OUTLIER=GROSS ERROR? DO ONLY GROSS ERRORS CAUSE OUTLIERS IN GEODETIC NETWORKS? ADDRESSING THESE AND OTHER QUESTIONS

Stefano Sampaio Suraci[1] - ORCID: 0000-0002-4453-9273

Leonardo Castro de Oliveira[1] - ORCID: 0000-0001-5290-5029

[1] Instituto Militar de Engenharia,
Seção de Engenharia Cartográfica, Rio de Janeiro, Brasil.
E-mail: stefanosampaio@hotmail.com; leonardo@ime.eb.br

Abstract:

This article has theoretically discussed some points regarding outliers caused by errors in geodetic observations (see consideration made). Comments have also been made on the usual *3σ-rule* to identify outliers and its common approachs in the simulation of outliers in geodetic networks. Three simulated experiments have been conducted to verify the elements discussed. In the first one, with the simulation of random errors, we have verified that it can have a magnitude large enough to generate outliers. In the second one, in scenarios of leveling network simulated by Monte Carlo methods, observations containing gross errors with a lower magnitude than their respective σ tended to not be identified as outliers by the iterative data snooping procedure. This has also occurred in the third experiment, in which gross errors of magnitude 3.1σ had their value masked by the random error of the respective observation. From the conceptual discussion presented, we have concluded that gross error and outlier are not synonyms, and neither is one a particular case of the other. From the obtained results, we have concluded that there are inconsistencies in how outliers have been simulated in geodetic networks, which indicates the need to continue with investigations.

Keywords: outlier; gross error; Monte Carlo simulation

1. Introduction

This work focuses on geodetic observations and products, but it can be broadly and comprehensively applied to other areas of knowledge that address measurement errors. Because of the sensitivity of the adjustment by the established Least Squares (LS) method to outliers, they need to have some treatment in order to reduce or even isolate their effects on the parameters to be estimated. Therefore, the treatment of outliers is essential for an appropriate adjustment of geodetic observations.

It is unlikely to guarantee the absence of outliers in data collected in the field in the practice of surveys (Knight et al. 2010). In addition, according to Lehmann (2013b), the cost of completely avoiding gross errors (the most frequent causes of outliers in geodetic networks) is economically unjustifiable, and it is more feasible to accept the risk of some for later treatment. Thus, the

How to cite this article:

treatment of outliers should always be performed. In this context, Gemael, Machado and Wandresen (2015) warn that gross errors can occur even in electronic equipment processing, such as in satellite positioning devices and electronic levels. Bustos (1981) concludes that the existence of a small number of outliers in observations, considering the most diverse applications, seems to be the rule, not the exception.

However, aiming at the treatment of outliers per se, a conceptual analysis of the related terms, their meanings, and aspects is necessary for a consistent and homogeneous understanding about them. As a contribution to this, some questions on the subject were raised in this work, specifically regarding the measurement errors of the observations generating outliers in geodetic networks. That is, it is based on the premise that there are no other problems that may cause outliers, such as wrong mathematical modeling, in the adjustment computations. The issue of configuration weaknesses in geodetic networks (Hekimoglu et al. 2011), relevant in the identification of outliers, was also not analyzed. Hence, in order to isolate the topic of measurement errors in observations, all experiments were performed with correct mathematical modeling and network configuration reliable against existing outliers.

1.1. Outliers x gross errors

Errors in geodetic observations are classified as random, systematic, and gross. Random are the inevitable measurement errors, present in all geodetic observations. Its frequency distribution tends to approximate the normal distribution as sample size increases. Systematic errors address the conditions of the collection of observations that can be modeled and have their effects neutralized, either with collection procedures or with appropriate mathematical models.

Gross errors are those that are neither random nor systematic. They have no known distribution or modeling and can be of any magnitude. They usually occur specifically, from human or equipment failures. Further details on gross, systematic, and random errors can be found in Gemael, Machado and Wandresen (2015) and Ghilani (2010).

In this work, the term "total error" e_T (Equation 1) will be used to denote the sum of the random error e_A (which always exists), systematic error e_S (if any), and gross error e_G (if any) of an observation.

$$e_T = e_A + e_S + e_G \tag{1}$$

The geodetic observations o_G (Equation 2) are composed of the exact value v_E (unknown in the practice of surveys) of the measured quantity and its total error. They can be classified as "good" observations or outliers. With this, the former are all those that are not outliers.

$$o_G = v_E + e_T \tag{2}$$

According to Klein (2011), the most mentioned definition of outlier in the literature is that of Hawkins (1980), by which "an outlier is an observation that deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism". From the definition presented, we can note that an outlier is not a type of error (it is an observation), nor is it related to any specific type of error. Thus, gross, not modeled systematic, and even random

errors (or combinations of them) may in theory be the cause of outliers, as long as the total observation deviation raises the suspicion of a different mechanism. Hence, outlier and gross error are different concepts.

Recalling that random errors are in all geodetic observations, Figure 1 illustrates the positioning of the outliers among them. All combinations of errors can generate outliers, or "good" observations, because the classification criterion is in the discrepancy in relation to the sample, not in the types of error that the observation contains.
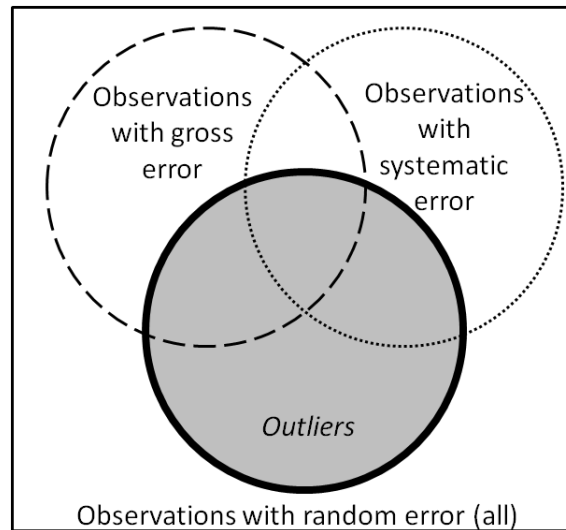
Figure 1: Outliers among geodetic observations.

Despite this, the concepts of outlier and gross error are sometimes confused. According to Lehmann (2013a), this occurs because "in geodesy, outliers are most often caused by gross errors and gross errors most often cause outliers". That is, although the most common occurrences contribute to this conceptual misunderstanding, there are exceptions to this general case. Therefore, in the adjustment of geodetic networks as well, not every outlier is caused by gross errors and not every gross error causes outliers. In addition, according to the author, an outlier is defined as "an observation that is so probably caused by a gross error that it is better not used or not used as it is". Rofatto et al. (2018) adopted the same definition for outlier. Hence, an outlier may or may not contain a gross error, since the criterion is the probability of this occurring, not the occurrence in fact. In this sense, it is in agreement with Hawkins' definition: the great probability of gross error can be seen as the fact that raises the suspicion that a different mechanism is involved (Lehmann 2013b).

In fact, the probability of a gross error (and not the actual occurrence) is considered because it is not possible to determine, without uncertainty, the occurrence of a gross error in the practice of surveys. Baarda (1968), who is the main reference on the subject related to geodetic applications, recognizes that "because of the random character of observations it is impossible to signalize gross errors with certainty". At the risk of being repetitive, it is fundamental for the chaining of ideas to perceive the subtle - but relevant - difference between an observation "having a high probability of containing gross error" (raising suspicion of a different mechanism) and "containing a gross error". In addition, it is reasonable to imagine that observations with relatively larger total error tend to be more likely to contain a gross error, even if they do not contain it.

Thus, for example, an observation without a gross error and without a systematic error may have significant discrepancy because of a large deviation in its random error, which may generate an outlier. Although unlikely, this is not impossible. We emphasize that random errors of observations tend to have normal distribution, whose density function, although relatively larger in the regions closer to the zero abscissa, is defined and not zero throughout the space $\mathbb{R}$ of the real numbers. In addition, an observation with no systematic effects and with random and gross errors of relatively small magnitude probably will not be discrepant, nor will it have a high probability of gross error, thus not setting an outlier, although containing a gross error. These exceptions to the general case mentioned above were addressed in the experiments of this paper.

1.2 More objective considerations on outliers

The definitions of outliers presented are subjective. More objectively, for the one-dimensional case, one of the options adopted in the sciences in general is to take as outliers the observations that deviate more than a certain amount of their standard deviation σ around its respective mean μ. This is based on the characteristics of normal distribution, which presents 99.73% and 95.44% of the data, for the $\mu\pm3\sigma$ and $\mu\pm2\sigma$ intervals, respectively, which are the two of the most used. The adoption of the first interval for identifying outliers is usually called the *3σ-rule*. According to Leys et al. (2013), in this context, although there are authors who defend different amounts of σ for the limit of deviation, this choice depends on the situation and the perspective defended by the researcher. However, it is noteworthy that, also in this objective approach to the outlier term, corroborating the conclusion of the previous section, the classification of an observation as outlier does not depend on the type of error it contains but on the magnitude and stipulated accepted tolerance of the deviation of its total error.

In the estimation of geodetic networks, already addressing multidimensional cases, Baarda (1968) pioneered the placement of an objective approach to the subject. He developed the data snooping (DS), a statistical hypothesis test to identify gross errors in observations. The observations rejected by the test were classified as outliers. Thus, for example, an observation without a gross error, but rejected in the test, is also considered an outlier. This makes sense according to the definition given by Lehmann (2013a), because, when rejected in the test, the observation, even if it did not contain a gross error, presented a high probability of having one and should therefore be classified as an outlier.

In the last 50 years, other outlier identification methods have emerged in the geodetic literature, through other statistical tests or through robust estimation techniques. A comparison of 7 of the most common can be seen in Klein et al. (2015). However, even today DS is considered to be one of the best outlier identification methods in geodetic networks (Rofatto, Matsuoka and Klein 2017), and it is also the most used in commercial softwares and recommended in related textbooks (Lehmann and Losler 2016).

Nevertheless, there is a continuous research on outliers in geodetic networks. Among others, in addition to those papers already mentioned in this article, we can also cite: (Klein, Matsuoka and Souza 2011), (Baselga 2011), (Hekimoglu, Erdogan and Tunalioglu 2012), (Klein et al. 2012), (Hekimoglu and Erdogan 2013), (Klein, Matsuoka and Monico 2013), (Erdogan 2014), (Guo

2015), (Klein, Matsuoka and Guzatto 2015), (Zhao and Gui 2017), (Rofatto, Matsuoka and Klein 2017), (Teunissen 2018) and (Rofatto, Matsuoka and Klein 2018).

In fact, DS itself was extended over the years. One of its adaptations in the case of multiple hypotheses is the iterative data-snooping (IDS) (Teunissen 2006), which identifies one outlier at a time and then restarts without the outliers already identified until none are found. For a review of the half-century of DS and its variations we recommend (Rofatto et al. 2018). As in this last reference, the IDS was applied in the experiments of the present article.

In order to evaluate the efficacy of the identification methods, outliers are generated in the geodetic networks analyzed by the intentional insertion of gross errors with "large" magnitude, enough to make the respective observation discrepant. Thus, the method succeeds if it correctly identifies the alleged outliers. Recalling the fact that most outliers in geodesy are caused by gross errors, although this procedure does not cover all possible outlier cases, it corresponds to the most common. For example, among others, it can be found in (Amiri-Simkooei 2003) and (Yetkin and Berber 2013).

In this context, authors usually consider $3*\sigma_i$, being σ_i the respective standard deviation of the observation, as the minimum magnitude of the gross error to be "large" enough to cause an outlier. This approach will be called in this work as *3σ-rule for gross error*. An example of the insertion of outliers by the *3σ-rule for gross error* for the evaluation of outlier identification methods can be found in Klein et al. (2015). It is similar to the *3σ-rule* presented above, but it applies the 3σ limit to a possible gross error of the observation, not to its total error.

Rofatto et al. (2018) and Hekimoglu and Erenoglu (2007) used another similar but different approach. In their simulations, they regarded as outliers the observations with gross error over 3σ and no random error. Hence, they guaranteed that the outliers had gross and total error over 3σ, being the latter in accordance to the *3σ-rule*. However, they considered observations with only random errors not to be outliers, even if these errors were above 3σ, something debatable that was also tested in the experiments. This last point is not in agreement with the *3σ-rule*.

Mean Successful Rate (MSR) is the most usual index to measure the performance of a method in identifying outliers. It represents the number of scenarios in which the method succeeded divided by the total of scenarios tested. With the increasing use of Monte Carlo simulations in the context of the identification of outliers in geodesic networks, for a more complete analysis, the measurement of the following indices has also become usual, as done by Rofatto, Matsuoka and Klein (2017) and Rofatto et al. (2018):

1) Type I error (%): probability of identifying an outlier when there is none;

2) Type II error (%): probability of non-identifying an outlier when there is at least one;

3) Type III error (%): probability of misidentification a non-outlying observation as an outlier, instead of the outlying one;

4) Over-identification+ (%): probability of identifying correctly the outlying observation and others;

5) Over-identification- (%): probability of identifying more than one non-outlying observation, whereas the "true outlier" remains in the dataset.

1.3 Exceptions to the *3σ-rule for gross error*

From the content presented, all types of errors and their combinations can generate outliers. However, when adopting the *3σ-rule for gross error*, it is considered that outliers are caused by gross errors greater than 3σ (in module). Thus, for example, two situations that seem to deserve more attention are ignored: 1) the random error may be discrepant to the point of generating an outlier, and 2) an observation may not be very likely to contain a gross error, even if it contains a gross error with an absolute value greater than 3σ. Actually, the first is an exception not only to the *3σ-rule for gross error*, but also to the approach presented of Rofatto et al. (2018) and Hekimoglu and Erenoglu (2007) to the *3σ-rule*.

As for the first point raised, the area under the standard normal curve limited by ± 3σ is 0.9973. Thus, even if there are no gross or systematic errors, there is a probability of 0.27% for the total error of a single observation extrapolating the ± 3σ because of its random error. In this case, the observation has a discrepant error (Figure 2), which suggests that it should be classified as an outlier (which would not be the case in the *3σ-rule for gross error*) although in reality (unknown in practice) it only contains a random error.
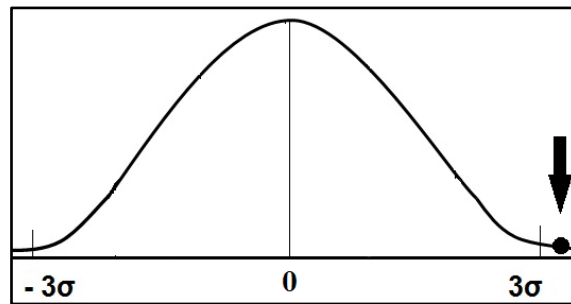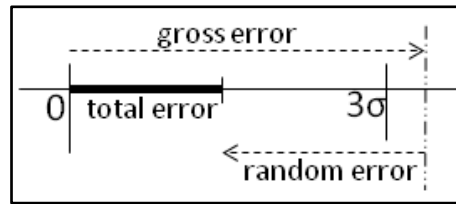

Figure 2: Discrepant random error.

It is worth noting that despite the relatively small probability (0.27%) for a single observation, in scenarios with a relatively high number of observations, such as the High Precision Altimetry Network of the Brazilian Geodetic System, the probability **p** of the random error of at least one observation having an absolute value greater than 3σ increases significantly. For example, for networks with 1,000, 2,000, and 10,000 observations, this probability is already 93.30%, 99.55%, and approximately 100% (Equation 3), respectively.

$$p_{1,000} = 1 - (99.73/100)^{1,000} = 93.30\%$$

$$p_{2,000} = 1 - (99.73/100)^{2,000} = 99.55\% \qquad (3)$$

$$p_{10,000} = 1 - (99.73/100)^{10,000} \cong 100.00\%$$

In addition, as for the second point raised in the antepenultimate paragraph, for example, an observation with a gross error of 3.1σ (above 3σ, i.e., which would be considered an outlier by the *3σ-rule for gross error*), but with a random error of 2.0σ of opposite sign, has a total error (1.1σ in module) that is not discrepant (Figure 3) by the limit considered of 3σ. Thus, it is expected to "not raise the suspicion of a different mechanism" and "not have a high probability of gross error", and should not therefore be considered an outlier. This was tested in the experiments of this paper.

Figure 3: Non-discrepant total error despite gross error above 3σ.

2. Methodology

For the proof of the elements discussed, the experiments were constructed in such a way as to show that: 1) discrepant observations can be caused by random errors, that is, not only by gross errors, 2) observations with a gross error of "small" magnitude tend not to present a high probability of containing a gross error, and 3) the previous item is also valid for observations with a gross error whose magnitude is above but close to 3σ and with a random error of opposite sign to the gross error.

In the first experiment, 1,000 different samples of 1,000, 2,000, and 10,000 random errors were simulated. From the mean occurrence of discrepant random errors, considering the magnitude limit of 3σ, we intended to verify that the random error of an observation can be discrepant and that tends to occur to 0.27% of the observations. Moreover, from the comparison between the means for 1,000, 2,000, and 10,000 random errors, we intended to demonstrate that the larger the number of observations, the greater the mean number of occurrences of discrepant random errors in the sample. We also computed the number of random errors of magnitude greater than 4σ and 5σ to verify the occurrence of "very" discrepant random errors.

The second experiment shows the case of the gross error having a relatively "small" magnitude, lower than the standard deviation σ_i of the observation itself. From **n** exact elevation differences between the vertices of an altimetric network, n*100 altimetric network scenarios with only random errors in their observations were simulated by Monte Carlo methods, as in the heteroscedascity case (different weights for observations) of "observations without outliers" of Hekimoglu and Erenoglu (2007). Then, a gross error with the specified characteristic was added separately for each of the **n** simulated observations of each of the n*100 network scenarios, amounting to n*n*100 scenarios with a gross error to be evaluated. The sign of the gross error was also randomly chosen. The IDS was applied to each of these scenarios to verify the occurrence of observations with a high probability of containing a gross error, which the method identifies as outliers. As in the network to be analyzed n=20, then 20*20*100=40,000 scenarios were tested. In the context of the identification of outliers, with this number of scenarios the obtained results had a standard deviation of about 0.21%, in the accuracy analysis as a function of the number of simulations performed by Rofatto et al. (2018).

For comparison, we applied IDS to the same 40,000 scenarios, but with no random error in the observations in which the "small" gross error was inserted. In fact, the random error was always replaced by the gross error, as in the case of simulations of "bad observations" with random sign for gross errors performed by Hekimoglu and Erenoglu (2007). Also for comparison, we applied IDS to 40,000 scenarios simulated by Monte Carlo methods with only random errors in their observations, as "observations without outliers" of Hekimoglu and Erenoglu (2007). This was part 1 of Experiment 2.

In part 2 of Experiment 2, a case of systematic error of relative "small" magnitude was verified. Also in a scenario of "observations without outliers", instrumental errors of small magnitude in common for some observations of the network were simulated. IDS was applied to this scenario to check if these errors would cause the identification of outliers.

The third experiment was similar to part 1 of the second experiment, but with a gross error of magnitude of 3.1σ (over 3σ) being added separately for each of the **n** simulated observations of each of the n*100 network scenarios, also amounting to n*n*100 scenarios with a gross error to be evaluated. At first, it was tested with the sign of the gross error always opposite to the random error one of the respective observation. The IDS was applied to each of these scenarios. Then, for comparison, we applied IDS in the same scenarios, but with no random error in the observations in which the gross error was inserted.

The experiments were conducted in the software Octave (GNU, 2018). The simulation of random numbers normally distributed was done using the *randn* routine of the same software. The reader can contact the authors to obtain the codes of the experiments.

3. Results and discussion

3.1 Experiment 1

Table 1 shows the mean number and percentage of random errors with magnitude greater than 3σ for 1,000 different samples of 1,000, 2,000, and 10,000 random errors. In addition to clarifying that in fact random errors in module can extrapolate the limit of 3σ, the percentage in the three cases was approximately 0.27% of the respective samples, which is consistent with the theoretical value. As emphasized by Rofatto, Matsuoka and Klein (2017), it represents the probability of type I error of the local test for a single alternative hypothesis (level of significance **α**) which is different from the probability of type I error of the IDS. Moreover, from these results we can infer that, the larger the number of observations, the greater the tendency of the mean number of occurrences of discrepant random errors in the sample.

Table 1: Mean and percentage of discrepant random errors in 1,000 samples.

Random errors per sample	Mean number above 3σ	Percentage (%)
1,000	2.75	0.27
2,000	5.42	0.27
10,000	27.23	0.27

Table 2 presents the number of random errors of magnitude greater than 4σ and 5σ for the same samples. In all cases there were "very" discrepant errors above 4σ and even above 5σ. Obviously, the undesired effect for the results of the adjustment of an observation with a total error of magnitude greater than 5σ is the same regardless of whether its total error is only a random error or also contains a gross error. The case of 5σ has been quoted to draw more attention from the reader to the possible discrepancy of a random error, but the same holds true in case 3σ is the stipulated limit. Therefore, although usually not interpreted in this way in the geodetic literature, it is also reasonable to understand that the minimum error magnitude

stipulated to classify an observation as outlier should not only apply to observations with gross error, but for any observation of the network, regardless of the types of error it contains. This would be in accordance to the *3σ-rule* presented.

Table 2: Quantity of "very" discrepant random errors in 1,000 samples.

Random errors per sample	Quantity above 4σ	Quantity above 5σ
1,000	58	1
2,000	122	1
10,000	644	8

3.2 Experiment 2

Figure 4 and Table 3 respectively show the geometry and exact elevation differences h_i of the initial altimetric network lines, which are the basis for the simulation of scenarios by the Monte Carlo technique as in the heteroscedascity case (different weights for observations) of Hekimoglu and Erenoglu (2007). During the peer-review process of the current article, Rofatto et al. (2018) applied Monte Carlo simulations without the need for initial exact elevation differences. This is only an operational detail that does not invalidate our results using elevation differences as performed by Hekimoglu and Erenoglu (2007). The standard deviation of observations with simulated random errors was given by Equation 4, where **K** (in km) is the length of the respective line.

$$\sigma_i = 1.0(mm) * \sqrt{K_i} \tag{4}$$
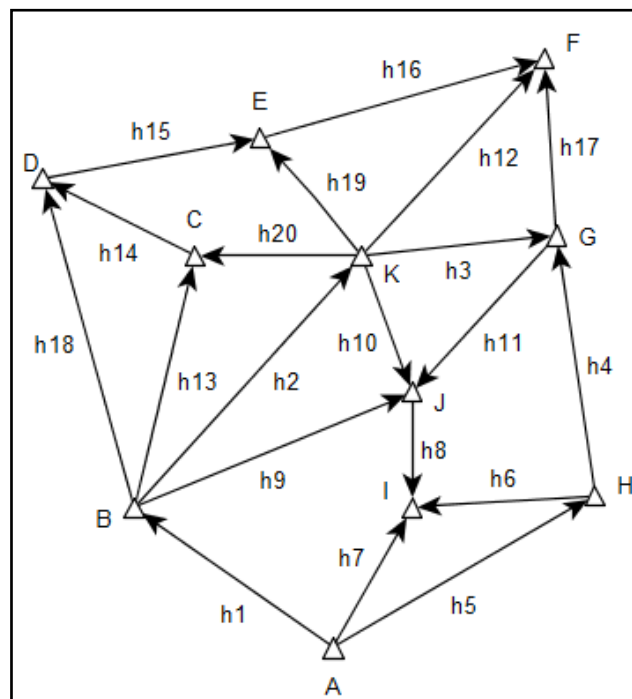
Figure 4: Geometry of the leveling network.

Table 3: Elevation differences.

h_i	Elevation difference (mm)	Distance (km)	h_i	Elevation difference (mm)	Distance (km)
h_1	163854.9	49	h_{11}	110227.2	62
h_2	6446.2	41	h_{12}	155928.2	50
h_3	57037.0	38	h_{13}	52875.0	35
h_4	126209.5	34	h_{14}	62904.2	43
h_5	101128.6	22	h_{15}	3889.5	20
h_6	296885.8	13	h_{16}	42705.7	28
h_7	398014.4	23	h_{17}	98891.2	19
h_8	60449.1	48	h_{18}	115779.2	39
h_9	173710.4	15	h_{19}	113222.5	27
h_{10}	167264.2	24	h_{20}	46428.8	21

In part 1 of Experiment 2, a typing error (gross error) of one unit in tenths of mm and random sign was added separately for each of the 20 observations in each of the 2,000 (20*100) altimetric network scenarios simulated by Monte Carlo methods with only random errors in their observations, totaling 40,000 (20*20*100) simulated scenarios. Considering that the respective observation was an outlier, Table 4 presents success and errors rates by IDS. The alleged outlier was identified by IDS (with α=0.0027) as an outlier in only 0.19% (sum of MSR and Over-identification+) of the scenarios evaluated. Thus, we can note that the classification of an observation that contains a gross error as an outlier also depends on the value of the first, which is not always sufficiently "large" to provoke the second. This confirms that observations with a gross error of "small" magnitude tend not to be very likely to contain a gross error and are therefore not identified as outliers. Besides, at least one outlier was identified in only 3.77% (100% minus Type II Error) of the scenarios evaluated.

Table 4: Experiment 2 – alleged outlier with random error and "small" gross error

MSR	Type II Error	Type III Error	Over-identification+	Over-identification-
0.18%	96.23%	3.54%	0.01%	0.04%

For comparison, we applied IDS to the same 40,000 scenarios, but with no random error in the observations in which the gross error was inserted. Considering that the respective observation was an outlier, Table 5 presents success and errors rates by IDS. The alleged outlier was not identified in any of the 40,000 scenarios, i.e., the simple occurrence of a gross error was not enough to cause outliers due to its "small" magnitude. Besides, at least one outlier was identified in only 3.29% (100% minus Type II Error) of the scenarios evaluated.

Table 5: Experiment 2 – alleged outlier with "small" gross error but no random error

MSR	Type II Error	Type III Error	Over-identification+	Over-identification-
0.00%	96.71%	3.26%	0.00%	0.03%

Also for comparison, we applied IDS to 40,000 scenarios simulated by Monte Carlo methods with only random errors in their observations. In Table 6, we can see that at least one outlier was identified in 4.79% (more than 3.77% and than 3.29%). This shows that scenarios with "small"

gross error in this experiment had even less outliers identified than scenarios with no gross errors at all, confirming that not always the occurrence of gross error is essential to cause outliers.

Table 6: Experiment 2 – all observations with random error but no gross error

At least one outlier identified	No outliers identified
4.79%	95.21%

In part 2 of Experiment 2, instrumental errors common to part of the observations (systematic errors) of one unit in tenths of mm were added to 6 (of 20) observations in a simulated scenario with only random errors. The observations with only random errors were also simulated in the same way as the "observations without outliers" of Hekimoglu and Erenoglu (2007). The systematic errors were added to the simulated observations correspondents to h_4, h_5, h_6, h_7, h_8 and h_9. No outlier has been identified by IDS, which shows that even systematic errors may not cause outliers. After that, we performed the same test, but with no random errors in the observations that systematic errors were inserted. Again no outliers were identified, confirming this conclusion.

3.3 Experiment 3

From the same 2,000 altimetric network scenarios with only random errors in their observations of part 1 of Experiment 2, a gross error of magnitude of $3.1*\sigma_i$ and opposite sign to the random error of the respective observation was inserted separately for each of the 20 observations of each scenario, totaling 40,000 scenarios simulated. Considering that the respective observation was an outlier, Table 7 presents success and errors rates by IDS. The alleged outlier was identified by IDS (with $\alpha=0.0027$) as an outlier in only 4.01% (sum of MSR and Over-identification+) of the 40,000 (20*20*100) scenarios evaluated. This also confirms the tendency of low probability of being identified as outliers the observations with gross error above (but close to) 3σ and random error of opposite sign to the gross error.

Table 7: Experiment 3 – alleged outlier with gross error and random error of opposite sign

MSR	Type II Error	Type III Error	Over-identification+	Over-identification-
3.98%	89.54%	6.39%	0.03%	0.06%

For comparison, we applied IDS to the same 40,000 scenarios, but with no random error in the observations in which the gross error was inserted. Considering that the respective observation was an outlier, Table 8 presents success and errors rates by IDS. The MSR was more than three times higher than when random errors were inserted together with gross error in the simulations. This emphasizes that even gross errors with magnitude over 3σ may be masked by random errors (and then not cause outliers), and that the magnitude of total error is more important than the magnitude of the gross error when it comes to outlier identification.

Table 8: Experiment 3 – alleged outlier with gross error but no random error

MSR	Type II Error	Type III Error	Over-identification+	Over-identification-
12.28%	79.29%	8.23%	0.10%	0.10%

4. Conclusions

The identification of outliers is essential for an appropriate adjustment of observations by LS. However, how can we identify something that sometimes does not seem to have an accurate objective definition? This work has raised some points regarding measurement errors in observations generating outliers in geodetic networks.

From the conceptual discussion presented, we have concluded that a gross error is a type of error, while an outlier is a type of observation. In addition, gross error and outlier are not synonyms, and neither is one a particular case of the other, as a gross error does not always cause an outlier and not all outliers are caused by gross errors. This was substantiated in the experiments, where we could see examples of random errors generating discrepant total errors and gross errors that did not cause outliers.

The classification of an observation as outlier does not depend on the type of error it contains but on the magnitude and stipulated accepted tolerance of the deviation of its total error. Based on the area under the standard normal curve, we have shown that the larger the number of observations in the sample, the greater the probability of the magnitude of the random error of at least one of them being over 3σ, as well as its average occurrence. The 3σ is a usual limit in the sciences, from which it is considered that the total error generates an outlier (*3σ-rule*). Hence, if this rule is applied, observations containing such errors should be considered outliers, despite not containing any gross or systematic errors. In geodesy, it is also common in the intentional insertion of outliers to evaluate identification methods the use of the herein called *3σ-rule for gross error*, which is similar to the first one but which applies the limit of 3σ to a possible gross error of the observation, not to its total error.

We have shown that, even with a gross error above 3σ, the observation tends not to set up an outlier, thus not being identified by IDS, if the gross error has a magnitude close to 3σ and an opposite sign to the random error of the respective observation. This suggests that the insertion of errors by the *3σ-rule for gross error* to evaluate outlier identification methods presents alarming inconsistencies, since this result shows that sometimes analysts can believe that they are generating an outlier in the network, when in fact they are not, thus contaminating the accuracy of the evaluation.

As seen in the theoretical discussion, the criterion of classification of an observation as outlier is conceptually in the discrepancy with respect to the sample, not in the types of error that the observation contains. Thus, a suggested alternative for simulated data for future works would be to understand that the limit of 3σ should not only be valid for the gross error but for the total error of the observations (*3σ-rule*). This approach could also be applied in the context of evaluating outlier identification methods. Thus, observations with discrepant random errors in module (above 3σ) would be considered outliers and those with a total error within the limits of 3σ, even if with gross error, would not be. That is, possibly, these two problems raised in this work would be bypassed.

Hence, future work, besides exploring aspects of outliers related to errors in mathematical modeling and configuration weakness in geodetic networks, not seen in this research, should be developed in order to arrive at an objective and consistent definition for outlier to minimize the inconsistencies presented. Moreover, the actual influence of observations with all kinds and

magnitudes of errors in the parameters estimated should de compared and analyzed, in order to provide more data to support a classification of an observation as outlier or not. Mainly the issue of systematic errors generating outliers, still little explored in geodetic literature, seems to deserve a more comprehensive analysis.

Finally, it is imperative to point out that although 3σ is a usual limit, others have already been or can be tested. For example, the limit of 3.29σ can be adopted, also seen in the context of geodetic observations. With it, despite increasing the acceptance range of the magnitude of the errors, the probability of the random error of a single observation extrapolating the limits of the rule decreases from 0.27% to 0.10%. However, there is no mathematically rigorous justification in the literature that places some limit as ideal for geodetic networks, which can also be approached in future works.

ACKNOWLEDGMENT

AUTHOR´S CONTRIBUTION

Definition of research problem (L.C.O.), Literature review (S.S.S.), Definition of methodological procedures (L.C.O. and S.S.S.), Experiments (S.S.S.), Analysis and interpretation of results (L.C.O. and S.S.S.), Manuscript writing (S.S.S.), Manuscript revision (L.C.O.).

REFERENCES

Amiri-Simkooei, A. 2003. Formulation of L1 Norm Minimization in Gauss-Markov Models. *Journal of Surveying Engineering*, 129 (1), pp.37–43.

Baarda, W. 1968. *A testing procedure for use in geodetic networks.* Publications on Geodesy 9, 2(5). Delft: Netherlands Geodetic Commission.

Baselga, S. 2011. Nonexistence of Rigorous Tests for Multiple Outlier Detection in Least-Squares Adjustment. *Journal of Surveying Engineering*, 137 (3), pp.109–112.

Bustos, O. 1981. *Estimação robusta no modelo de posição*. Rio de Janeiro: Instituto de Matemática Pura e Aplicada.

Erdogan, B. 2014. An outlier detection method in geodetic networks based on the original observations. *Boletim de Ciências Geodésicas*, 20 (3), pp.578–589.

Gemael, C. Machado, A. M. L. and Wandresen, R. 2015. *Introdução ao ajustamento de observações: aplicações geodésicas*. 2nd ed. Curitiba: Ed. UFPR.

Ghilani, C. D. 2010. *Adjustment Computations: Spatial Data Analysis*. 5th ed. Hoboken: John Wiley & Sons, Inc.

GNU, 2018. *GNU Octave*. [online] Available at: <https://www.gnu.org/software/octave/> [Accessed 16 June 2018].

Guo, J. 2015. A note on the conventional outlier detection test procedures. *Boletim de Ciências Geodésicas*, 21 (2), pp.433–441.

Hawkins, D. 1980. *Identification of Outliers*. New York: Champman and Hall.

Hekimoglu, S. Erenoglu, R. 2007. Effect of heteroscedasticity and heterogeneousness on outlier detection for geodetic networks. *Journal of Geodesy*, 81, pp.137–148.

Hekimoglu, S. et al. 2011. Detecting configuration weaknesses in geodetic networks. *Survey Review*, 43 (323), pp.713–730.

Hekimoglu, S. Erdogan, B. Tunalioglu, N. 2012. Elimination of some unknown parameters and its effect on outlier detection. *Boletim de Ciências Geodésicas*, 18 (3), pp.548–557.

Hekimoglu, S. Erdogan, B. 2013. Application of median-equation approach for outlier detection in geodetic networks. *Boletim de Ciências Geodésicas*, 19 (4), pp.347–362.

Klein, I. 2011. *Controle de qualidade de observações geodésicas*. MD. Universidade Federal do Rio Grande do Sul.

Klein, I. Matsuoka, M. T. Souza, S. F. 2011. Teoria de confiabilidade generalizada para múltiplos outliers: apresentação, discussão e comparação com a teoria convencional. *Boletim de Ciências Geodésicas*, 17 (4), pp. 519–548.

Klein, I. et al. 2012. Planejamento de redes geodésicas resistentes a múltiplos outliers. *Boletim de Ciências Geodésicas*, 18 (1), pp. 480–507.

Klein, I. Matsuoka, M. T. Monico, J. F. G. 2013. Proposta para a estimativa da acurácia de redes geodésicas horizontais integrando análise de robustez e de covariância. *Boletim de Ciências Geodésicas*, 19 (4), pp. 525–547.

Klein, I. et al. 2015. On evaluation of different methods for quality control of correlated observations. *Survey review*, 47 (340), pp. 28–35.

Klein, I. Matsuoka, M. T. Guzatto, M. P. 2015. Como estimar o poder do teste mínimo e valores limites para o intervalo de confiança do data snooping. *Boletim de Ciências Geodésicas*, 21 (1), pp. 26–42.

Knight, N. L. Wang, J. Rizos, C. 2010. Generalised measures of reliability for multiple outliers. *Journal of Geodesy,* 84, pp. 625–635.

Lehmann, R. 2013a. On the formulation of the alternative hypothesis for geodetic outlier detection. *Journal of Geodesy*, 87, pp. 373–386.

Lehmann, R. 2013b. The 3σ-rule for outlier detection from the viewpoint of geodetic adjustment. *Journal of Surveying Engineering*, 139(4), pp. 157–165.

Lehmann, R. Losler, M. 2016. Multiple Outlier Detection: Hypothesis Tests versus Model Selection by Information Criteria. *Journal of Surveying Engineering*, doi:10.1061/(ASCE)SU.1943-5428.0000189.

Leys, C. et al. 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), pp. 764-766.

Rofatto, V. F. Matsuoka, M. T. Klein, I. 2017. An Attempt to Analyse Baarda's Iterative Data Snooping Procedure based on Monte Carlo Simulation. *South African Journal of Geomatics*, 6 (3), pp. 416-435.

Rofatto, V. F. Matsuoka, M. T. Klein, I. 2018. Design of geodetic networks based on outlier identification criteria: an example applied to the leveling network. *Bulletin of Geodetic Sciences*, 24 (2), pp. 152-170.

Rofatto, V. F. et al. 2018. A half-century of Baarda's concept of reliability: a review, new perspectives, and applications. *Survey Review*, doi: 10.1080/00396265.2018.1548118.

Teunissen, P. J. G. 2006. *Testing theory: an introduction*. 2nd ed. Delft: Delft University Press.

Teunissen, P. J. G. 2018. Distributional theory for the DIA method. *Journal of Geodesy*, 92, pp. 59-80.

Yetkin, M. Berber, M. 2013. Application of the Sign-Constrained Robust Least-Squares Method to Surveying Networks. *Journal of Surveying Engineering*, 139 (1), pp. 59-65.

Zhao, J. Gui, Q. 2017. Outlier detection in partial errors-in-variables model. *Boletim de Ciências Geodésicas*, 23 (1), pp.1–20.