An approximation to outliers in GNSS traces

Ariza-López, FJ[1]*; Rodríguez-Avi, J*[1]*; Reinoso-Gordo, JF*[2]

[1] Universidad de Jaén (España), [2] Universidad de Granada (España).

Abstract

The issue of outliers in GNSS-traces set is presented. A basic classification of outlier situations in three cases is proposed. Some graphical tools (box-plots, bag-plots, etc.) are explored in order to detect this situation.

Keywords: GNSS multitraces, outlier detection, positional accuracy.

1. Introduction

Multitraces of GNSS devices (3D line strings) are currently an important data input for automatic road maps generation. In order to obtain a valid road axis, or the center of a lane or track, mining methods (e.g. k-means algorithm, trace merging and the kernel density estimation) are needed. But a problem that arises before the determination of a mean axis is the detection and treatment of outliers (see outlier situations in Figure 1).

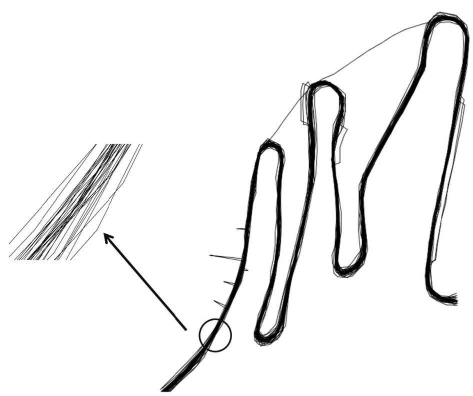

Figure 1: Example of GNSS traces with different outliers situations.

An outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism (Hawkins 1980). From a statistical perspective outliers have been seen as a problem, but they are of great value for some cases (e.g. fraud detection, public health, business intelligence, etc.). There are many approaches (model based, proximity based, specific solutions) and algorithms and statistical test or criteria (Chauvenet's, Grubbs', Peirce's, etc.) for outliers detection, but an important consideration is that there is no rigid mathematical definition of what constitutes an outlier; determining whether or not an observation is an outlier is ultimately a subjective exercise (Wikipedia, 2014).

There are some studies dealing with outliers in spatial data, but none have dealt with outliers in GNSS traces. For this reason, as a first approximation we propose some graphical analysis in order to explore and obtain an intuitive idea of the problem before starting with any analytical or computational method development.

2. Outliers in GNSS traces

We can consider that GNSS trace positions follow a normal generating procedure, and that outliers come from an abnormal generating mechanism. This mechanism is not unique and includes: ionospheric affectations, occlusion situations, multipath and device errors. When observing GNSS traces positional outliers can appear as (Figure 2):

- **Points**. Isolated points within a trace that show a dramatic positional change in relation to their neighbourhood. Instantaneous device errors are the main cause (e.g. receiver's noise).
- **Sections**. A continuous sequence of points of a trace with bias in relation to the main stream of GNSS traces determines an outlier section in a trace. Multipath errors are the main cause (e.g. an embankment with large slopes).
- **Line string**. An entire trace can be considered as an outlier if it separates from the main stream of a GNSS set of traces, for example if there is a mean offset. A line string can be considered an outlier line string if the percentage of length accumulated by out-layered sections is greater than a given tolerance. The main cause is bad configuration, resulting in offsets (e.g. erroneous ionosphere models, etc.) or strange traces.

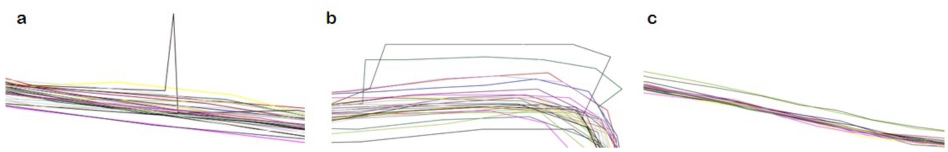

Figure 2: Example of 3 outlier categories in GNSS traces: a) point, b) section, c) entire string.

3. Graphical analysis tools for outliers in GNSS traces

In order to detect outliers in GNSS traces we have two basic sources: position and heading. We can calculate errors (a trace versus the mean trace), increments (between sequential values of a trace), or increments of errors or increments of increments. Any variable (positions, headings and positional or angular errors/increments) can be used in isolated or in combined ways and analyzed by graphics. Graphical tools (e.g. box-plot) are widely used for outlier detection and analysis in statistical analysis because of their simplicity and efficiency. The box-plot or box and whiskers plot is typically used for 1D data but can be easily adapted for 2D and 3D data. Other ideas come from circular statistics where circular representations are used to show the angular arrangement. Sequences of differences between coordinates, errors or headings can also be used to detect changes by means of high pass filters. In all the cases we need to determine previously the mean axis for the entire GNSS-traces set. In this way some possibilities are:

- **Unique plot**. Any kind of box-plot (e.g. box-plot, bag-plots, violin-plots) or circle-plot (e.g. rose-plot) can be used for any variable (positions, headings and positional or angular errors/increments). The complete multitrace set is summarized in a unique plot. Box-plots represent linear dispersions and circle-plots angular dispersions. When possible, concentric layered plots can be used. The problem is the congestion of the representation if many traces are represented, if the length of the traces is great, or if density of points is high. To reduce this problem a selection of cases can be done.

– **Sequence along the rectified mean axis**. A variable or an ordered sequence of plots, of any kind, can be represented along the rectified mean axis. If the scale is the same for all the plots a better understanding can be achieved. See Figure 3 for 2D box-plots series representing the 2D RMSE of a multitrace set. An interesting variable is the scaled increments of absolute differences. For instance, Figure 4 shows scaled differences in coordinates between sequential points for three traces. The peaks denote an outlier situation.

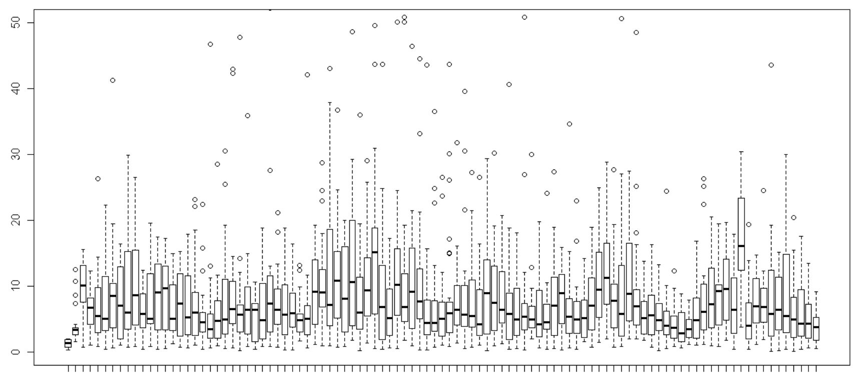

Figure 3: Sequence of box-plots for 3D errors (multitrace set) (X-axis: length, Y-axis: error [m]).

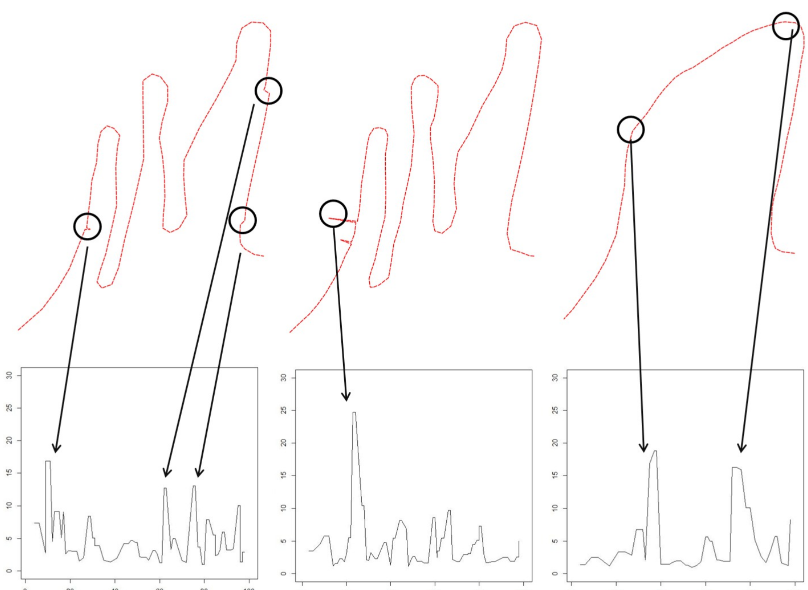

Figure 4: Sequence of scaled increments (GNSS trace) (X-axis: length, Y-axis: dimensionless).

– **Sequence along the mean axis**. A variable or an ordered sequence of plots, of any kind, can be represented along the planimetric or altimetric representation of the mean axis. This case is similar to the previous one but offers richer contextual information (mainly curvature).

- **Polygraph**. Several variables are represented together. The representation can be along the rectified mean axis or along the mean axis. This case offers the possibility of analysing jointly several variables (e.g. 3D position error, heading errors, increments of errors, scaled differences of errors, etc.). This case offers complementary visions together (e.g. position and heading).
- **In situ**. A variable or an ordered sequence of plots can be displayed along the rectified mean axis when traveling through 3D terrain visualization. This case offers richer contextual information (e.g. curvature, presence of embankments, trees and other kind of obstacles). Figure 5 shows is an example.

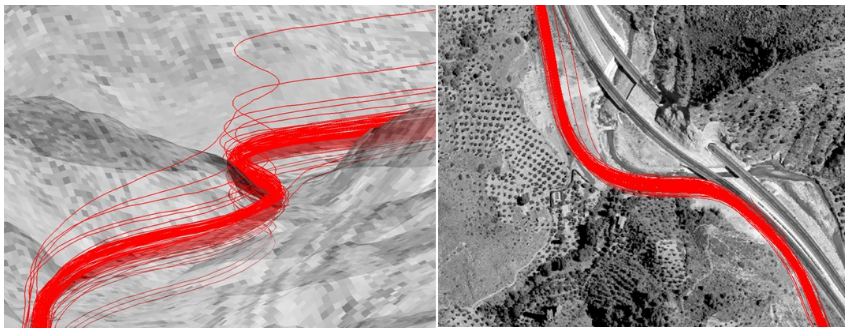

Figure 5: In situ visualization of outliers (3D perspective and planimétrico projection).

3. Conclusion

Outlier detection and analysis is an important issue when working with GNSS multitraces. In this paper we have presented a classification of outliers in this scope and some methods for representing and analysis in order to evidence the presence of outliers in sets of GNSS traces. Graphical methods have been classified with respect to the context information they provide. We believe that the graphical approach to outliers is of interest to obtain a prior intuitive idea to any numerical analysis.

Acknowledgments

This work has been funded by the Ministry of Science and Technology (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 the research groups with code PAIDI-TEP-164 and PAIDI-FQM-245.

References

Hawkins, D. (1980). Identification of Outliers. Chapman and Hall.

Wikipedia (2014). Outlier. http://en.wikipedia.org/wiki/Outlier.