To cite this article: Antonio Tomás Mozas-Calvache & Francisco Javier Ariza-López (2019) Analysing the positional accuracy of GNSS multi-tracks obtained from VGI sources to generate improved 3D mean axes, International Journal of Geographical Information Science, 33:11, 2170-2187, DOI: 10.1080/13658816.2019.1645335

Analysing the positional accuracy of GNSS multi-tracks obtained from VGI sources to generate improved 3D mean axes

The sharing of Global Navigation Satellite System (GNSS) tracks on the Internet is increasing enormously. Every day a great number of users capture routes using different devices, such as navigators or smartphones, and share these data. Individually these tracks present a poor positional accuracy because these devices obtain positions with accuracy of about 5-10 metres. In addition, they are usually captured for navigation purposes and not for surveying. However, we can take advantage of the great quantity of tracks of the same linear element in order to obtain a more accurate solution. This study analyses this possibility using a wide set of tracks obtained from a circuit of roads in known conditions. More concretely, we emulated those tracks obtained by Volunteered Geographic Information (VGI) users and we compared the mean axis obtained using all tracks with others obtained from a more accurate source. Additionally, we analyse the displacement of other axes obtained by varying several parameters such as the number of tracks and their length or by dividing the route into sections in function of sinuosity, slope, type of road, etc. The results have shown an improved 3D mean axis and as a consequence the viability of the method proposed in this study in order to use axes obtained from several tracks in maps at certain scales.

Keywords: GNSS tracks, positional accuracy, quality, VGI

1. Introduction

In recent years data acquisition and the use of geographic information (GI) has experienced a great revolution. We can highlight two main causes, the global access to GI given by the Internet applications along with the evolution and worldwide distribution of devices using GI, such as smartphones, GPS navigators, etc. As an

example, Google Earth was launched in 2005 allowing users access to worldwide images and GI, opening up to the general public some of the more straightforward capabilities of the Geographic Information Systems (GIS) (Goodchild 2007). In this sense Google Earth has been called 'the democratization of GIS' (Butler 2006). In addition, the development of Spatial Data Infrastructures (SDIs) provides GI shared on the Internet by institutions and enterprises using web services, such as those standardized by the Open Spatial Consortium (OGC). Secondly, a new paradigm of data acquisition is appearing. From geographic data exclusively obtained by institutions and enterprises, a user-based data acquisition has appeared and been developed. As an example, Open Street Map (OSM) allows users to create and distribute free geographic data on the Internet (OSM 2018). Goodchild (2007) proposed the term Volunteered Geographic Information (VGI) to reference user-generated geographic data.

Until now, one of the GI datasets most widely acquired and shared on the Internet by VGI users has been GNSS tracks (VGI-GNSS tracks). These tracks were collected using GNSS navigators or smartphones along a determined route. As an example, GNSS tracks can be derived from vehicle trajectories on roads, bicycles following tourist routes, etc., capturing positions (longitude, latitude and height) at a certain frequency (usually 1Hz). So tracks are composed of positions (p1, p2, …, pn) spaced at a certain time interval. We have to note that other information can be available at each position, such as the dilution of GNSS precision (DOP) (based on satellite geometry), speed, etc. However, VGI users do not usually share this information because applications focus on the geographical position of the vehicle at a certain timestamp (which is sufficient for navigation purposes). There are several sites that allow sharing VGI-GNSS tracks on the Internet (as examples, OSM, Wikiloc, Bikemap, Ridewithgps, JustGoRide, etc.). Until now, several studies have used tracks

from VGI for different purposes. As examples, there are several studies on obtaining and assessing trajectories using data mining techniques (Cao and Krumm 2009; Liu et al. 2012; Basiri et al. 2016a; Basiri et al. 2016b), detecting anomalous behaviours in trajectories (Huang et al. 2014; Ivanovic et al., 2016), detecting problems of tracks obtained from VGI (Gil de la Vega et al. 2016), and which analyse the behaviour of vehicles using GNSS tracks from OSM (Mozas-Calvache 2016). Huang et al. (2013) described a more detailed analysis of VGI trajectories based on GPS. In addition, other studies described the use of GNSS tracks in map inference. As an example, Biagioni and Eriksson (2012) described a survey and comparative evaluation of map inference algorithms based on the k-means (Edelkamp & Schrödl 2003), trace merging (Cao & Krumm 2009) and kernel density estimation (Davies et al. 2006).

One important aspect of GI, and consequently of VGI, is that final users demand reliable data. In recent times there has been an increase of interest in the quality of VGI data. ISO (1999) described several quality components of GI (positional accuracy, thematic accuracy, temporal accuracy, logical consistency and completeness). Some studies have analysed these components of VGI data (Cooper et al. 2011; Antoniou and Skopeliti 2015; Senaratne et al. 2017). Among others, the positional component of quality has a great importance in GI and more concretely in GNSS tracks obtained by VGI users because of the necessities of correct positioning. This component is difficult to analyse in shared tracks because of the diversity of acquisition procedures and data. As examples, there are a great variety of devices with different technical specifications and configuration possibilities, managed by non-professional users for non-surveying purposes. Additionally, data could be shared in different applications using several formats (GPX, KML, etc.), with several levels of privacy, etc. Finally, as indicated by

Gil de la Vega et al. (2016), another problem is that there is not always a great availability of traces of the same route.

The GNSS devices usually used by VGI contributors have a positional accuracy of about 5-10 metres (Haklay and Weber 2008; Zhang et al. 2010). This value is sufficient for navigation purposes because applications include map-matching algorithms that correct the position captured on a defined route (White et al. 2000). However, this accuracy is not sufficient for using tracks for other purposes such as mapping (at scales higher than 1:20000 [ASPRS 1990]) or controlling maps. Thus GNSS tracks shared by VGI contributors present a priori a low positional accuracy. However, we can take advantage of the large number of tracks of the same linear element shared by VGI contributors in order to obtain a more accurate solution. A large number of low accurate values could derive into a more accurate mean value.

To summarize, our assumption is that a mean track derived from a set of tracks will provide a more accurate (positional) solution than each individual track. This hypothesis is based on the fact that the variance of the average of n independent measurements is l/n of the variance of an individual measurement. Our goal is not to demonstrate the last relationship exactly. Our objective is to demonstrate empirically that there is a reduction and, from there, determine the number and length of the set of tracks or sections of tracks to be used. So we have developed an observational study under known conditions where we emulate the surveys carried out by VGI users in capturing a large number of tracks of a selected road and compare several mean axes, determined using different parameters, with a more accurate one. The main goals are to demonstrate the above-mentioned assumption and establish these parameters (the number and length of the tracks) considering the accuracy of the axis obtained. In addition, this study was applied to different types and conditions (sinuosity, slope) of

roads in order to analyse their possible influence on the accuracy achieved. Considering this accuracy, the axes obtained using VGI-GNSS tracks could be used to update cartography, control maps, etc.

This document is structured as follows. Firstly, a description of the method implemented and its application to real data is defined in section 2. The main results obtained and the discussion of them are shown in section 3. Finally, section 4 describes the main conclusions of this study and future work derived from it.

2. Method and application

The method is conditioned by the observational study, and the design is focussed on accomplishing our hypothesis and objectives. Prior to the presentation of the method, some definitions and clarifications are necessary:

- GNSS tracks (t_1, t_2, …, t_m) of a route are considered as linear elements composed of a sequence of edges (e_1, e_2, …, e_{n-1}) defined by vertexes (v_1, v_2, …, v_n). The vertexes are related to those positions ([x_1, y_1, z_1], [x_2, y_2, z_2], …, [x_n, y_n, z_n]) captured by GNSS devices during the route, ordered by acquisition time. A GNSS track section S_i is a continuous part of a complete track determined by a sectioning criterion. Additionally, tracks and sections can be assigned attributes (e.g. slope, sinuosity, road type) that can be used to make selections.
- Sample of tracks. In our study the population of interest are the tracks (understood as complete tracks or as sections of a complete track). This population is infinite and is estimated by a sample of tracks obtained under known conditions.

- Reference axis. This is a linestring that corresponds to the axis of the roads used in this study. This axis has been surveyed using more accurate methods and is considered as the ground truth for our study.
- Variable of interest (response). In our study the variable of interest is the positional accuracy of the mean of a set of tracks/sections. The response is measured by the comparison of mean axes derived from the sets of tracks/sections versus a source of greater positional accuracy (the reference axis). The accuracy is measured using the Vertex Influence Method (VIM) (Mozas and Ariza-López 2011, Mozas and Ariza-López 2015). Mean and deviations are computed in order to describe the accuracy achieved.

- Factors (or inputs to the process). Factors are all the several aspects which we consider can influence the results. In our case the factors to be considered are: number of tracks/sections, length of tracks/sections (by sectioning original tracks), slope of tracks/sections and sinuosity of tracks/sections.
- Levels (settings of each factor in the study). In our study the levels are the following:
    - For the quantity, we considered nine levels [5, 10, 15, 20, 25, 30, 40, 50, 60]
    - For the length, we considered four levels [1000, 2000, 5000, 10000]
    - For the slope, we considered three levels (Low, Medium and High) [L, M, H]
    - For the sinuosity, we considered three levels (Low, Medium and High) [L, M, H]

2.1. Method

In this study, we develop five analyses that have a common method part. These analyses are:

A1. Analysis of the positional accuracy achieved by the mean axis calculated from the set of all the complete tracks. The objective is to generate a mean axis using the greatest and most complete possible sample.

A2. Analysis of the positional accuracy achieved by the mean axis calculated from a certain number n = [5, 10, 15 ...] of complete tracks. In this case, a bootstrap process is applied (resampling with replacement). The objective is to see how positional accuracy varies according to the number of tracks.

A3. Analysis of the positional accuracy achieved by the mean axis calculated from homogeneous sections for different criteria (e.g. slope, sinuosity, etc.). The objective is to learn whether these aspects (e.g. slope, sinuosity, etc.) have an influence on positional accuracy.

A4. Analysis of the positional accuracy achieved by the mean axis calculated from a certain number n = [5, 10, 15 ...] of homogeneous sections for different criteria (e.g. slope, sinuosity, etc.). In this case, a bootstrap process is applied (resampling with replacement). The objective is to know how the positional accuracy varies depending on the number of tracks for each type of criterion.

A5. Analysis of the positional accuracy achieved by the mean axis calculated from a certain number n = [5, 10, 15 ...] of sequences of sections of a given total length lt = [1000, 2000, 5000 ...]. In this case, a bootstrap process is applied (resampling with replacement). The objective is to know how positional accuracy varies according to the total length of the set of sections.

The five analyses mentioned above share the calculation method. The method is summarized in Figure 1. This figure shows the calculation method (in grey) and the

criteria and parameters that are established on the data (GNSS sections) for each of the five analyses. In the case of bootstrapping, the number of iterations is 100. As can be seen, all analyses are performed against the reference data. Below the calculation process will be explained in more detail in a general way (valid for the five analyses).

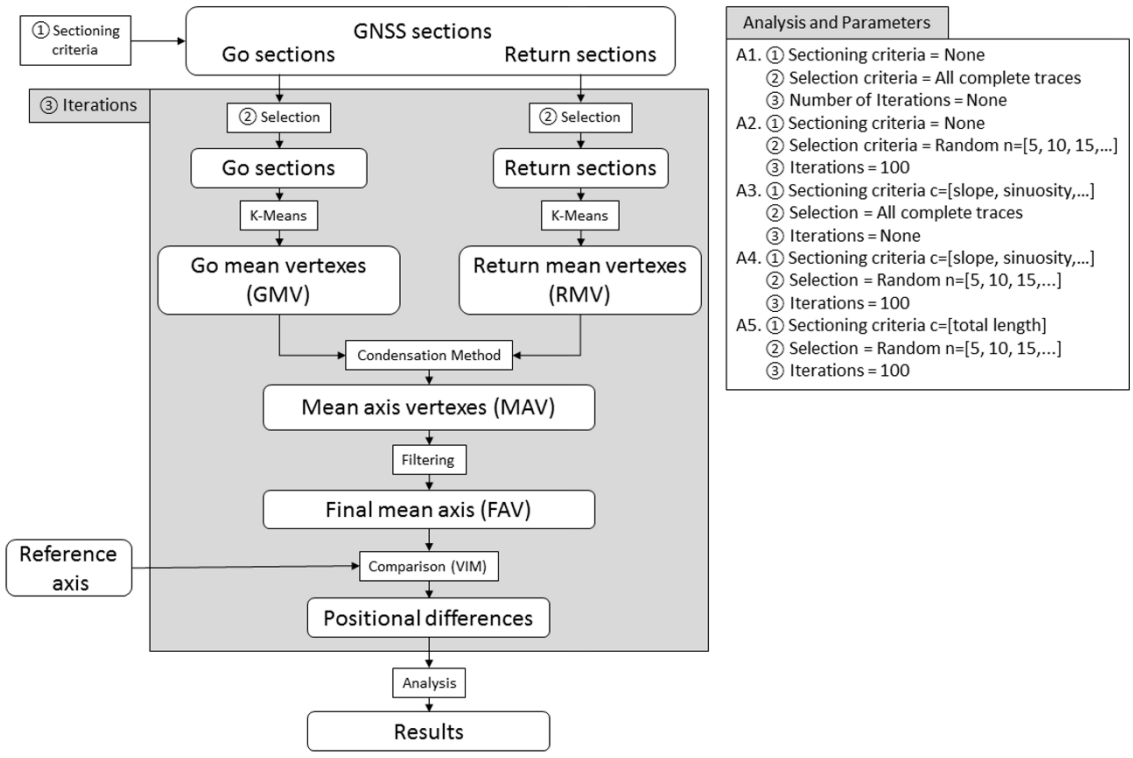


Figure 1. Method developed in this study

The procedure for obtaining a mean axis from several tracks/sections is described below. Firstly, we obtained a mean linestring from the selected tracks/sections developed in the go direction and another mean linestring using the tracks/sections of the return direction. In this case, we used the K-means algorithm (MacQueen 1967; Edelkamp & Schrödl 2003) that consists of the determination of some cluster seeds along the element to be determined. The vertexes of all tracks/sections were assigned to each cluster depending on their distance to the seeds and afterwards the cluster seeds were recalculated using all the vertexes that compose each cluster. This procedure is repeated until all vertexes are definitively assigned to one cluster and there are no

changes in the assignation. The mean linestring will be composed of the collection of the centre of the clusters obtained (Go Mean Vertexes [GMV] and Return Mean Vertexes [RMV]). The selection of K-means algorithm was based on the characteristics of tracks shared by VGI contributors. Thus these tracks are usually very heterogonous, with variable density (captured at different frequencies), interruptions, differences in starting and ending points, etc. These conditions imply using algorithms based on clustering in contrast to those based on the sequence of vertexes of tracks.

Secondly, once there are two mean linestrings defined by GMV and RMV (go and return), we obtain the final axis of the track/section using the Condensation Method (Mozas-Calvache and Ariza-López 2017). For two linestrings this method consists of the determination of the vertexes of the mean axis (Mean Axis Vertexes [MAV]) by ordering all the crosses between both lines and the middle points of the lines that connect each vertex of one line to another and vice versa. After that, a filtering process based on a minimum distance between consecutive vertexes could be carried out in order to reduce the quantity of vertexes that composes the axis obtained. The final axis (FAV) is composed of the MAVs that are not removed after this filtering procedure. So far, two averages have been calculated. The first is to obtain each of the vertices of the GMV and RMV axes. These vertices are the centroids of the clusters. In each of them different quantities of cases (n = [5, 10, ...]) intervene. Therefore, an estimate of its error is σ^2/n; And the second is to obtain each one of the vertexes of the MAV. Here only two vertices of the previous type intervene. For this reason, the error is $\sigma^2/2n$. As indicated previously, our goal is not to demonstrate the last relationship exactly but demonstrate empirically that there is a reduction and, from there, determine the number and length of the set of tracks or sections of tracks to be used.

Finally, the FAV derived from the considered set of tracks/sections was compared to the reference axis. This comparison was based on the determination of the positional differences using the Vertex Influence Method (VIM) both in 2D and in 3D. This method consists of the determination of the distances from the vertexes of the more accurate linestring (reference) to the linestring to be controlled (FAV). The mean displacement is obtained by weighting all distance values by the length of the adjacent segments (reference linestring) to the implicated vertex.

2.2. Data and application

2.2.1 Data

The application of this method was developed using a field survey based on a GNSS device capturing positions of several roads that emulate those captured by VGI contributors. More concretely, we selected a closed circuit of about 12136 metres composed of two types of roads (main and local) with great variability in sinuosity and slope (Figure 2a and Figure 2c). Table 1 describes the main aspects of the roads used in this study. The N323 main road (Figure 2a) was located along a valley, the JV3231 local road connected this main road to several villages and the JV2227 local road was an old road used infrequently (Figure 2b). This circuit was surveyed 69 times using a GNSS navigator (Columbus V-990) (Figure 3a) such as those commonly used by VGI contributors. The main characteristics of the sample of GNSS tracks used in this study are shown in Table 2.

Table 1. Characteristics of the roads used in this study

	Width	Road shoulder	Sinuosity	Slopes	Roadside vegetation

N323	7 m	1.5 m	Low	Low	Abundant
JV3231	6 m	No	Low-Medium	High	Occasional
JV2227	5 m	No	High	High	Occasional

Table 2. Main characteristics of the sample of GNSS tracks used in this study.

	Go	Return
Device	Columbus V990	Columbus V990
Accuracy	3m CEP; 5m CE (95%)	3m CEP; 5m CE (95%)
Number of complete tracks	69	69
Number of kilometres	837	837
Number of positions	55200	55800
Mean distance between positions	15.1 m	15.0 m
Average capture speed	55 km/h	55 km/h

The procedure for obtaining a mean linestring using more than two tracks was carried out using the K-means algorithm (Edelkamp and Schrödl 2003). We used the vertexes of the tracks with lower displacements (using VIM) with respect to the other tracks as seeds. In addition, a maximum distance of 20 metres (about four times the value of 95% CE) was used as filtering in each cluster calculation to avoid the inclusion of undesirable values. After that, an axis was determined using pairs of linestrings from both directions (go and return) through the Condensation Method. Finally, we compared these cases to another axis obtained from a more accurate source that was used as

control using VIM.

The base data for the reference axis were obtained through a GNSS accurate survey developed using a geodetic-surveying GNSS system. More concretely, we used a Leica System 1200+ GNSS device with 10mm+1ppm and 20mm+1ppm both in horizontal and vertical components for kinematic post-processing accuracy. The GNSS processing was developed using data downloaded from the UJAEN GNSS reference station located 18 kilometres from the study zone (Berrocoso et al. 2006). The GNSS observation data were captured at 1Hz, surveying both edges of the roads using a device that guaranteed the levelling and height of the GNSS antenna (Figure 3b) (Ariza-López et al. 2015). The GNSS surveying was carried out by one operator walking along the edge of the roads, and as a consequence the number of points captured presented a great density. After GNSS processing, two linestrings referring to both edges of the roads were obtained. The reference axis of the road was calculated using the Condensation Method. After filtering, more than 11300 vertexes composed the final reference axis, which supposed a mean value of 1 metre between contiguous vertexes. Both the GNSS tracks dataset and the accurate GNSS dataset used in this study are available in Ariza-López et al. (2018).

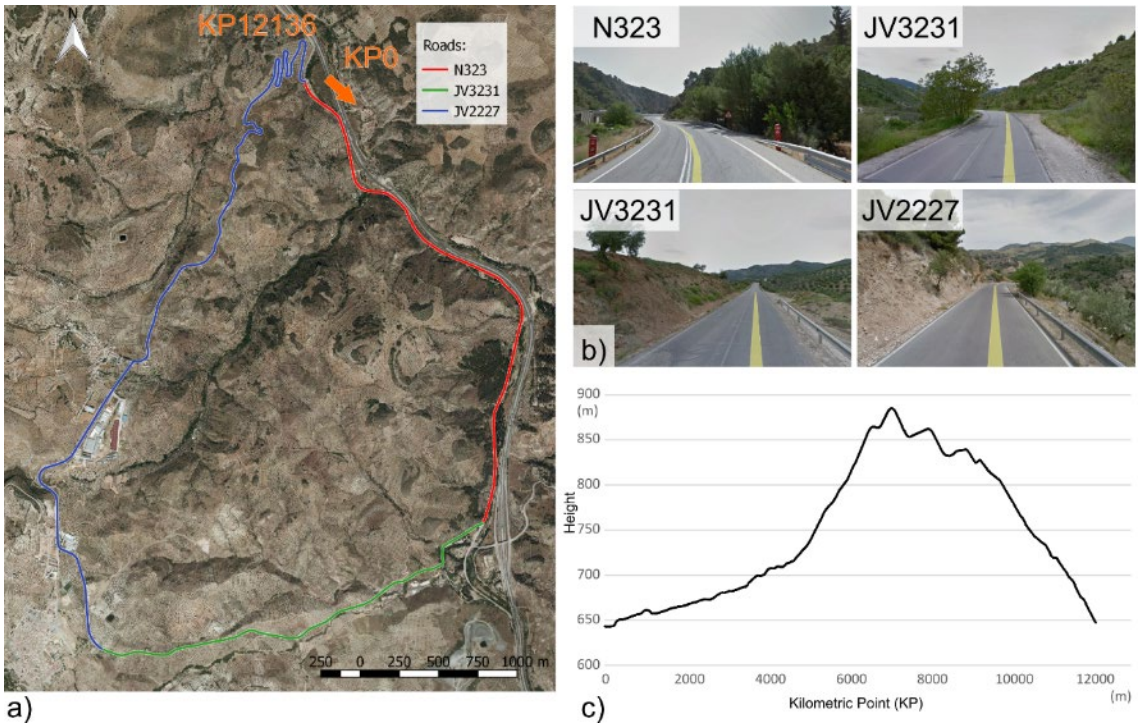


Figure 2. Zone of study: a) roads used in this study (background: PNOA orthoimage provided by Instituto Geográfico Nacional of Spain); b) views of roads (views provided by Google Earth); c) longitudinal profile of the route.

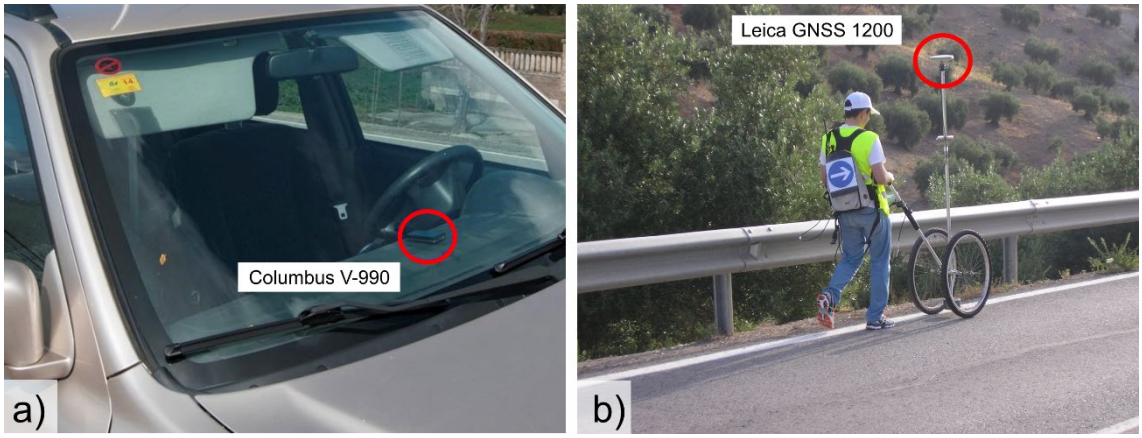


Figure 3. GNSS surveys: a) GNSS navigator used in tracks survey; b) Leica GNSS 1200+ used in the accurate survey.

2.2.2 Application

After obtaining both the sample of GNSS tracks and the reference axis (Figure 4a and Figure 4c), we applied the previously-described method using the complete sample of

tracks/sections and several resamples, depending on the analysis case. All the studies were developed in order to analyse the positional accuracy achieved with the FAV (which is a mean axis) derived from the data under analysis versus the reference axis. Below are some details of each of the analyses (A1, A2…A5).

A1. Firstly, we obtained a mean axis using all Columbus GNSS tracks (69 of go and return) (Figure 4b). The linestrings obtained from go and return surveys derived a FAV that is compared to the reference axis using VIM.

A2. Secondly, we used a random selection with replacement of 5, 10, 15, 20, 25, 30, 40, 50 and 60 GNSS tracks of all routes, obtaining an FAV for each case. These axes were compared to the reference axis. This process was repeated 100 times, obtaining 100 values of displacements both in 2D and in 3D.

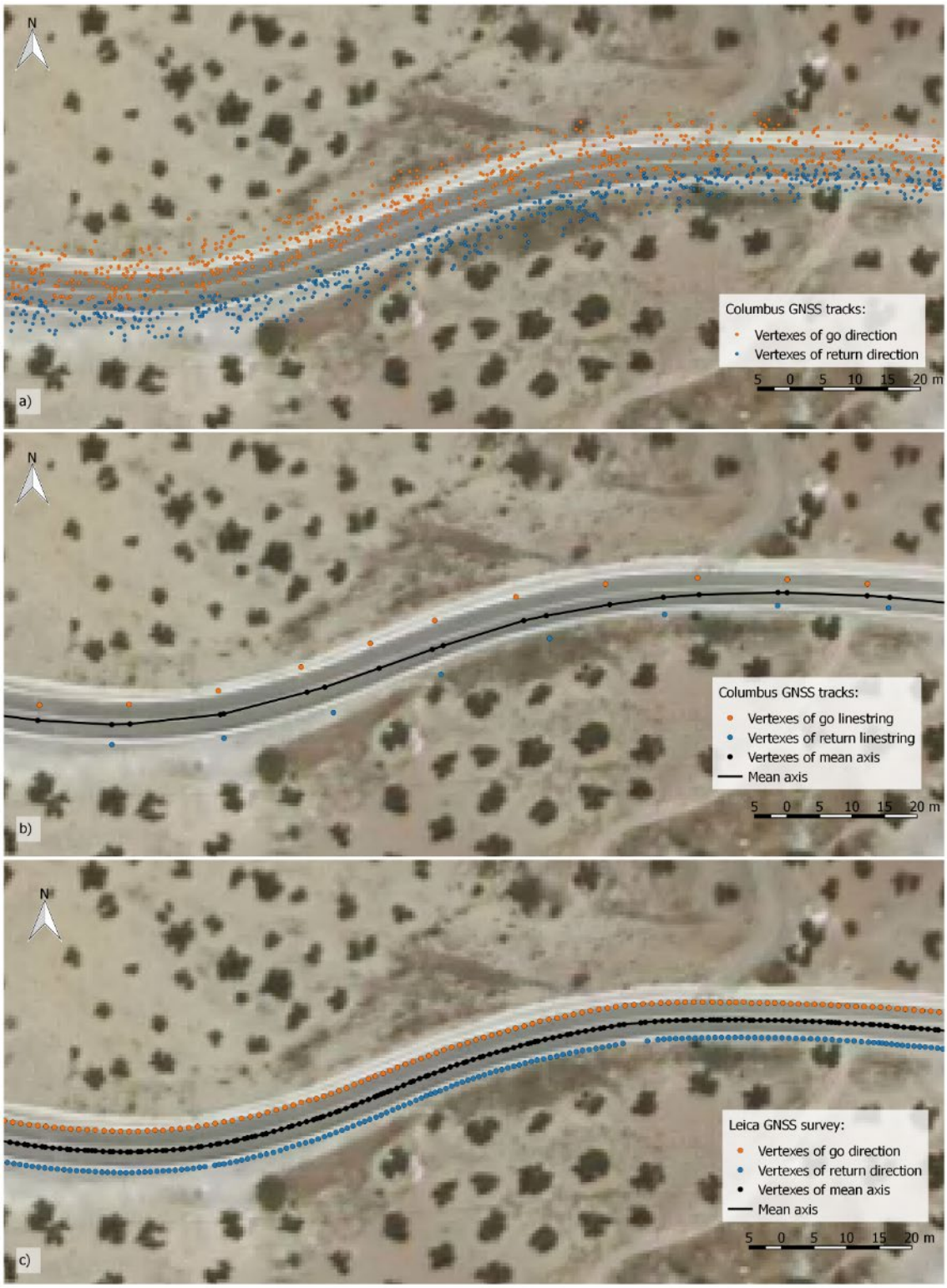


Figure 4. Examples of datasets used in this study: a) Columbus GNSS tracks (69 of go direction and 69 of return direction); b) Columbus go and return linestrings and mean axis for 69 tracks; c) Leica accurate survey and mean reference axis obtained. (Background: PNOA orthoimage provided by Instituto Geográfico Nacional of Spain).

A3. Thirdly, we divided the route into homogeneous sections (Table 3) by considering three types of road (N323, JV3231, JV2227), three classes of average slope (low, medium and high slope, <2%, 2-6% and >6% respectively) and three classes of sinuosity (low, medium and high sinuosity). The slope and sinuosity classes definition were based on the obtaining of similar sets in length. The selection of sections based on slope and sinuosity was performed by dividing the route into sections of 50 metres of length and computing the average slope and sinuosity of these sections. The calculation of sinuosity was carried out using the "Calculate Sinuosity" ArcGIS plugin, selecting values in the interval (0.9994, 1] for the Low Sinuosity class, (0.997; 0.9994] for the Medium Sinuosity class and values lower than 0.997 for the High Sinuosity class.

A4. Fourthly, as in the analysis A1, we applied this 100 times to the random selection with replacement of 5, 10, 15, 20, 25, 30, 40, 50 and 60 homogeneous sections.

Table 3. Characteristics of datasets used in the analysis by sections of the route

Case	Initial K.P.	Final K.P.	Length (m)
N323	0	3508	3508
JV3231	3508	6281	2773
JV2227	6281	12136	5855
Low slope (%) (<=2)	-	-	3850
Medium slope (%) (2, 6)	-	-	4250
High slope (%) (>=6)	-	-	4036
Low sinuosity (0.9994, 1])	-	-	4150
Medium sinuosity (0.997, 0.9994]	-	-	3986
High sinuosity (≤0.997)	-	-	4000

A5. Fifthly, in order to analyse the influence of the length of the tracks we selected several sets of sections of tracks whose characteristics are displayed in Table 4. The set of sections are composed by 10, 20, 50, 100 randomly selected sections of 100 m length; 2, 4, 10, 20 randomly selected sections of 500 m length and 1, 2, 5, 10 randomly selected sections of 1000 m length. The main goal was to determine whether the length of the track presented any influence on the accuracy obtained. As in previous analysis, we applied this 100 times to the random selection with replacement of 5, 10, 15, 20, 25, 30, 40, 50 and 60 tracks from which the random section of the sets of sections is performed.

Table 4. Definition of cases for the analysis by length.

Case	Number of sections	Length of each section (m)	Total length (m)
L10x100	10	100	1000
L2x500	2	500	1000
L1x1000	1	1000	1000
L20x100	20	100	2000
L4x500	4	500	2000
L2x1000	2	1000	2000
L50x100	50	100	5000
L10x500	10	500	5000
L5x1000	5	1000	5000
L100x100	100	100	10000
L20x500	20	500	10000
L10x1000	10	1000	10000

3. Results and discussion

The results of each of the analyses performed in this study are described below. The results are shown in graphs that display the displacements obtained (vertical axis of the graph) with respect to the number of tracks/sections (horizontal axis of the graph), both in 2D and in 3D. This type of graph was previously used in other studies such as Ariza-López et al. (2011) in order to analyse the influence of the sample size in positional assessments based on lines. We have to note that the 2D and 3D graphs present the vertical axis displaced but equally scaled.

3.1. Mean axis of all elements (analysis A1)

First, we obtained the mean value of VIM displacement between the reference axis and that derived from all Columbus tracks (69 tracks of each go and return direction) for the full route of 12136 metres. The mean value of displacement was 1.28 m in 2D (VIM2D) and 3.71 m in 3D (VIM3D). This 2D value supposes that this axis could be used for planimetric purposes at scales lower than 1: 7000. The Columbus GNSS device used in this study presented an a priori accuracy of about 3 metres in horizontal positioning (50% CEP) (Columbus V990 specifications). By using a more accurate and independent reference for the calculation of positional accuracy, these results have demonstrated our main hypothesis related to the possibility of obtaining a more accurate mean axis from a set of GNSS tracks with respect to that derived from individual tracks.

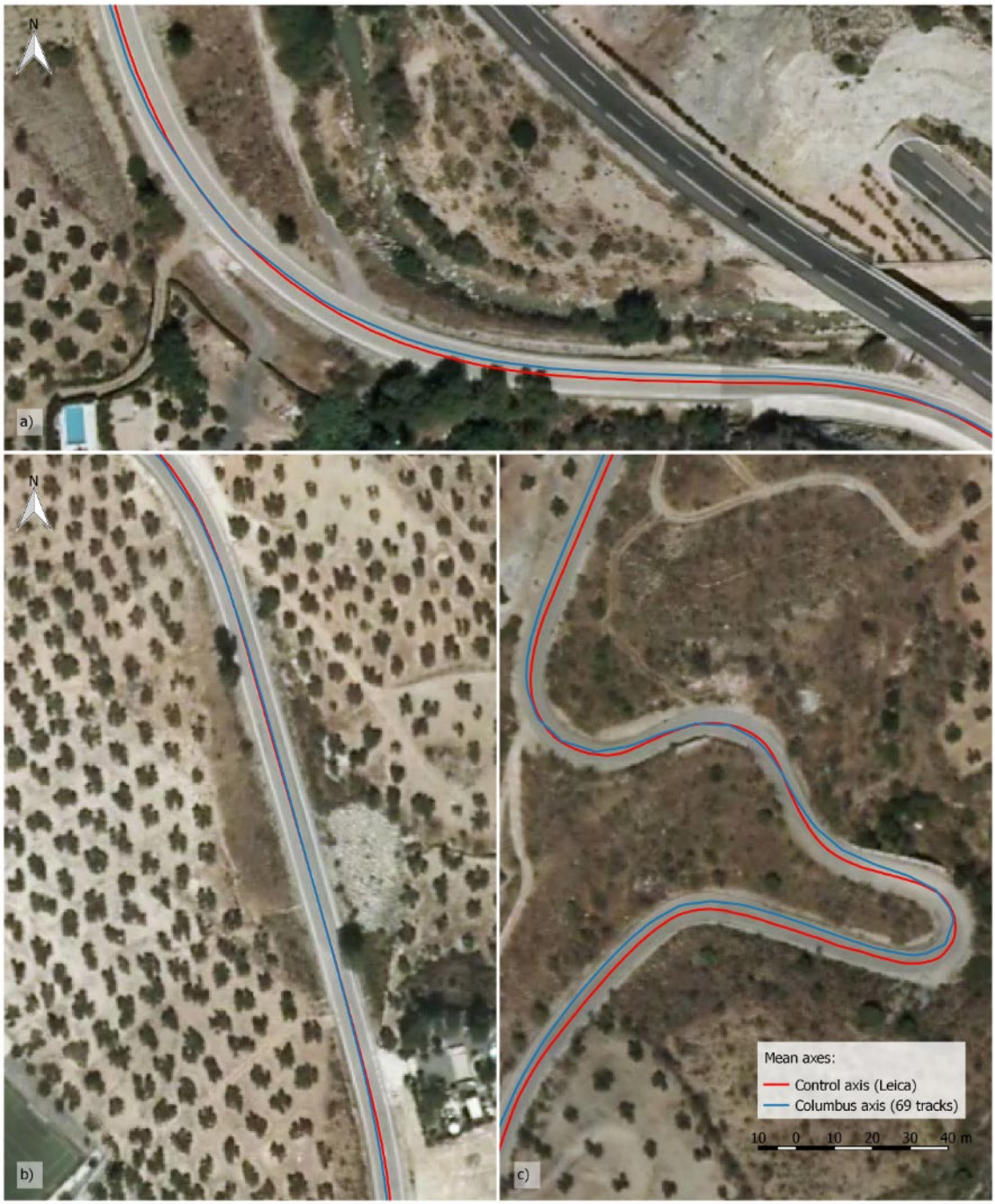


Figure 5. Examples of the comparison of axes obtained (reference axis vs. Columbus axis [using 69 GNSS tracks]): a) road N323; b) road JV3231; c) road JV2227. (Background: PNOA orthoimage provided by Instituto Geográfico Nacional of Spain).

Figure 5 shows some examples that compare the reference axis and the axis obtained from the sample of 69 GNSS tracks. These examples show clear differences depending on the sinuosity of the route and the type of road. In this sense, the greatest displacements appeared in narrow roads and in zones of curves mainly based on

capturing problems derived from occlusions. In addition, driving behaviour is another aspect to consider in curve zones. In these zones, there is a clear tendency for drivers to increase the radius of curves (see examples in Figure 5).

Figure 6 shows some examples of individual differences (distances) obtained between the vertexes of the reference axis with respect to the FAV obtained using 69 GNSS tracks. These examples shows that the greatest differences mainly appear in curve zones with high sinuosity (XY case) and built-up areas (Z case).

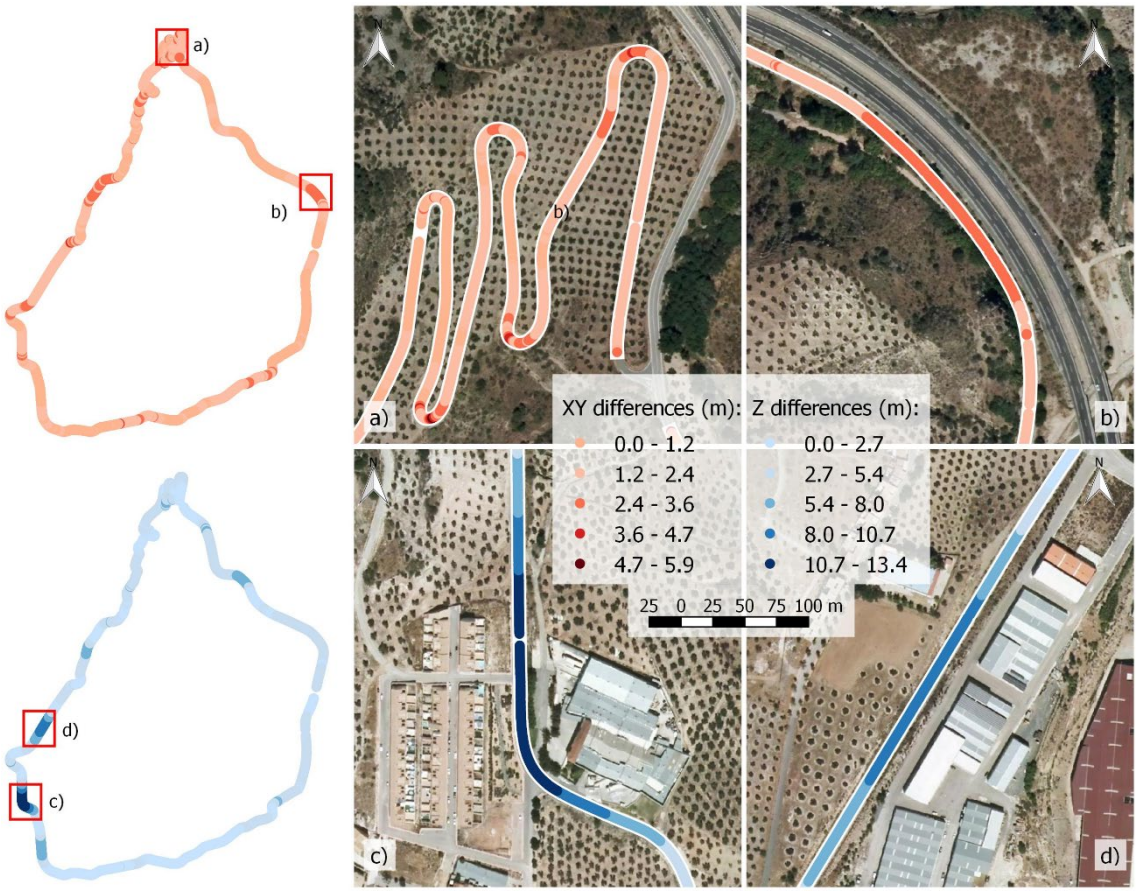


Figure 6. Examples of XY and Z differences obtained between vertexes of reference axis and Columbus axis: a) and b) XY differences; c) and d) Z differences (Background: PNOA orthoimage provided by Instituto Geográfico Nacional of Spain).

3.2. Number of complete tracks (analysis A2)

Second, the mean results for the case A2 analysis are displayed in Figure 7. These results shows curves of mean displacements that approached the results of the full

sample (69 tracks) and dispersions that were reduced with the increase of the number of tracks. However, there were great differences between 2D and 3D, mainly in dispersions. The results have shown that it is not necessary to use 69 tracks. As Figure 7 reflects, a very similar accuracy can be achieved using a reduced number of GNSS tracks (mainly in 2D). Based on this graph, we could select a lower number of tracks considering the purpose and accuracy requirements of the mean axis. As a general suggestion we recommend using a minimum of 15 tracks in order to limit the variability to the order of ±0.1metre in the 2D case. In the 3D case, the dispersions are greater than in the 2D case. Thus, the dispersion achieves 0.60 metres using 15 tracks. To reduce the dispersions of the 3D case we suggest using a large number of tracks (e. g. with 50 tracks the variability is in the order of ±0.2 metres). To establish this value we must consider that the Z accuracy (using GNSS) is usually worse than the X and Y accuracies.

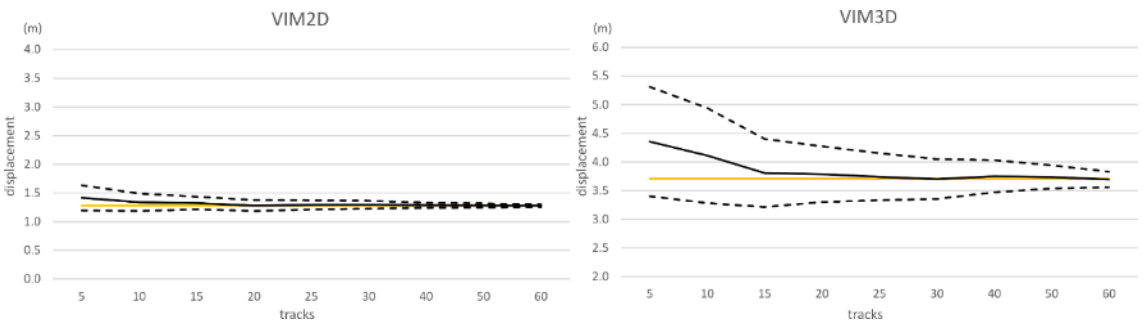


Figure 7. Results of VIM displacements (2D and 3D) of complete tracks depending on the number of tracks: Mean VIM value of all tracks (69) on yellow line (as reference), mean value of VIM displacement on continuous black line and the dispersion given by the standard deviation values on discontinuous black lines.

3.3. Homogeneous sections (analysis A3)

The results obtained after dividing the tracks into homogeneous sections considering several parameters are shown in Table 5 (69 GNSS tracks). They correspond to mean

values of displacement between the reference axis and the FAV axes of these sections. These results show differences between the displacements obtained both in the 2D and in 3D cases.

Table 5. VIM displacements for each homogeneous section.

Criteria	Case	2D (m)	3D (m)
Complete sample	Complete tracks	1.28	3.71
Type of road	N323	1.01	3.64
	JV3231	1.18	2.37
	JV2227	1.50	4.39
Slope	Low slope (%) (≤2)	1.21	3.86
	Medium slope (%) (2, 6)	1.29	3.68
	High slope (%) (≥6)	1.33	3.60
Sinuosity	Low sinuosity (0.9994, 1])	1.19	4.11
	Medium sinuosity (0.997, 0.9994]	1.33	3.56
	High sinuosity (≤0.997)	1.31	3.44

3.4. Homogeneous sections and quantity (analysis A4)

Figure 8 shows the mean results of obtaining 100 axes using samples of 5, 10, 15, 20, 25, 30, 40, 50 and 60 tracks selected randomly with replacement considering these sections.

The analysis by the type of road (Figure 8a) shows that the main road N323 presented the lowest displacement and the local road JV2227 shows the greatest values in 2D while the local road JV3231 shows the best behaviour in 3D. In the 3D case, there were greater differences between all cases. In this case, the JV2227 shows the greater displacements.

The study by slopes (Figure 8b) shows similar displacements in 2D but with an increase of the differences with the slope. The lowest displacements were related to the lowest slopes and the highest to the highest slopes. In the 3D case, the results show inverse tendencies with respect to the 2D case. Thus, the highest slopes show the lowest differences and the lowest slopes show the greatest differences.

Finally, the analysis by sinuosity (Figure 8c) also shows differences between 2D and 3D cases. In the 2D case, the results show the lowest displacements in the low sinuosity case (0.9994, 1]. In contrast, the 3D case show that the sections with the highest sinuosity (≤0.997) presented lower displacements with respect to the sections of lower sinuosity.

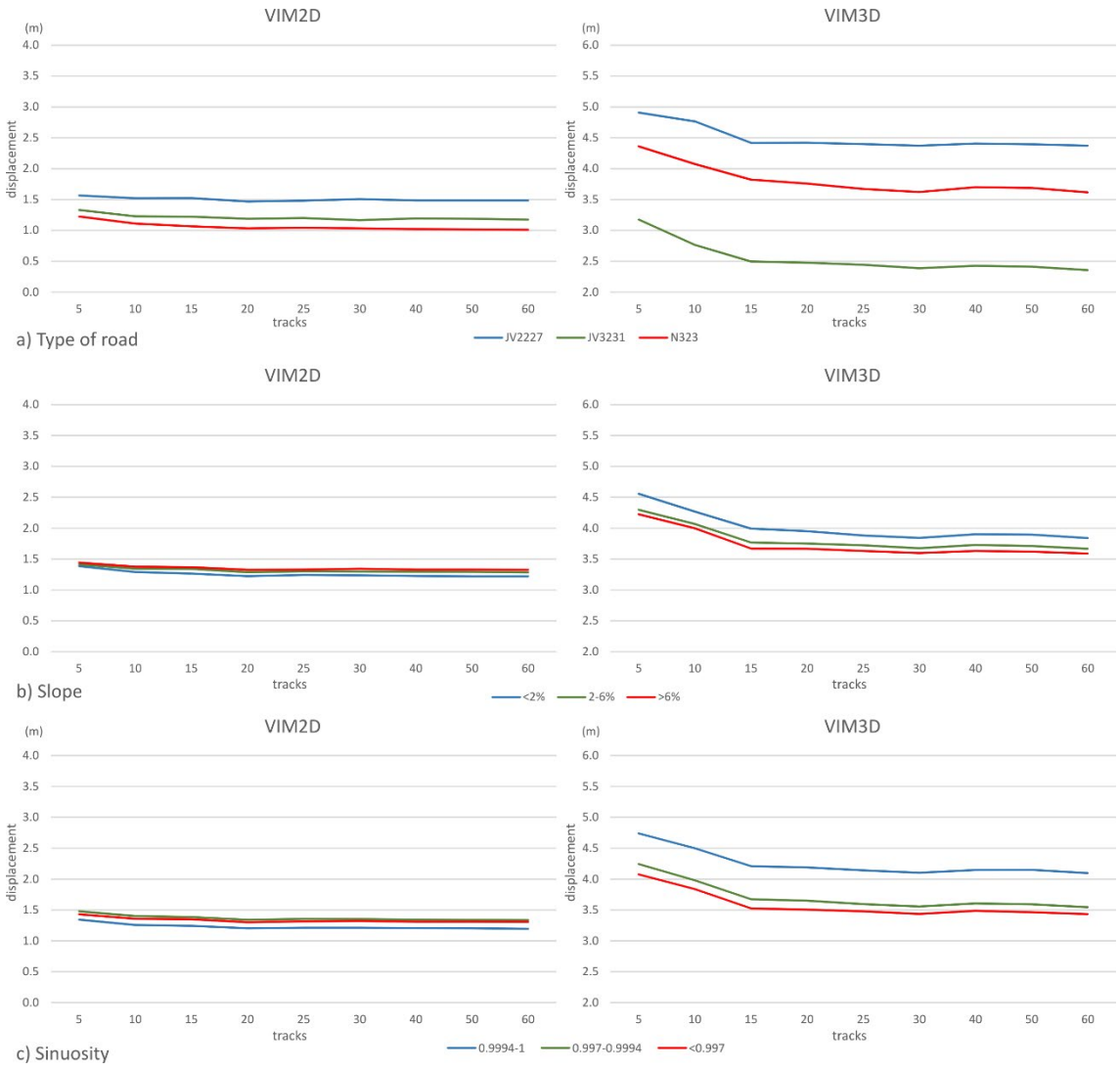


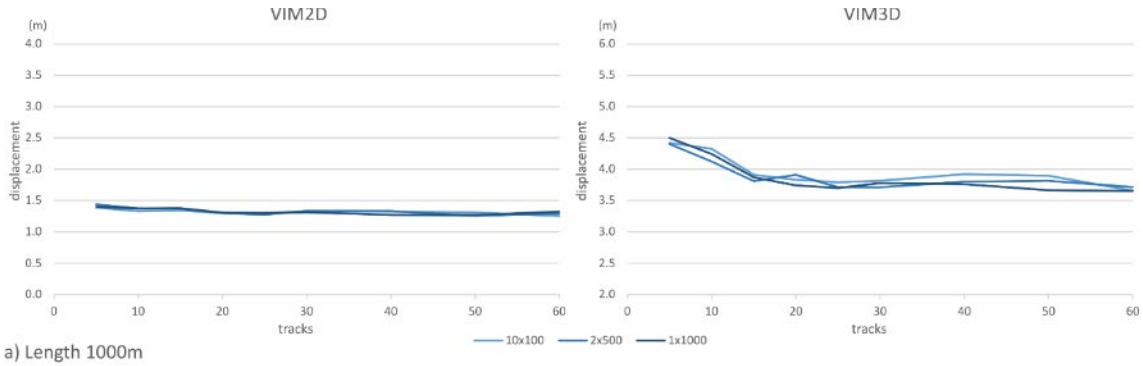
Figure 8. Results of VIM displacements (2D and 3D) by sections and depending on the number of tracks.

The analysis of the results of displacements of several sections of the route shows differences between 2D and 3D. In the 2D case, the results of mean displacement show that the N323 road presented the best behaviour while the JV2227 road had greater values of displacements. This is coherent if we take into account the possible influence of occlusions in positioning. Occlusions are more probable in narrow roads (JV2227). However, the slope and sinuosity showed lower influence in the 2D case. Therefore, the main factor is related to the type of road and more concretely, the presence (or not) of occlusions. On the other hand, the case of 3D displacements shows some differences with respect to the 2D cases. In the case of JV2227 the displacements were the greatest, as occurred in the 2D case. However, there was a great reduction of displacement in the JV3231 case. These results of slope and sinuosity were opposite in the 3D case with respect to the 2D case. The best behaviour was obtained in the highest slopes and the highest sinuosity cases. Despite the differences in horizontal and vertical accuracies of GNSS devices, we have detected anomalous behaviours depending on the section analysed between 2D and 3D displacement. We suppose that these results and differences between 2D and 3D results were widely conditioned by the presence of occlusions and multipath effects.

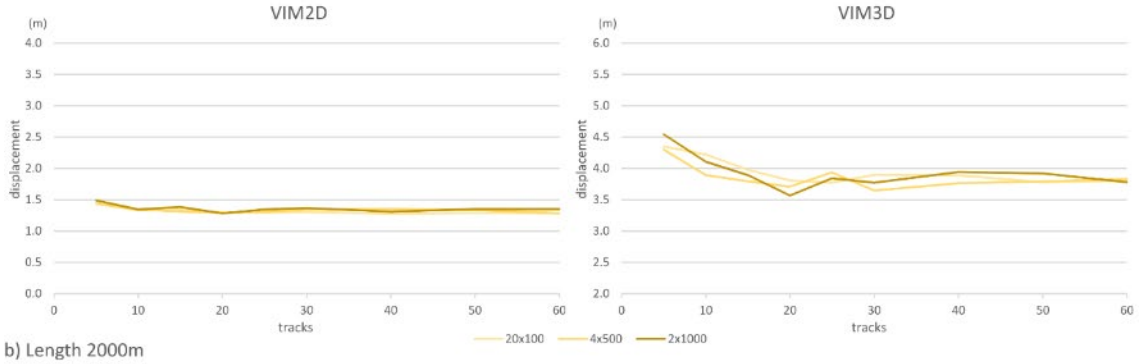
3.5. Total length of a sequence of sections (analysis A5)

The results of the analysis by total length of sections are displayed in Figure 9. In all cases, graphs show more stable curves with the increase in length and number of sections. More concretely, the variability of curves is greater in 3D (cases of lengths of 1000 metres and 2000 metres) (Figure 9a and Figure 9b). The analysis of the total length of the sections shows some variability in the cases of 1000 metres and 2000

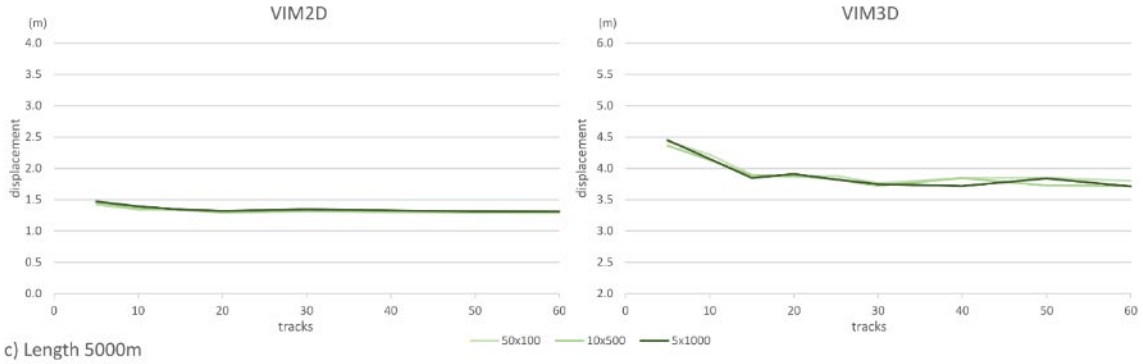
metres (more clearly in 3D), but we cannot establish clear differences in order to determine a minimum number of tracks. Therefore, we consider that this aspect is independent in selecting a sample of sections.



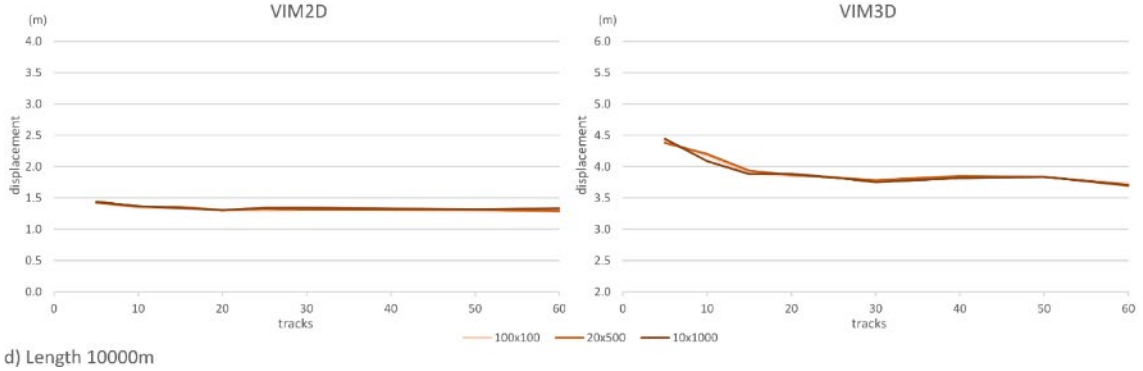
a) Length 1000m



b) Length 2000m



c) Length 5000m



d) Length 10000m

Figure 9. Results of VIM displacements (2D and 3D) by length of tracks and depending on the number of tracks.

4. Conclusions

This study has analysed the positional accuracy of mean axes of roads obtained from GNSS tracks such as those acquired and shared on the Internet by VGI contributors. The analysis has been based on the comparison of these axes with respect to another axis obtained from a more accurate source. In addition to the method used, the great amount of data and the analysis developed, the main novelty of this study has been the application both in 2D and in 3D. More concretely, the method, which has been applied on a large sample of GNSS tracks, has demonstrated the possibility of increasing the positional accuracy of the result by using a mean axis obtained from a set of tracks instead of individually. This aspect is very important considering the new paradigm of GI where users have become active contributors. Moreover, this study has also analysed the consequences from the perspective of the positional accuracy results of using several parameters in the determination of the axes, such as the number (sample size) and the length of sections, and the variability of the displacements related to some aspects such as the type of road, slopes, sinuosity of the route, etc.

In conclusion, the results of applying the method proposed in this study have demonstrated our assumptions, allowing us to suggest a minimum number of about 15 tracks for general purposes in 2D. The results also confirm that height data obtained from GNSS tracks presented discrepancies up to 3 times greater than horizontal components. In addition, we obtained more variability in mean displacements of 3D case. We suggest a minimum number of about 50 tracks for general purposes in 3D.

Another important aspect to take into account is related to the surveying of data. The type of the road and the presence of occlusions have a great influence on the results

obtained. In addition, driving behaviour is also important because the usual trend of drivers is to increase the radius of curves. Finally, the analysis of the length of GNSS sections has not shown important differences. In this case, we do not consider it a relevant parameter to be taken into account.

Future work will focus on the specific analysis of the aspects suggested in this study that could affect the positional accuracy of the original GNSS tracks, such as the presence of occlusions caused by vegetation or road cut-slopes. The application to urban areas is another interesting study to be developed due to the potential increment of issues derived from occlusions and multipath effects. Additionally, we could analyse the improvements in accuracy derived from the appearance of new GNSS devices that include new positioning systems (Galileo, etc.).

References

Antoniou, V. and Skopeliti, A., 2015. Measures and indicators of VGI quality: an overview. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1, 345-351.

American Society for Photogrammetry and Remote Sensing (ASPRS), 1990. ASPRS Accuracy Standards for Large-Scale Maps. *Photogrammetric Engineering and Remote Sensing*, 56 (7), 1068-1070.

Ariza-López, F. J., Mozas-Calvache, A. T., Ureña-Cámara, M. A., Alba-Fernández, V., García-Balboa, J. L., Rodríguez-Avi, J. and Ruiz-Lendínez, J. J., 2011. Influence of sample size on line-based positional assessment methods for road data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66 (5), 708-719.

Ariza-López, F.J., García-Balboa, J.L., Ureña-Cámara, M.A. 2015. ES2530686 (A1) - Dispositivo autonivelado para el levantamiento GNSS de elementos lineales. European Patent Office. Available in: https://worldwide.espacenet.com/publicationDetails/biblio?FT=D&date=20150304&DB=&locale=en_EP&CC=ES&NR=2530686A1&KC=A1&ND=4#

Ariza-López, F.J., Mozas-Calvache, A., Ureña-Cámara, M.A. and Gil de la Vega, P. (2019). Dataset of three-dimensional traces of roads. *Scientific Data*. In press

Basiri, A., Jackson, M., Amirian, P., Pourabdollah, A., Sester, M., Winstanley, A., Moore, T. and Zhang, L., 2016. Quality assessment of OpenStreetMap data using trajectory mining. *Geo-spatial Information Science*, 19, 56-68.

Basiri, A., Amirian, P. and Mooney, P., 2016. Using crowdsourced trajectories for automated OSM data entry approach. *Sensors*, 16 (9), 1510.

Berrocoso, M., Paez, R., Jijena B. and Caturla, C., 2006. The RAP net: A geodetic positioning network for andalusia (South Spain), *EUREF Publication*, 16, 364-368.

Biagioni, J. and Eriksson, J., 2012. Inferring road maps from global positioning system traces: Survey and comparative evaluation. *Transportation research record*, 2291(1), 61-71.

Butler, D., 2006. Virtual g1obes: the web-wide world. *Nature*, 439, 776-778.

Cao, L. and Krumm, J., 2009. From GPS traces to a routable road map. *In: 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM SIGSPATIAL GIS 2009)*, 4-6 November 2009, Seattle, WA, 3-12.

Cooper, A., Coetzee, S., Kaczmarek, I., Kourie, D., Iwaniak, A. and Kubik, T., 2011. Challenges for quality in Volunteered Geographical Information. *In: Proceedings of the AfricaGEO 2011 Conference*, 31 May-2 June 2011, Cape Town, South Africa, 1-13.

Davies, J. J., Beresford, A. R. and Hopper, A., 2006. Scalable, distributed, real-time map generation. *IEEE Pervasive Computing*, 5(4), 47-54.

Edelkamp, S. and Schrödl, S., 2003. Route planning and map inference with global positioning traces. *In:* R. Klein, H.-W. Six, L. Wegner, eds. *Computer science in perspective*. Berlin, Heidelberg: Springer, 128-151.

Gil de la Vega, P, Ariza-López, F.J. and Mozas-Calvache A.T., 2016. Problemas que presentan las trazas GNSS procedentes de información geográfica voluntaria. *Geofocus*, 17.

Goodchild, M. F., 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal,* 69 (4), 211-221.

Haklay, M. and P. Weber, 2008. OpenStreetMap: User-generated street maps. *IEEE Pervasive Computing*, 7 (4), 12-18.

Huang, R., Huang, C., Shan, J., Xiong, L. and Yan, J., 2013. Evaluation of GPS trajectories on VGI and social websites. *In: IEEE 21st International Conference on Geoinformatics*, 20-22 June 2013, Kaifeng, China. 1-5.

Huang, H., Zhang, L. and Sester, M., 2014. A recursive Bayesian filter for anomalous behavior detection in trajectory data. *In:* Huerta J., Schade S. and Granell C., eds. *Connecting a Digital Europe Through Location and Place*. Cham, Switzerland: Springer International Publishing. 91-104

International Standard Organization ISO, 1999. *ISO-15046-13: Geographic Information-Quality principles. ISO/TC 211/WG 1*. Geneva, Switzerland: International Standard Organization.

Ivanović, S. S., Raimond, A. M. O., Mustière, S. and Devogele, T., 2016. Detection of outliers in crowdsourced GPS traces. *In: Proceedings of Spatial Accuracy 2016*, 5-8 July 2016, Montpellier, France, 130-135.

Liu X., Biagioni J., Eriksson J., Wang Y., Forman G. and Zhu Y., 2012. Mining large-scale, sparse GPS traces for map inference: Comparison of approaches. *In: Proceedings of KDD12*, 12-16 August 2012, Beijing, China.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. *In: Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1(14), 281-297.

Mozas-Calvache, A. T. and Ariza-López, F. J., 2011. 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.

Mozas-Calvache, A. T. and Ariza-López, F. J., 2015. Adapting 2D positional control methodologies based on linear elements to 3D. *Survey Review*, 47 (342), 195-201.

Mozas-Calvache, A. T., 2016. Analysis of behaviour of vehicles using VGI data. *International Journal of Geographical Information Science*, 30 (12), 2486-2505.

Mozas-Calvache, A. T. and Ariza-López, F. J., 2017. An iterative method for obtaining a mean 3D axis from a set of GNSS traces for use in positional controls. *Survey Review*, 49 (355), 277-284.

OpenStreetMap (OSM), 2018. *OSM-Wiki* [online] Available from: http://wiki.openstreetmap.org/wiki/ [Accessed 29 Oct 2018].

Senaratne, H., Mobasheri, A., Ali, A. L., Capineri, C. and Haklay, M., 2017. A review of volunteered geographic information quality assessment methods. *International Journal of Geographical Information Science*, 31 (1), 139-167.

White, C. E., Bernstein, D. and Kornhauser, A. L., 2000. Some map matching algorithms for personal navigation assistants. *Transportation Research Part C: Emerging Technologies*, 8(1), 91-108.

Zhang, L., Thiemann, F. and Sester, M., 2010. Integration of GPS traces with road map. *In: Proceedings of the Second International Workshop on Computational Transportation Science*, 3-5 November 2010, San Jose, CA, 1-6.

Figure captions:

Figure 1. Method developed in this study.

Figure 2. Zone of study: a) roads used in this study (background: PNOA orthoimage provided by Instituto Geográfico Nacional of Spain); b) views of roads (views provided by Google Earth); c) longitudinal profile of the route.

Figure 3. GNSS surveys: a) GNSS navigator used in tracks survey; b) Leica GNSS 1200+ used in the accurate survey.

Figure 4. Examples of datasets used in this study: a) Columbus GNSS tracks (69 of go direction and 69 of return direction); b) Columbus go and return linestrings and mean axis for 69 tracks; c) Leica accurate survey and mean reference axis obtained. (Background: PNOA orthoimage provided by Instituto Geográfico Nacional of Spain).

Figure 5. Examples of the comparison of axes obtained (reference axis vs. Columbus axis [using 69 GNSS tracks]): a) road N323; b) road JV3231; c) road JV2227. (Background: PNOA orthoimage provided by Instituto Geográfico Nacional of Spain).

Figure 6. Examples of XY and Z differences obtained between vertexes of reference axis and Columbus axis: a) and b) XY differences; c) and d) Z differences (Background: PNOA orthoimage provided by Instituto Geográfico Nacional of Spain).

Figure 7. Results of VIM displacements (2D and 3D) of complete tracks depending on the number of tracks: Mean VIM value of all tracks (69) on yellow line (as reference), mean value of VIM displacement on continuous black line and the dispersion given by the standard deviation values on discontinuous black lines.

Figure 8. Results of VIM displacements (2D and 3D) by sections and depending on the number of tracks.

Figure 9. Results of VIM displacements (2D and 3D) by length of tracks and depending on the number of tracks.