figure 2(d)).

Due to this bias, we consider only split the quality criterion  $Q$  only in shape component and the quality of trace  $\mathcal{A}\mathcal{X}$  is evaluated with the distance to nearest neighbour. The graph curve (Cf. Figure 2(e)) is noisy but clearly shows convergence: the error decreases from 7 meters (without aggregation) to around 3 meters with approximately fifteen traces.

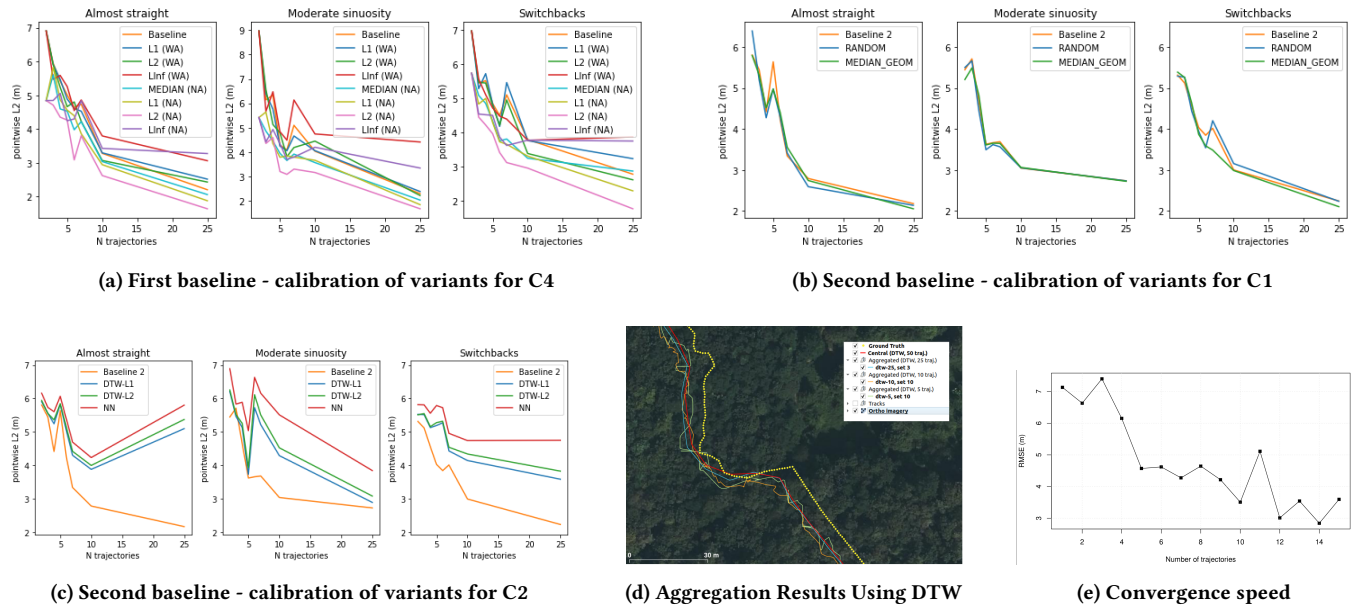


Figure 2: Algorithm's performance from a metrological perspective

Further details on the data and results are described in [11].

## 5 Conclusion

We expanded upon the iterative aggregation trajectories algorithm proposed by [4] by introducing a modular framework based on four components which allows analyzing its ability to reconstruct an accurate aggregated trajectory and easily adjusting measures and parameters based on the data and objectives. Additionally, a module for generating realistic trajectories and a protocol for collecting experimental and ground truth data are implemented in open source. We also conducted a thorough analysis by comparing the results obtained with the original algorithm [4] with various variants. A future direction in this regard is to examine the sensitivity of trajectories to different types of noise.

Finally, our next step is to use the MIAA algorithm to compute a route network from the different existing data taking into account temporal dimension (*i.e.* trajectories from a given period) or semantic (*i.e.* engaging in specific activities: walking, snowshoeing, biking). The particularity of the network is that it must summarize spatially and as fine and exhaustively as possible the activity of human seasonal practices and be able to enrich him with contextual, semantic and temporal information. Indeed, determining the granularity of the network poses a significant challenge. Whether to define the granularity of the network as part of the aggregation algorithm during the aggregation process or as a post-processing step remains an open question.

## Acknowledgments

This work was supported by the ANR under grant agreement no. ANR-23-CE55-0003. The authors would like to thank Professors Franck Tertre and Emmanuel Clédât for supervising the students'

project, which helped establish the ground truth for the Multi-Sensor Trajectories dataset. The authors thank Professors Franck Tertre and Emmanuel Clédât for supervising the student project that established the ground truth for the Multi-Sensor Trajectories dataset and the reviewers.

## References

- [1] M. Brankovic and K. Buchin et al. (k, l)-medians clustering of trajectories using continuous dynamic time warping. In *Proceed. of the 28th Int. Conference on Advances in GIS, SIGSPATIAL '20*, page 99–110, 2020.
- [2] K. Buchin and M. Buchin et al. Median Trajectories. *Algorithmica*, 66(3):595–614, July 2013.
- [3] P. Calloch, B. Labbé, and P. Lorine. Établissement d'une méthode low-cost de vérité terrain pour l'étude de la précision de traces gns de randonneurs en forêt. Technical report, 2024.
- [4] L. Etienne and T. Devoegele. Trajectoires médianes. In *14<sup>ème</sup> 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, Jan. 2014.
- [5] D. Grejner-Brzezinska, C. Toth, and Y. Yi. On improving navigation accuracy of gps/ins systems. *Photogrammetric engineering & remote sensing*, 71(4):377–389, 2005.
- [6] N. Le Corre and A. Saint-Pierre et al. Outdoor recreation in french coastal and marine protected areas. exploring recreation experience preference as a way for building conservation support. *Journal of Outdoor Recreation and Tourism*, 33:100332, 2021.
- [7] Y. Méneroux and M.-D. Van Damme. Tracklib: a python library with a variety of tools, operators and functions to manipulate GPS trajectories, Dec. 2022.
- [8] Reinoso, Moncayo, and Ariza-López. A new iterative algorithm for creating a mean 3d axis of a road from a set of gns traces. *Mathematics and Computers in Simulation*, 118:310–319, 2015.
- [9] B. D. Ripley. *Stochastic simulation*. John Wiley & Sons, 2009.
- [10] J. M. Rubio-Iglesias, T. Edovald, R. Grew, T. Kark, A. E. Kideys, T. Peltola, and H. Volten. Citizen science and environmental protection agencies: Engaging citizens to address key environmental challenges. *Frontiers in Climate*, 2, 2020.
- [11] M.-D. Van Damme, Y. Méneroux, and A.-M. Olteanu-Raimond. An extensive analysis and calibration of the modular aggregation algorithm across three categories of for gns trajectories data sources. Technical report, 2024.
- [12] F. Vauglin. *Modèles statistiques des imprécisions géométriques des objets géographiques linéaires*. PhD thesis, Université de Marne-la-Vallée (1991-2019), 1997.
- [13] Y. Zhang and Y. Yang. Cross-validation for selecting a model selection procedure. *Journal of Econometrics*, 187(1):95–112, 2015.