

SCIENTIFIC DATA

OPEN

Dataset of three-dimensional traces of roads

DATA DESCRIPTOR

Francisco Javier Ariza-López , Antonio Tomás Mozas-Calvache ,
Manuel Antonio Ureña-Cámara  & Paula Gil de la Vega

Received: 7 March 2019

Accepted: 11 July 2019

Published online: 05 August 2019

We present a dataset consisting of three-dimensional traces, captured by Global Navigation Satellite System techniques with three-dimensional coordinates. It offers 138 traces (69 going and 69 returning), in addition to the actual mean axis of the road determined by precise surveying techniques to be used as ground truth for research activities. These data may serve as a test bed for research on data mining applications related to Global Navigation Satellite System multitraces, particularly the development and testing of algorithms intended for mining mean axis data from road multitraces. The data are suitable for the statistical analysis of both single-trace and multitrace datasets (e.g., outliers and biases).

Background & Summary

Roads are an important component of national infrastructure and are traditionally represented in maps. Currently, road networks represent a key element of national and regional spatial databases, for example, Euroglobalmap¹ or the transport network theme of Inspire². Additionally, there are many applications for cars (e.g., TomTomTM and HereTM) and for mobile devices and desktop computers (e.g., Google Maps[®] and OpenStreetMaps[®]) that offer different services (e.g., tracking, routing, and navigating) based on road networks.

The automatic generation and updating of road networks has received considerable research attention^{3–8}. In this automatic environment, two processes are of great importance: the mining of all available data to derive road sections or networks^{4,9–14} and the assessment of the positional accuracy of the results using line-based methods^{15–23}.

Road networks are recorded in digital spatial databases as node-edge structures, where each edge represents the axis of a road section. In the majority of cases, the edge geometry has only two-dimensional (2D) coordinates and is supported by a string of lines (polygonal), which represent the simplest geometric primitive for recording the axes of linear features. From a conventional perspective, this geometry can be generated with different technologies, for example, by a topographic or Global Navigation Satellite System (GNSS) survey or by photogrammetric or mobile mapping techniques. Data mining techniques, however, can be used at present to obtain these road axes. For example, TomTomTM and HereTM have communities of users that are allowed to upload their actual GNSS traces, and after a mining process, these traces are used to update the road database; this is known as information of communities (IC). Another possibility is the use of information known as volunteered geographical information (VGI), which is created or collected by volunteers²⁴. VGI allows the generation and maintenance of bulky sets of geographic information in the form²⁵ of points of interest, comments, images (photographs), navigation traces, etc. IC and VGI are similar because many contributors with a substantial number of different configurations (e.g., the branch and type of GNSS devices) are possible. Nevertheless, when IC comes from a community organized by users of a specific trademark, VGI is considered to have more variability. In the case of VGI, several GNSS trails can be downloaded from the same path on some platforms, e.g., Wikiloc (<https://www.wikiloc.com/>) and Wikirutas (<http://www.wikirutas.es/>). Nevertheless, VGI regarding paths presents several problems²⁶ when trying to use it to determine path axes.

Given a dataset of multiple traces coming from IC or VGI sources, the derivation of road axes that conform to the road network structure represents a complex data mining procedure^{4,7,11–13,27}. In addition, another frequent problem is the absence of ground truth with which to evaluate the geometric quality and positional accuracy of the results.

The 3D GNSS-road trace dataset that we present is the first database with multiple actual road axes captured by GNSS techniques that have three-dimensional coordinates (3D). This dataset is not VGI or IC because it has

Department of Cartographic, Geodetic Engineering and Photogrammetry, University of Jaén, Jaén, Spain. Correspondence and requests for materials should be addressed to F.J.A.-L. (email: fjariza@ujaen.es)



Fig. 1 Example illustrating multitracess. Each coloured line represents a different trace. Basemap: Orthophoto from Instituto Geográfico Nacional.

been created under a controlled design within a research project. This dataset offers 138 traces (69 going and 69 returning). The trajectories were surveyed on a set of roads that define a circular circuit with high altimetric differences, slopes and sharp curves. The actual mean axis of the road, which was determined by precise survey techniques, is supplied to be used as ground truth for research activities. Statistics about the multitracess and the axis dataset are included.

This 3D GNSS-road trace dataset could be of great help to researchers in data mining related to GNSS multitracess (e.g., algorithm development). The main application of this dataset is the development and testing of algorithms intended for mining mean axis data from road multitracess. Additionally, the dataset is suitable for the statistical analysis of single-trace and multitrace datasets, including the determination of outliers, biases and so on. The inclusion of the actual mean axis of the road determined by precise techniques allows the development of quality controls for the results. This dataset facilitates the abovementioned work because it is not necessary to invest time and money to obtain such expensive data. In addition, this dataset will facilitate comparisons with future studies and their results, which is considered extremely important for the advancement of research in this field.

Methods

In this section, we provide a definition of the data, an explanation of the design and a description of the production methods.

Multitracess and control axis. A trace is a recorded GNSS path, and a multitrace set is a set of such recorded paths, as can be observed in Fig. 1.

We must define some terms to provide a better explanation of the content within the 3D GNSS-road trace dataset. The terms to be defined are GNSS trace, multitrace, axis and control axis.

Given a geographical phenomenon such as the centreline of a road, railway, shoreline, border, river or stream, a line string L or polygonal representation is determined by a set $\{P_1, \dots, P_n\}$ of n ordered points (vertexes) that determine an ordered set S of $n - 1$ segments ($S_{i,i+1}$) formed by two consecutive vertexes (P_i, P_{i+1}) (see Fig. 2). Each vertex is represented by a 3D point with coordinates $\{X, Y, Z\}$ in a specific coordinate reference system. In an analytical way:

$$\begin{aligned}
 L &= \{S_{0,1}, S_{1,2}, \dots, S_{n-2,n-1}\} \\
 S_{i,i+1} &= \{P_i, P_{i+1}\} \\
 P_i &= \{X_i, Y_i, Z_i\}, \quad i \in [0, 1, \dots, n - 1] \\
 P_j &= \{X_j, Y_j, Z_j\}, \quad j = i + 1
 \end{aligned}$$

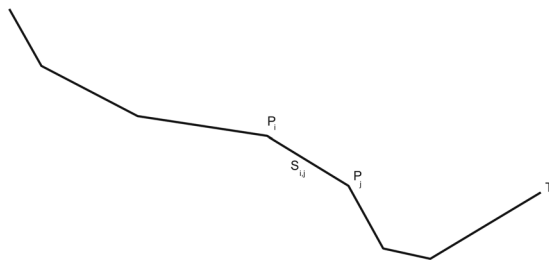


Fig. 2 Definition of a line string L .

Section	Length (km)	Mean slope
N-323	3.5	4.3%/–3.3%
JV-3231	2.8	6.4%/–4.1%
JV-2227	5.9	6.0%/–8.8%
All	12.2	6.2%/–5.7%

Table 1. Definition of the linear feature.

where:

n : number of vertexes.

If a line string L represents the path through a road section of a vehicle (e.g., a car or a motorcycle) captured by a GNSS device, L is called a GNSS trace (simply called a trace) and is denoted by T . If several traces T are available for the same road section, a multitrace MT set is established.

If a line string L represents the mean axis of a phenomenon (e.g., the central axis of a paved road), L is called the axis and is denoted by A . If an axis A is determined in such an accurate way (i.e., at least three times more accurate than an MT set), this axis can be used as a control axis for the MT set and is denoted CA_{MT} .

Design. The purpose of this dataset is two-fold; on the one hand, the dataset exists to offer a set of actual GNSS multitraces to work with and to examine all the problems corresponding to their use (omissions, outliers, bias, axis mining, etc.); on the other hand, the dataset is intended to offer an accurate mean axis for a road that can be used to control the result of mining an estimated axis from the multitrace dataset. The design covers both the area and device selection and the data processing method.

Area selection and description. To offer an MT dataset characterized by sufficient complexity and variability, an appropriate study area was sought. The criteria for the selection of the study area were as follows:

- Proximity to the continuously operating GNSS reference station at the University of Jaén to obtain differential corrections for the precise GNSS survey of the accurate axis.
- Circular travel route and design to facilitate the logistics of the survey.
- Roads with little traffic to facilitate field work.
- Roads exhibiting a considerably variable slope with areas of both curved and straight segments.

Finally, a circular path was found on the outskirts of two small villages (Cárcchel and Carchelejo) near Jaén (Spain). The total length of the path is 12.2 km, and it has a mean slope of 6%. This path is composed of three different road sections from primary and tertiary roads (Table 1). Figure 3 shows a general view of the area, and Fig. 4 presents a profile of the circular path.

Multitrace production. The MT dataset was captured using a Columbus V990 device (http://cbgps.com/v990/index_en.htm), which is a GNSS data logger that allows the use of a memory card to record a vast number of points and is designed for navigation and in-car applications. The accuracy specifications of the device indicate a 5.0 m circle of error probable (CEP) (95%) for non-differential GNSS applications using the Global Positioning System (GPS) and a 2.5 m CEP (95%) for differential GNSS applications using either EGNOS (European Geostationary Navigation Overlay Service) or WASS (Wide Area Augmentation System) as supplementary systems. However, the device is able to report only on the dilution of precision (DOP) of each point (no precision is offered). The device was placed on the front dashboard of a car, and the survey was performed using the non-differential configuration.

The car was driven in a normal way while taking into account all traffic signals and road conditions (e.g., slopes). A total of 69 traces were obtained in each direction (going and returning). Each trace was assigned an identifier and a label indicating whether it was a going trace or a returning trace.

Control axis production. The centreline of a road section is the objective sought through the MT sets; for this goal, we need ground truth to assess the positional accuracy. This ground truth is what we call the control

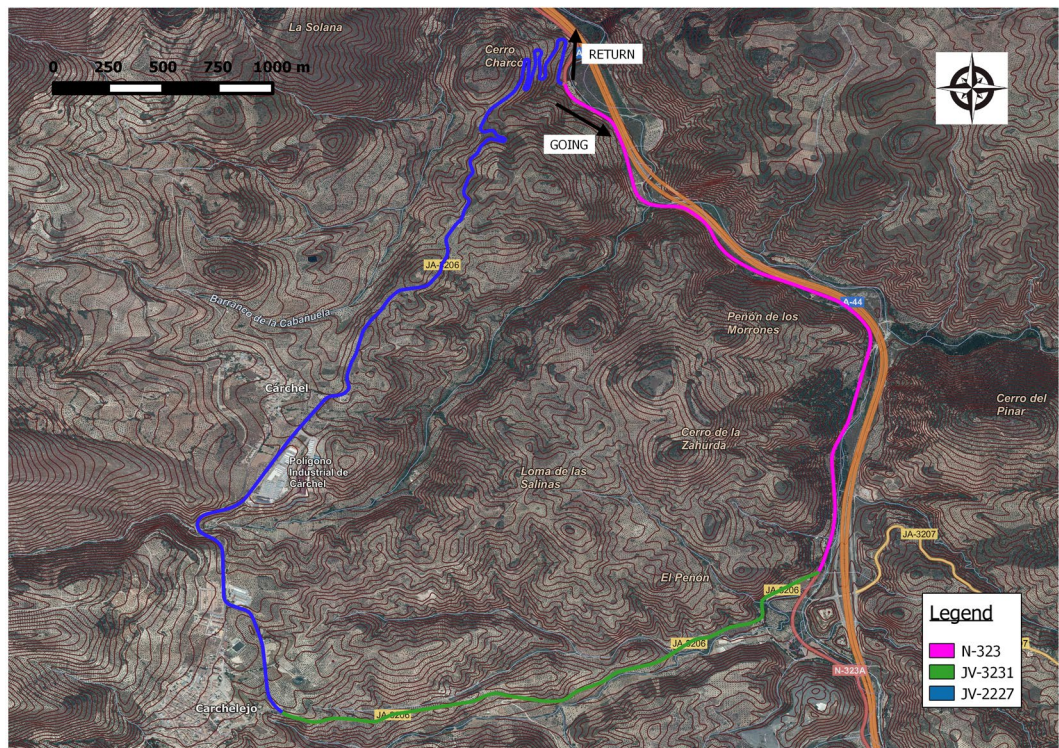


Fig. 3 Round trip, topography and roads.

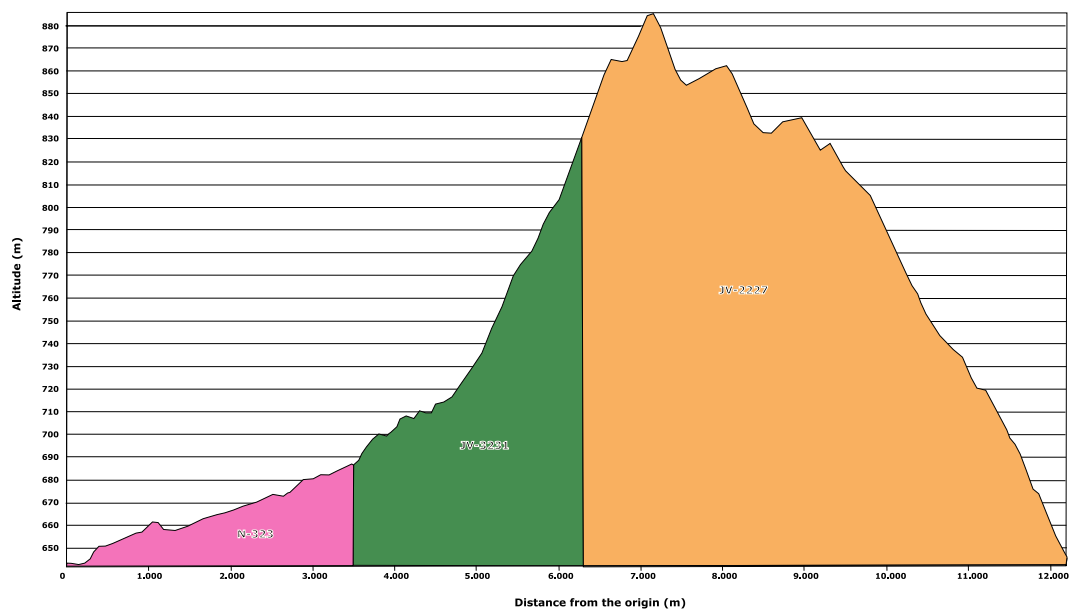


Fig. 4 Altimetric profile.

axis. The production of the control axis involves two main steps: first, a precise differential GNSS survey is conducted following the white roadside lines; then, the mean axis to be used as the ground truth (control axis) is calculated.

The GNSS survey was performed using a Leica 1200, which has a horizontal accuracy of 10 mm + 1 ppm and a vertical accuracy of 20 mm + 1 ppm for kinematic surveying²⁸. Post-processing was carried out using the corrections provided by the continuously operating GNSS reference station at the University of Jaén. This survey was executed on foot using the device²⁹ shown in Fig. 5 that allowed the GNSS antenna to be kept vertical. The post-processed points had a planimetric precision better than 40 mm (1 sigma) and an altimetric precision better than 50 mm (1 sigma).



Fig. 5 Precise GNSS survey using the Leica 1200 system.

Field	Data Type	Description
3DGRT_ID	Integer	Unique identifier.
3DGRT_IDT	Integer	Trace unique identifier.
3DGRT_DI	Character	Direction: G = going, R = return.
3DGRT_DE	Character String	Device: Columbus_V990
3DGRT_SO	Integer	Sequential order of point.
3DGRT_LAT	Double	Latitude: decimal degrees. CRS = EPSG4979
3DGRT_LON	Double	Longitude: decimal degrees. CRS = EPSG4979
3DGRT_h	Double	Ellipsoidal height: metres. CRS = EPSG4979
3DGRT_HDOP	Double	Horizontal dilution of precision.
3DGRT_VDOP	Double	Vertical dilution of precision.

Table 2. Attributes of traces in the multitrace dataset (EPSG: 4979).

Field	Data Type	Description
3DGRT_ID	Integer	Unique identifier.
3DGRT_IDT	Integer	Trace unique identifier.
3DGRT_DI	Character	Direction: G = Going, R = Return.
3DGRT_DE	Character String	Device: Columbus_V990
3DGRT_SO	Integer	Sequential order of point.
3DGRT_X	Double	Easting UTM: metres. CRS = EPSG25830
3DGRT_Y	Double	Northing UTM: metres. CRS = EPSG25830
3DGRT_h	Double	Ellipsoidal height: metres. CRS = EPSG25830
3DGRT_HDOP	Double	Horizontal dilution of precision.
3DGRT_VDOP	Double	Vertical dilution of precision.

Table 3. Attributes of traces in the multitrace dataset (EPSG: 25830).

Once both white roadside lines had been obtained, the mean road axis was derived using the Fréchet distance³⁰ (see Gil de la Vega³¹ for more details). After this, a visual inspection of the mean axis was performed to check for the inexistence of artefacts. Figure 6 shows an example of the two white roadside lines (in blue) and the control axis (mean axis) derived (in red). A precision value was obtained for each point along the mean axis by composing the precision of the post-processed GNSS data from the survey of the two white roadside lines. The points selected for each composition were determined through an interpolation process.



Fig. 6 The results of the survey of the two white roadside lines (in blue) and the mean axis or control axis (in red).

Field	Data Type	Description
3DGRT_ID	Integer	Unique identifier.
3DGRT_DI	Character	Direction: A = Axis
3DGRT_SO	Integer	Sequential order of point.
3DGRT_LAT	Double	Latitude: decimal degrees. CRS = EPSG4979
3DGRT_LON	Double	Longitude: decimal degrees. CRS = EPSG4979
3DGRT_h	Double	Ellipsoidal height: metres. CRS = EPSG4979
3DGRT_QP	Double	Quality of planimetric position: root mean square error (RMSE) in metres
3DGRT_Qh	Double	Quality of ellipsoidal height: root mean square error (RMSE) in metres
3DGRT_QPh	Double	Quality of 3D position: composition of 3DGRT_QP and 3DGRT_Qh in metres.

Table 4. Attributes of the control axis (EPSG: 4979).

Data Records

The 3D GNSS-road trace dataset is composed of three files: two contain the spatial data, while the third file contains the metadata. The spatial data files are provided in SHP (shapefile) format³². The data are offered in two different coordinate reference systems (LatLon + UTM projected) to facilitate their use in research applications (the coordinate reference systems are EPSG4979 and EPSG25830, respectively). The ISO 19115-1 metadata are provided following the ISO 19115-1 standard³³ and include the purpose, lineage and usage, in addition to many other technical characteristics. The dataset is available from Figshare³⁴ as a compressed (.zip) file that contains the geospatial data in shapefile format. The fields of the SHP files and their explanations are provided in Tables 2 through 5.

To describe some characteristics of the dataset, some statistical information about both the *MT* set and the control axes can be seen in Table 6. Please note that the *MT* data are raw data, and for this reason, values such as the minimum distance are 0 while the slope variation rises to 1136%. These artefacts in the *MT* data have not been removed for the reason indicated in the Introduction and will be noted in the Technical Validation.

Technical Validation

This section presents relevant information on the 3D GNSS-road trace *MT* dataset and the control axis data and processing method to assure the readability and quality of the dataset. Following the quality description framework of the ISO 19157 standard³⁵, there are five categories of quality elements employed to describe the quality of geospatial data: completeness, positional accuracy, thematic accuracy, temporal quality, and logical consistency. However, neither the *MT* dataset nor the control axis data can be evaluated with respect to temporal quality because the complete dataset was surveyed over a very short period of time, and thus, no temporal quality aspect

Field	Data Type	Description
3DGRT_ID	Integer	Unique identifier.
3DGRT_DI	Character	Direction: A = Axis
3DGRT_SO	Integer	Sequential order of point.
3DGRT_X	Double	Easting UTM: metres. CRS = EPSG25830
3DGRT_Y	Double	Northing UTM: metres. CRS = EPSG25830
3DGRT_h	Double	Ellipsoidal height: metres. CRS = EPSG25830
3DGRT_QP	Double	Quality of planimetric position: root mean square error (RMSE) in metres
3DGRT_Qh	Double	Quality of ellipsoidal height: root mean square error (RMSE) in metres
3DGRT_QPh	Double	Quality of 3D position: composition of 3DGRT_QP and 3DGRT_Qh in metres.

Table 5. Attributes of the control axis (EPSG: 25830).

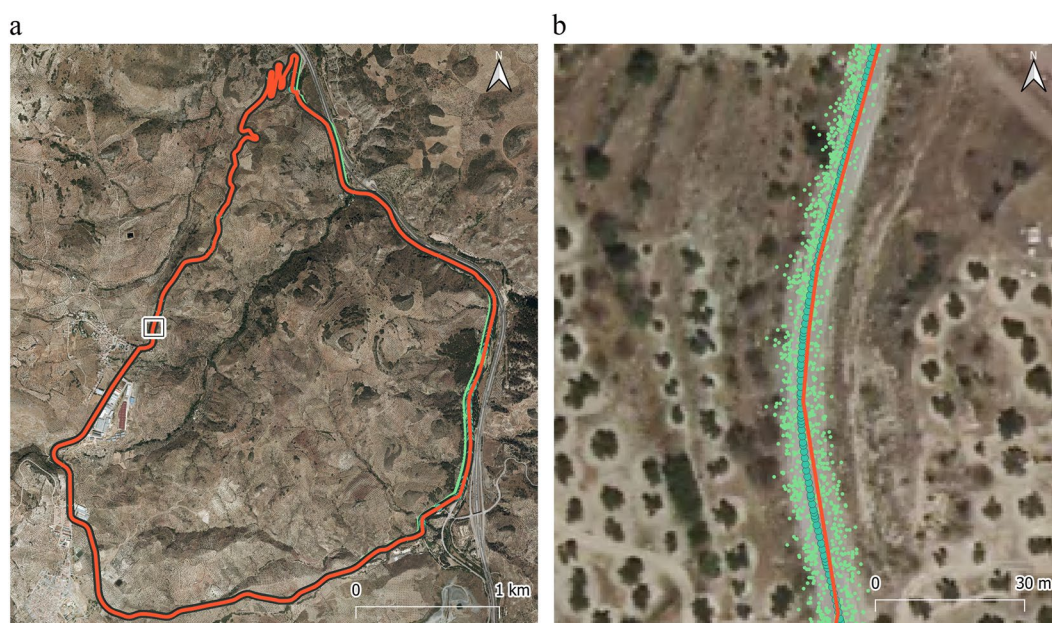


Fig. 7 Visual test among the control axis (green dots with solid line), multitrace points (small green points) and the axis that exists on the official cartography of the IGN (red line). **(a)** General view; **(b)** detailed view. Basemap: Orthophoto from IGN (2014).

is involved or relevant to the whole dataset. However, we can provide a discussion about the other four quality elements for the *MT* dataset and for the control axis:

1. **Completeness.** Raw GNSS data points from the data logger have been provided. These data are offered without processing to ensure the presence of all possible artefacts that may occur in any survey. This is of great interest for future research (e.g., the treatment of outliers and errors). However, the *MT* points were inspected visually, as can be seen from the general view in Fig. 7a and the detailed view in Fig. 7b. In addition, the control axis has been visually inspected by three operators and is complete (Fig. 8).
2. **Positional accuracy.** The accuracy of the raw GNSS data points has been provided as registered by the data logger. With respect to the control axis, the surveying device and processing method guarantee a positional accuracy that is at least three times greater than the positional accuracy of raw GNSS data; hence, the raw data can be used as a control axis. Moreover, the per point accuracy has been provided as the RMSE compositions of the points used to define each vertex of the mean axis. In addition, a visual control has been executed, and a complete comparison of the mean axis with an independent official source (at a 1:25000 scale) (Fig. 8) clearly shows that the provided control axis boasts a better performance.
3. **Thematic accuracy.** The only thematic attribute that appears in this dataset is the 3DGRT_DI field. The assignment of all the values has been reviewed and found to be correct in 100% of the cases.
4. **Logical consistency.** Among all the logical consistency aspects of geospatial data, only the format consistency can be checked in this dataset. To test the Shapefile format, all files were loaded using ArcMap from ESRI™ and GDAL (<http://gdal.org>). Both pieces of software loaded the Shapefiles, including the height values from the geometry, thereby validating the file structure and compatibility. However, we have included the geometry tuple in the attributes of each feature to ensure the ability to analyse this third coordinate.

	MT	Control Axis
Number of points	111113	11304
2D distance between points [m] (min/mean/max/stdev)	0/15.05/221.32/4.63	0.01/1.07/112.14/1.64
3D distance between points [m] (min/mean/max/stdev)	0/15.07/221.32/4.62	0.01/1.08/112.14/1.64
Mean velocity [km/h]	55	—
Altimetric statistic [m]	581.02/747.62/983.02/74.51	642.95/759.95/884.90/76.35
Turning value [°]	−166.95/0.18/160.03/8.86	−59.18/−0.02/54.54/3.46
Absolute slope [%]	0.00/4.58/1136.36/6.97	0.00/4.96/40.70/3.41
HDOP (provided by GNSS) (min/mean/max/stdev)	1.1/2.1/10/0.7	—
VDOP (provided by GNSS) (min/mean/max/stdev)	0.8/1.3/10.0/0.4	—
Calculated planimetric RMSE [m] (min/mean/max/stdev)	—	0/0.02/0.73/0.05
Calculated altimetric RMSE [m] (min/mean/max/stdev)	—	0/0.02/1.16/0.10
Calculated 3D RMSE [m] (min/mean/max/stdev)	—	0/0.04/1.36/0.11

Table 6. Statistics of the dataset.



Fig. 8 Details of the differences between the control axis (green dots with solid line) and the axis that exists on the official cartography of the IGN (red line) in a sinuous section. Basemap: Orthophoto from IGN (2014).

Usage Notes

The dataset is distributed as Shapefiles that contain the data organized as described in Tables 2 through 5. Shapefiles are the standard for exchanging and storing spatial data. Many Geographic Information System tools (e.g., ArcGIS®, QGIS®, and GRASS®) are able to load such files. The use of scripting languages (e.g., Python) within these software tools can be of great help for the processing of multitrace datasets because there are no standard capabilities for such types of data. The GDAL library for R (<https://www.R-project.org/>) allows access to the data and the use of many other R packages with powerful capabilities for dealing with trace and multitrace datasets.

The 3D GNSS-road trace dataset is available for free use/reuse. There are no restrictions to support the widest possible use. The main application of this dataset is the development and testing of algorithms intended for mining mean axis data from road multitraces^{4,7,11–13,27,36}. Multitraces are provided as raw data, and thus, this dataset is suitable for the statistical analysis of both single-trace and multitrace datasets, including the determination of outliers and biases and the development and testing of filtering algorithms focused on problems pertaining to traces^{31,37}. Finally, the number of traces allows the use of simulation procedures (Monte Carlo, Bootstrap, etc.) to derive and estimate the distributions of some characteristics; for instance, these simulation techniques can be applied to determine the adequate sample size.

References

1. Eurogeographics. *EuroGlobalMap*, <https://eurogeographics.org/products-and-services/open-data/> (European National Mapping, Cadastral and Land Registry Authorities Association, 2018).
2. Inspire. *D2.8.1.7 Data Specification on Transport Networks – Technical Guidelines*. (INSPIRE Thematic Working Group *Transport Networks*, Joint Research Centre, 2014).
3. Rogers, S., Langley, P. & Wilson, C. Mining GPS data to augment road models. In *Proceedings of the fifth ACM SIGKDD International Conference on Knowledge discovery and data mining* 104–113, <https://doi.org/10.1145/312129.312208> (1999).
4. Edelkamp, S. & Schrödl, S. Route Planning and Map Inference with Global Positioning Traces. In *Computer Science in Perspective* **2598**, 128–151, https://doi.org/10.1007/3-540-36477-3_10 (2003).
5. Guo, T., Iwamura, K. & Koga, M. Towards high accuracy road maps generation from massive GPS traces data. In *Proceedings of the Geoscience and Remote Sensing Symposium* 667–670, <https://doi.org/10.1109/IGARSS.2007.4422884> (2007).
6. Shi, W., Shen, S. & Liu, Y. Automatic generation of road network map from massive GPS vehicle trajectories. In *Proceedings of the Intelligent Transportation Systems* 48–53, <https://doi.org/10.1109/ITSC.2009.5309871> (2009).
7. Liu, X., Wang, Y., Biagioni, J., Eriksson, J. & Zhu, Y. Mining large-scale sparse GPS traces for map inference: comparison of approaches. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* 669–677, <https://doi.org/10.1145/2339530.2339637> (2012).
8. Biagioni, J. & Eriksson, J. Inferring Road Maps from Global Positioning System Traces. Survey and comparative Evaluation. *Transportation Research Record: Journal of the Transportation Research Board* **2291**, 61–71, <https://doi.org/10.3141/2291-08> (2012).
9. Schrödl, S., Wagstaff, S., Rogers, P., Langley, P. & Wilson, C. Mining GPS Traces for Map Refinement. *Data Mining and Knowledge Discovery* **9**(1), 59–87, <https://doi.org/10.1023/B:DAMI.0000026904.74892.89> (2004).
10. Davies, J. J., Beresford, A. R. & Hopper, A. Scalable, distributed, real-time map generation. *Intelligent Transportation Systems* **5**(4), 47–54 (2006).
11. Lima, F. & Ferreira, M. Mining spatial data from GPS traces for automatic road network extraction. In *Proceedings of the 6th International Symposium on Mobile Mapping Technology* 1–7, http://www2.fct.unesp.br/docentes/carto/JoaoFernando/Artigos_MMT_2009/110_Lima_MMT09.pdf (2009).
12. Cao, L. & Krumm, J. From GPS Traces to a Routable Road Map. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* 3–12 (2009).
13. Ahmed, M. & Wenk, C. Constructing Street-Maps from GPS Trajectories. *20th Annual European conference on Algorithms* **1**, 60–71 (2012).
14. Ariza-López, F. J., Barrera, D., Reinoso, J. F. & Romero-Zalaz, R. Inferring mean road axis from Big Data: sorted points cloud belonging to traces. *Modelling, Computation and Optimization in Information Systems and Management Sciences* **1**, 443–453 (2015).
15. Goodchild, M. F. & Hunter, G. J. A simple positional accuracy measure for linear features. *International Journal of Geographical Information Science* **11**(3), 299–306, <https://doi.org/10.1080/136588197242419> (1997).
16. Kagawa, Y., Sekimoto, Y. & Shibasaki, R. Comparative study of positional accuracy evaluation of line data. In *Proceedings of the ACRS* (1999).
17. Tveite, H. & Langaas, S. An accuracy assessment method for geographical line data sets based on buffering. *International Journal of Geographical Information Science* **13**(1), 27–47, <https://doi.org/10.1080/136588199241445> (1999).
18. Johnston, D., Timlin, D., Szafoni, D., Casanova, J. & Dilks, K. *Quality Assurance/Quality Control Procedures for ITAM GIS Databases* (US Army Corps of Engineer Research and Development Centre, 2000).
19. Van Niel, T. G. & McVicar, T. R. Experimental evaluation of positional accuracy estimates from a linear network using point- and line-based testing methods. *International Journal of Geographical Information Science* **5**(16), 455–473, <https://doi.org/10.1080/13658810210137022> (2002).
20. Mozas-Calvache, A. T. & Ariza-López, F. J. Methodology for Positional Quality Control in Cartography Using Linear Features. *The Cartographic Journal* **47**(4), 371–378 (2010).
21. Mozas-Calvache, A. T. & Ariza-López, F. J. New method for positional quality control in cartography based on lines. A comparative study of methodologies. *International Journal of Geographical Information Science* **25**(10), 1681–1695, <https://doi.org/10.1080/13658816.2010.545063> (2011).
22. Haklay, M. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B: Planning and Design* **37**, 682–703, <https://doi.org/10.1068/b35097> (2010).
23. Ariza-López, F. J. & Rodríguez-Avi, J. A Method of Positional Quality Control Testing for 2D and 3D Line Strings. *Transactions in GIS* **19**(3), 480–492, <https://doi.org/10.1111/tgis.12117> (2015).
24. Goodchild, M. F. Citizens as sensors: the world of volunteered geography. *GeoJournal* **69**(4), 211–221, <https://doi.org/10.1007/s10708-007-9111-y> (2007).
25. Castelein, W., Grus, L., Crompvoets, J. & Bregt, A. A characterization of Volunteered Geographic Information. In *Proceedings of 13th AGILE International Conference on Geographic Information Science* 1–10 (2010).
26. Gil de la Vega, P., Ariza-López, F. J., Mozas-Calvache, A. T. Detection of outliers in sets of GNSS tracks from Volunteered Geographic Information. In *Proceedings of the 18th AGILE International Conference on Geographic Information Science*, https://agile-online.org/conference_paper/cds/agile_2015/posters/86/86_Paper_in_PDF.pdf (2015).
27. Devogele, T. A new merging process for data integration based on the discrete Fréchet distance. In *Proceedings of the 10th International Symposium on Spatial Data Handling* 167–181 (2002).
28. Leica. *Leica GPS1200 Series High performance GNSS System*, http://www.surveyequipment.com/PDFs/GPS_1200.pdf (Leica Geosystems, 2007).
29. Ariza-López, F. J., Gracia-Balboa, J. L. & Ureña-Cámara, M. A. ES2530686 (A1) - Dispositivo autonivelado para el levantamiento GNSS de elementos lineales, https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20150304&DB=&locale=en_EP&CC=ES&NR=2530686A1&KC=A1&ND=4# (Oficina Española de Patentes y Marcas, 2015).
30. Alt, H. & Godau, M. Computing the Fréchet distance between two polygonal curves. *International Journal of Computational Geometry and Applications* **5**, 75–91 (1995).
31. Gil de la Vega, P. *Generación de Ejes precisos 3D a partir de multitrizas GNSS y control posicional* (Servicio de Publicaciones de la Universidad de Jaén, 2017).
32. ESRI. *ESRI Shapefile Technical Description. An ESRI White Paper*, <https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf> (Environmental Systems Research Institute, 1998).
33. ISO. *Geographic information – Metadata – Part 1: Fundamentals*. ISO 19115-1:2014 (International Standardization Organization, 2014).
34. Ariza-López, F. J., Mozas-Calvache, A. T., Ureña-Cámara, M. A., Gil de la Vega, P. 3D GNSS-road traces dataset v2. *Figshare*, <https://doi.org/10.6084/m9.figshare.c.4375493> (2019).
35. ISO. *Geographic information – Data quality*. ISO 19157:2013 (International Standardization Organization, 2013).
36. Ariza-López, F. J., Mozas-Calvache, A. T. & Gil de la Vega, P. Tratamiento de multitrizas GNSS 3D para la obtención de ejes medios. *XVI Congreso Nacional de Tecnologías de la Información Geográfica* **2014**, 557–563 (2014).
37. Ariza-López, F. J., Rodríguez-Avi, J. & Reinoso, J. F. An approximation to outliers in GNSS traces. *Proceedings of Spatial Accuracy* **2014**, 186–189 (2014).

Acknowledgements

The work was supported by the National Ministry of Economy and Competitiveness of Spain under Grant No. BIA2011-23271. The authors also acknowledge the Regional Government of Andalusia (Spain) for the financial support offered since 1997 to their research group (Ingeniería Cartográfica) under code PAIDE-TEP-164. The University of Jaén contributed the funds for the open publication of this work.

Author Contributions

Francisco Javier Ariza-López proposed the design and participated in the data capture. Paula Gil de la Vega participated in the data capture. All authors collaborated in the data preparation and writing the paper.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2019