Detection of outliers in sets of GNSS tracks from volunteered geographic information

Paula Gil de la Vega
University of Jaen
Department of Cartographic Engineering, Geodesy and Photogrammetry,
Campus Las Lagunillas, Edificio A3, Jaén (España), 23071
pgil@ujaen.es

Francisco J. Ariza-López
University of Jaen
Department of Cartographic Engineering, Geodesy and Photogrammetry,
Campus Las Lagunillas, Edificio A3, Jaén (España), 23071
fjariza@ujaen.es

Antonio T. Mozas-Calvache
University of Jaen
Department of Cartographic Engineering, Geodesy and Photogrammetry,.
Campus Las Lagunillas, Edificio A3, Jaén (España), 23071
antmozas@ujaen.es

Abstract

The use of navigation systems has spread widely, so you can dispose of bulky sets of traces GNSS from voluntary contributions (Volunteered Geographic Information, VGI). The data obtained by the GNSS navigators can be influenced by many factors that affect their accuracy (time of day, skyline of satellites, weather conditions, etc.). Therefore, a set of traces from VGI may have a different behaviour that must be analyzed. This paper presents a methodology to detect and remove outliers' tracks. The availability of the methodology have been tested over a real set of tracks.
Keywords: VGI, outlier tracks, mean axis.

1 Introduction

The accuracy of GNSS tracks can be affected by many factors which affect to the usability of the data. The exploitation of these data sets is complex due to both volume and variability of data. With the methodology presented below we can debug data as before being used in other processes (e.g. positional quality control, updating mapping, etc.).

There are many studies dealing with outliers in spatial data (e.g. in positional quality, in treating specific observations GNSS data), but the treatment of atypical sets of GNSS tracks is novel. Outlier tracks GNSS, when do not denote an error measurements but a different way, might be interesting to analyze suspicious behavior in groups of people [1], for detecting anomalous taxi routes [2], etc.

In this paper we develop a methodology for the visual and statistical analysis of the behavior of a set of GNSS tracks for detecting outliers and anomalous situations.

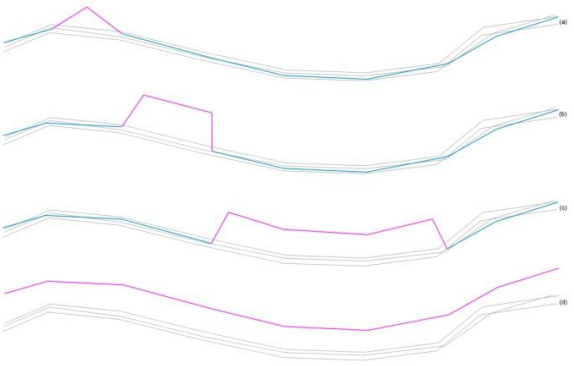
2 Outliers in GNSS tracks

A set of 3D GNSS tracks (SoT) is a set (T) of tracks (t_i) composed by 3D waypoints (V_j). All tracks from T were captured with GNSS equipment on the same linear element, so have a nearby beginning and ending.

There are errors that have a significant effect on the position solution on GNSS tracks, as device and multipath errors, ionospheric and tropospheric delays, etc. Outliers in GNSS tracks can appear as [3]:

- Points (Figure 1.a): when a single waypoint of a track differs from the general path of the rest of the tracks.
- Segment or multisegment (Figure 1.b and Figure 1.c): when two or more consecutive waypoints of a track differ from the general path of the rest of the tracks.
- Track (Figure 1.d): when an entire track differs from the general path of the rest of the tracks.

Figure 1: Classification of track outliers.



3 Outlier detection methodology

The outline of the methodology is contained in the Figure 2. The steps are:

- **Mean 3D axis calculation**: Outliers are determined in an iterative process based on a threshold marking the distance to an average value. Therefore, the first step is to determine a median axis as reference for the detection of outliers. For this we have developed an algorithm based on the intersection of the tracks with planes along

the path. With each plane intersections we determine a central position. Joining the central positions in each plane we form the mean 3D axis.

- **Outlier waypoints detection**: We determine if the waypoints of the tracks are outliers. A waypoint is an outlier if the distance to the mean axis exceeds a given threshold.
- **Outlier waypoints classification**: When outlier waypoints are detected, we analyze the frequency of outlier waypoints in a track and whether or not they are consecutive, and we classify outliers as we saw above.
- **Outliers elimination**: This methodology is iterative. We eliminate the outliers and repeat all the steps until there are not outliers. The process of removing is carried out as follows:
    - Point: we remove the affected waypoint and conserve the rest of the track (Figure 3.a).
    - Segment: we remove the affected segment and conserve the rest of the track (Figure 3.b).
    - Multisegment: we remove the affected segment and the track is divided into two tracks, one on each side of the affected multisegment (Figure 3.c).
    - Track: we remove the whole track (Figure 3.d).

4 Case of study

The methodology has been tested on a SoT (Figure 4) from a section of approximately 9km of the route "Vía Verde del Aceite".

The accuracy of the equipment normally used by the volunteers may vary between 5 to 30 meters approximately. According to these values we have established a threshold of 20m.

We have determined a mean 3D axis and we have calculated distance between waypoints and this axis. The waypoints with distances greater than this threshold are considered outliers. In the first iteration, we have found that 22.85% of the waypoints are outliers, classified as shown in Table 1. We have removed these outliers as it is described in Section 4 and repeated the process. In the second iteration, the percentage of waypoints affected is 0,45%, classified as shown Table 1. We have removed these waypoints again and repeated the process. In the third iteration we have not found outlier waypoints, so the process has finished.

Table 1: Outlier classification.

Type of outlier	Number of outliers 1st iteration	Number of outliers 2nd iteration
Waypoint	12	2
Segment	6	2
Multisegment	34	4
Track	4	0

At the end of the process, the SOT is composed by 25 tracks instead of the 32 tracks of the original set.

5 Conclusions

There are many studies on the detection of outliers in spatial data (e.g. in positional quality in treating specific observations GPS data), but the detection of outliers in sets of GNSS tracks is novel in the VGI treatment. In this paper we have developed a methodology that allows the detection and elimination of outliers for any set of tracks. The application on a real dataset has shown the feasibility of the method.

6 Aknowledgements

This work has been funded by the Ministry of Science and Technology of Spain and the European Regional Development Fund under grant no. BIA2011-23217. The authors also acknowledge the Regional Government of Andalusia (Spain) for the financial support since 1997 for their research group (Ingeniería Cartográfica) with code PAIDI-TEP-164.

References

[1] V. C. Fontes y V. Bogorny, «Discovering Semantic Spatial and Spatio-Temporal Outliers form Movin Object Trajecotires,» *CoRR*, pp. 1-28, 2013.

[2] C. Chen, D. Zhang, P. S. Castro, N. Li, L. Sun y S. Li, «Real-time Detection of Anomalous Taxi Trajectories from GPS traces,» *Transactions on Intelligent Transportation Systems*, vol. 14, nº 2, pp. 806-818, 2013.

[3] F. J. Ariza-López, J. Rodríguez-Avi y J. F. Reinoso-Gordo, «An approximation to outliers in GNSS traces,» de *Spatial Accuracy 2014*, Michigan, 2014.

Figure 2: Outlier detection methodology.

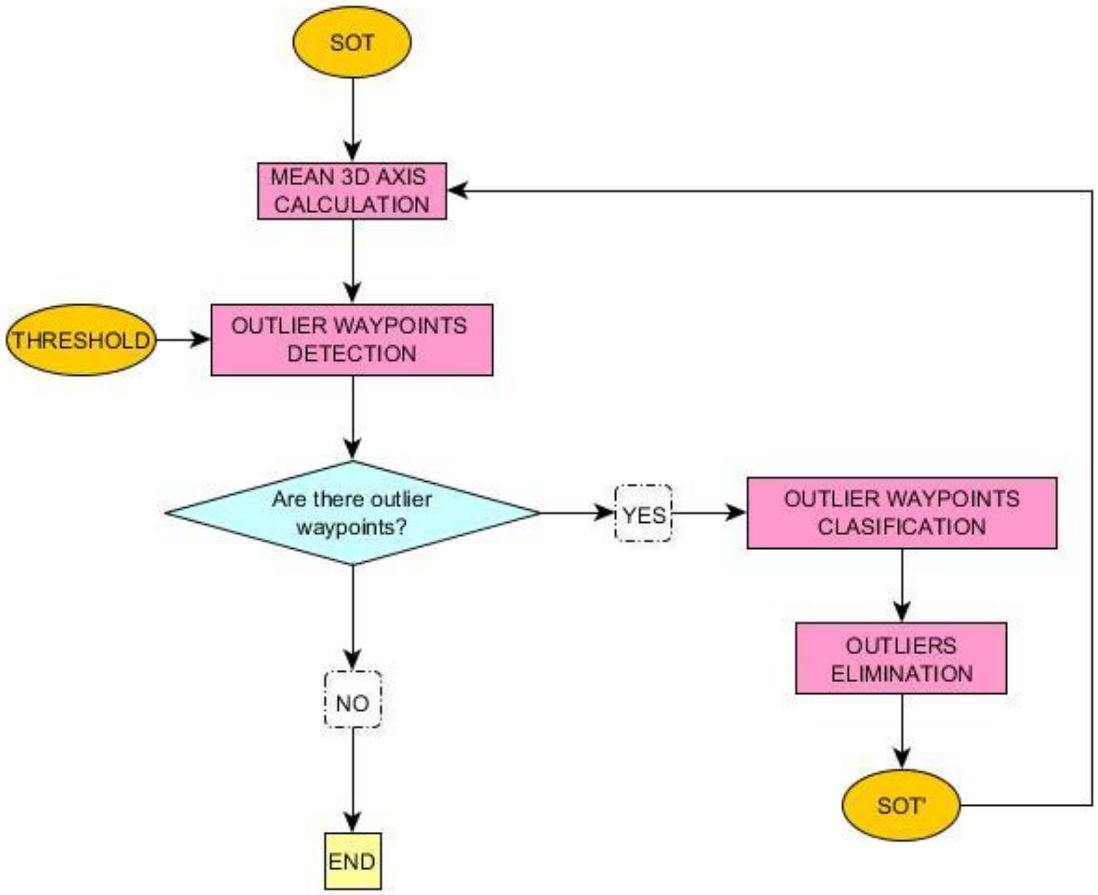


Figure 3: Outliers elimination.

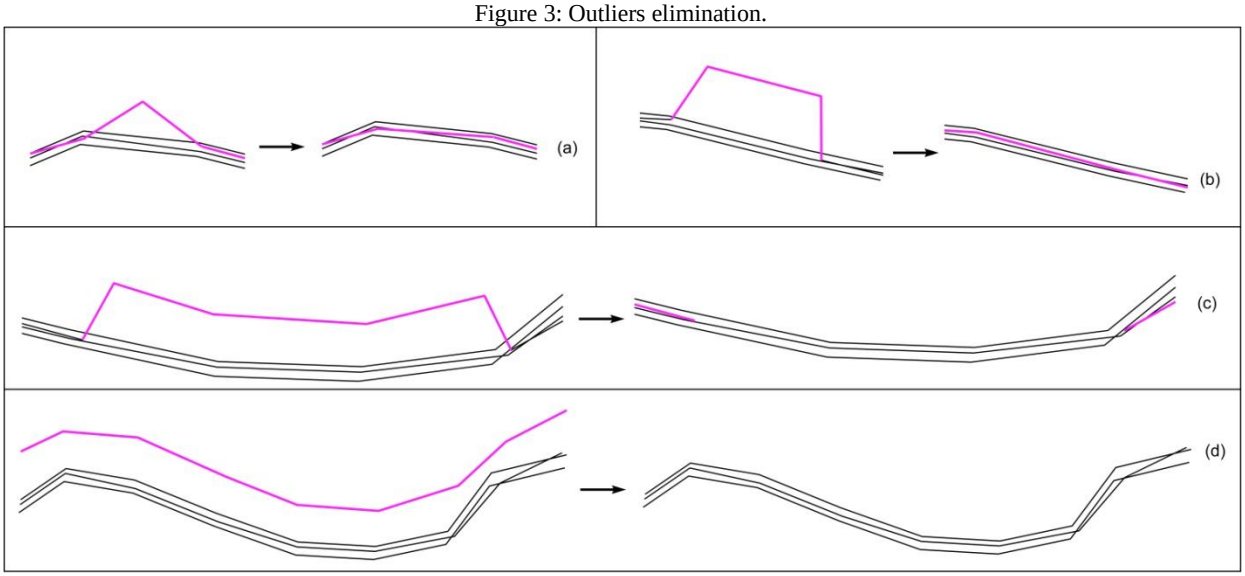


Figure 4: XY representation of the SOT downloaded.