South African Journal of Geomatics, Vol. 6. No. 3, October 2017

An Attempt to Analyse Baarda's Iterative Data Snooping Procedure based on Monte Carlo Simulation

Vinicius Francisco Rofatto[1,2], Marcelo Tomio Matsuoka[1,2], Ivandro Klein[3]

[1]Federal University of Uberlandia (UFU), Institute of Geography, Surveying and Cartographic Engineering, Monte Carmelo, Minas Gerais, Brazil, vinicius.rofatto; tomio@ufu.br
[2]Federal University of Rio Grande do Sul (UFRGS), Graduate Program in Remote Sensing, Porto Alegre, Rio Grande do Sul, rofattaum; tomiomatsuoka@gmail.com
[3]Federal Institute of Science and Technology Education of Santa Catarina (IFSC), Landing Surveying Program, Florianópolis, Santa Catarina, ivandroklein@gmail.com

DOI: http://dx.doi.org/10.4314/sajg.v6i3.11

Abstract

William Sealy Gosset, otherwise known as "Student", Fisher's disciple, was one of the pioneers in the development of modern statistical method and its application to the design and analysis of experiments. Although there were no computers in his time, he discovered the form of the "t distribution" by a combination of mathematical and empirical work with random numbers. This is now known as an early application of the Monte Carlo simulation. Today with the fast computers and large data storage systems, the probabilities distribution can be estimated using computerized simulation. Here, we use Monte Carlo simulation to investigate the efficiency of the Baarda's iterative data snooping procedure as test statistic for outlier identification in the Gauss-Markov model. We highlight that the iterative data snooping procedure can identify more observations than real number of outliers simulated. It has a deserved attention in this work. The available probability of over-identification allows enhancing the probability of type III error as well as probably the outlier identifiability. With this approach, considering the analysed network, in general, the significance level of 0.001 was the best scenario to not make mistake of excluding wrong observation. Thus, the data snooping procedure was more realistic when the over-identifications case is considered in the simulation. In the end, we concluded that for GNSS network that the iterative data snooping procedure based on Monte Carlo can locate an outlier in the order of magnitude 4.5σ with high success rate.

1. Introduction

The reliability of outlier identification is one of the major challenges in the quality control of geodetic measurements. In the sense of Least Squares Estimation (LSE), the outliers are nuisance observations that spoil both estimated parameters and their standard deviations, causing incorrect results. Thus, we often try to minimize the magnitude of undetectable outliers in the observations as well as to reduce the effect of the undetected ones on the estimated parameters. Two categories for the treatment of observations contaminated by outliers have been developed: robust adjustment procedures (for an overview see e.g. Wilcox, 2012; Klein et al. 2015a) and outlier detection based

on statistical tests (e.g. Baarda, 1968; Pope, 1976; Lehmann and Lösler, 2016; Klein et al. 2017). The first one is outside the scope of this paper. Besides the undoubted advantages of robust adjustment, the outlier tests are also used. The following advantages of outlier analysis are mentioned by Lehmann (2013):

- Detected outliers provide the opportunity to investigate causes of gross measurement errors;
- Detected outliers can be re-measured;

One of the best methods for outlier identification in geodetic data analysis is Baarda's testing procedure. This method is due to Baarda (1968). This method consists of three steps (see e.g. Teunissen 2006): detection (also known as overall model test), identification (also known as data snooping) and adaption (a corrective action, such as elimination of identified observation as an outlier).

At each iteration, only a single observation can be identified in the data snooping procedure. Once an identified observation is excluded, the LSE adjustment is restarted without the rejected observation and again the identification step (data snooping) is applied. Of course, if redundancy permits, this procedure is repeated until none identification. This procedure is called "iterative data snooping" (e.g. Teunissen, 2006). In this paper we are exclusively concerned with iterative data snooping procedure.

Since data snooping is based on a statistical hypothesis testing with an alternative hypothesis for each observation, it may lead to a false decision as follows:

- Type I error or ***false alert*** (probability level α) – Probability of identifying an outlier when there is none;
- Type II error or ***missed detection*** (probability level β) – Probability of non-identifying an outlier when there is at least one; and
- Type III error or ***wrong exclusion*** (probability level κ) – Probability of misidentification a non-outlying observation as an outlier, instead of the outlying one. This type of error decision was introduced by (Hawkins 1980; Förstner 1983).

The rate of type I decision error in a binary hypothesis test (i.e., with a single alternative hypothesis) can be selected by the user. The rate of type II decision error cannot. Lehmann and Voß-Böhme (2017) also point out that a test statistic with a low rate of type II is said to be powerful in the binary hypothesis case, when only a single alternative hypothesis is considered. However, in case of multiple alternative hypotheses (i.e., data snooping), without considering the Type III error, there is a high risk of over-estimating the successful identification probability (see e.g. Yang et al. 2013). On other hand, the confidence level is the probability that a non-outlying observation is correctly ignored; the power of the test is the probability that an outlier is correctly identified. Therefore, the confidence level and the power of the test are the probabilities of the test result leading to correct decisions, as opposed to the occurrence of type I, II and III errors (see, for example, Förstner, 1983; Teunissen, 2006; Klein et al. 2015b).

Here, we extended the decision errors above; we highlight that "iterative data snooping" procedure can identify more observations than real number of outliers (we call here "over-identification"). The later has a deserved attention in this work. The aim of this paper is to compute the statistical quantities (Power of test, type II error, Type III error and "over-identification") for iterative data snooping by means of the well-established Monte Carlo Simulation (MCS). This simulation technique in geodesy has been widely applied since the pioneering idea of Hekimoglu and Koch (1999).

Many of the relevant probabilities in this contribution are multivariate integrals over complex regions (Teunissen, 2017). They therefore need to be computed by means of numerical simulation such as MCS. MCS methods are used whenever the functional relationships are analytically not simple tractable, as is the case for data snooping testing procedure (Lehmann, 2013). MCS method replaces random variables by computing excessive random experiments. In other words, the statistical quantities can be determined by frequency distributions of computer random experiments performed using random numbers. The MCS has already been applied in outlier detection (e.g. Lehmann and Scheffler, 2011; Lehmann, 2012; Klein et al. 2012; Klein et al. 2015a, 2015b; Erdogan, 2014; Niemeier and Tengen, 2017). Following this line of thought, here our goal was to apply the MCS to analyse the efficiency of the iterative data snooping procedure for the correct identification (or not) of a single simulated outlier at time.

The rest of the paper is organised as follows: Section 'Theoretical overview' brings a theoretical background about data snooping statistical tests for outlier identification in the LSE. Section 'Monte Carlo approach for data snooping procedure' presents the method for determining the statistical quantities numerically. Section 'Experiments and results' contains the experiments setup, results and discussions. Finally, last section offers conclusions and recommendations for future studies.

2. Theoretical overview

The mathematical model generally adopted in geodetic data analysis is the linear(ised) Gauss-Markov model, given by (Koch, 1999):

$$e = y - A\,x \qquad [1]$$

Where e is the $n\,x1$ random error vector, A is the design(or Jacobian) matrix, x is the $u\,x1$ unknown parameters vector and y is the $n\,x1$ observations vector. The most employed solution for a redundant system of equations ($n > \mathrm{rank}\,(A)$) is the weighted least squares estimator (WLSE) for the vector of unknowns ($\hat{x}$):

$$\hat{x} = (A^T W A)^{-1}(A^T W y) \qquad [2]$$

W is the *n x n* weight matrix of the observations, taken as $W = \sigma_0^2(\Sigma_y)^{-1}$, where σ_0^2 is the variance factor and Σ_y is the covariance matrix of the observations. Teunissen (2003) demonstrates the geometric interpretation of the WLSE. More details about WLSE estimation can be seen in Ghilani (2010).

If there are only random errors in the observations, the WLSE is the best linear unbiased estimator (BLUE) for the unknown parameters; if the observational errors follow the multivariate normal distribution with mean $\mu = [0]$ and covariance matrix Σ_y, the WLSE coincides with the maximum likelihood estimator (Teunissen, 2003). However, the WLSE is no longer optimal in the presence of systematic and/or gross errors (blunders) in the observations. In other words, despite optimal properties for WLSE, they lack robustness or insensitivity to outliers in observations (Huber, 1964; Rousseeuw and Leroy, 1987; Lehmann, 2013). Therefore, statistical testing procedures for detection and identification of outliers have been developed.

Quality control to identify outliers in geodetic measurements has been widely investigated since the pioneering work of Baarda (1968). In the sense of LSE, statistical testing procedures for detection and identification of outliers are based on maximum likelihood ratio. Consider a null hypothesis H_0 for the parameters of the population probability distribution of an observation vector *y*. Consider further an alternative hypothesis H_A for these parameters, constructed in a way that H_0 is a subset of H_A. Thus, in the general case, the maximum likelihood ratio between H_0 and H_A is given by (Larson, 1974):

$$\lambda(y) = \frac{\max\, p(y \mid H_0)}{\max\, p(y \mid H_A)} \qquad [3]$$

Where $\max\, p(y \mid H_0)$ is the maximum of the probability density function (pdf) of *y* under H_0 and $\max\, p(y \mid H_A)$ is the maximum of the pdf of *y* under H_A. As the null hypothesis is defined so that its sample space is contained in the sample space of the alternative hypothesis, the ratio in Equation 3 lies in the interval of $0 \le \lambda(y) \le 1$ (Teunissen, 2006). The test criterion for the maximum likelihood ratio is given by (Larson, 1974):

$$\text{Do not reject } H_0 \text{ if } \lambda(y) \ge c \qquad [4]$$

Where $c > 0$ is the critical value for the test according to the significance level α stipulated (for more details, see Larson, 1974; Teunissen, 2006).

Assuming normally distributed observation errors, a general case of hypothesis testing in linear models is formulated as Teunissen (2006):

$$H_0 : E\{y\} = Ax \qquad \text{vs} \qquad H_A : E\{y\} = Ax + C_y\nabla; \nabla \neq 0 \tag{5}$$

Where H_0 is the null hypothesis (namely, absence of outliers in the observations) and H_A is an alternative hypothesis (presence of "*q*" outlying observations in at certain known locations). The quantity C_y defines the non-random error model (in this context called outlier model) with dimension n x *q* and ∇ is the corresponding vector of *q* outliers. The dimension of ∇should be comprised between$1 \le q \le n-u$. For example, if n = 5 and *q* = 2, then a possible outlier model is $C_{y_{5x2}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$and $\nabla_{2x1} = \begin{bmatrix} \nabla_1 \\ \nabla_3 \end{bmatrix}$. (one outlier in each observation y_1 and y_3). For more details about error models, see e.g. Lehmann and Lösler (2016).

Considering the maximum of the pdf of y under H_0 and H_A, the maximum likelihood ratio λ between the two hypotheses becomes (Teunissen, 2006):

$$\text{Reject } H_0 \text{ if: } T_q = \hat{e}_0^T \sum_y^{-1} C_y (C_y^T \sum_y^{-1} \sum_{\hat{e}_0} \sum_y^{-1} C_y)^{-1} C_y^T \sum_y^{-1} \hat{e}_0 > K_\alpha \tag{6}$$

Where $\hat{e}_0$ and $\sum_{\hat{e}_0}$ is the estimated random error vector and a posteriori covariance matrix of the estimated random error computed by LSE into H_0, respectively. K_α is the critical value for the test according to the significance level α. For more details see Koch (1999) and Teunissen (2003). Under the null hypothesis, the test statistic T_q follows the central chi-squared distribution with *q* degrees of freedom; under the alternative hypothesis, the test statistic T_q follows the non-central chi-squared distribution with *q* degrees of freedom and non-centrality parameter$\delta = \nabla^T C_y^T \sum_y^{-1} \sum_{\hat{e}_0} \sum_y^{-1} C_y \nabla$.

Data snooping procedure is a particular case of maximum likelihood ratio test when only one outlier (i.e. *q* = 1) is present in the data set at a time (see e.g. Baarda, 1968; Pope, 1976; Berber and Hekimoglu, 2003; Lehmann, 2012). Thus, it is formulated by the following test hypotheses (called individual model test or w-test) (Baarda, 1968; Teunissen, 2006):

$$H_0 : E\{y\} = Ax \qquad \text{vs} \qquad H_A : E\{y\} = Ax + c_y\nabla; \nabla \neq 0 \tag{7}$$

Where c_y is outlier model for q=1, i.e. the n x 1 unit vector with 1 in its i^{th} entry and zeros in the remaining (e.g. $c_{y_{nx1}} = \begin{bmatrix} 0 & 0 & \cdots & 1 & 0 & \cdots & 0 & 0 \end{bmatrix}^T$), and ∇ is a scalar value with the gross error (*outlier*) at i^{th} observation being tested. Therefore, in the null hypothesis (H_0), it is assumed that there are no outliers in the observations, while in the alternative hypothesis (H_A), is it assumed that the i^{th} observation being tested (for $i = 1, \ldots, n$) is contaminated by gross error of magnitude ∇.

If we consider one outlying observation in at certain known locations (q = 1), then the maximum likelihood ratio test for data snooping ($T_q = 1$) is given by (Teunissen, 2006):

$$T_{q=1} = \hat{e}_0^T \sum_y^{-1} c_y (c_y^T \sum_y^{-1} \sum_{\hat{e}_0} \sum_y^{-1} c_y)^{-1} c_y^T \sum_y^{-1} \hat{e}_0 \quad [8]$$

Under H_0, observation errors are zero-mean (multivariate) normally distributed. The null hypothesis is rejected if the following test statistic ($T_{q=1}$) of the i^{th} observation being tested exceeds a given critical value K_α , i.e.:

$$\begin{cases} \text{Reject } H_0 \text{ if: } T_{q=1} > K_\alpha \\ H_0 : T_{q=1} \sim \chi^2_{(1,0)};\ H_A : T_{q=1} \sim \chi^2_{(1,\delta)},\ \text{with } \delta = c_y^T \sum_y^{-1} \sum_{\hat{e}_0} \sum_y^{-1} c_y \nabla^2 \end{cases} \quad [9]$$

It should be noted that the test statistic T_q in Equation [8] is a particular case of generalized test statistic, when *q=1*. Important to mention also that the critical value follows from a chi-squared distribution with one degree freedom at a significance level of α in a one-tailed test. Baarda (1968) and Teunissen (2006) demonstrate that if q = 1, then the test statistics (Equation 9) can also be formulated based on a standard normal distribution in a two-tailed test (so-called *w-test*). Both the chi-squared and normal distribution tests are equivalent. Usually in geodesy, the value of α is set between 0.1% and 1% (Kavouras, 1982; Aydin and Demirel, 2004; Lehmann, 2013). Furthermore, data snooping contains multiple alternative hypotheses, as each observation is individually tested. Therefore, the only observation considered contaminated by outlier is the one whose test statistic satisfies the inequalities $T_{q=1} > K_\alpha$. In the case that two or more observations exceed the critical value K_α only the observation with the largest $T_{q=1}$ is flagged as an outlier. After having identified the observation most suspected of being an outlier (at given α), it is excluded usually from the model, and WLSE and data snooping are applied iteratively until there are no further outliers identified in the observations (iterative data snooping procedure) (Teunissen, 2006; Berber and Hekimoglu, 2003).

However, three types of incorrect decisions may occur into data snooping and its occurrence rates are associated with probability levels: the significance level α is the probability (when H_o is true) of a non-outlying observation be misidentified as an outlier (type I error or false positive); β is the probability that an outlying observation not be identified as outlier (type II error or false negative); finally, a non-outlying observation is misidentified as an outlier, instead of the outlying one (type III error given by κ). On other hand, the confidence level (CL) is the probability that a non-outlying observation is correctly ignored, therefore, CL = 1 – α; the power of the test (γ) is the probability that an outlier is correctly identified, i.e. γ = 1 – (β + κ). Therefore, the CL and the γ are the probabilities of the test result leading to correct decisions, as opposed to the occurrence of type I, I and III errors (see, for example, Förstner, 1983; Teunissen, 2006; Klein et al. 2015b). For example, the Figure 1 shows these relationships in the data snooping procedure (considering a single H_A) for CL=0.999, γ_0 = 0.80, so α_0 = 0.001 and β_0 = 0.20, leading to a pre-set non-centrality

parameter $\delta_0 = 17.075$ and a pre-set critical value of K_{α_0} = 10.83. The type III error does not appear in Figure 1; it would be linked with another alternative hypothesis concerning another observation.

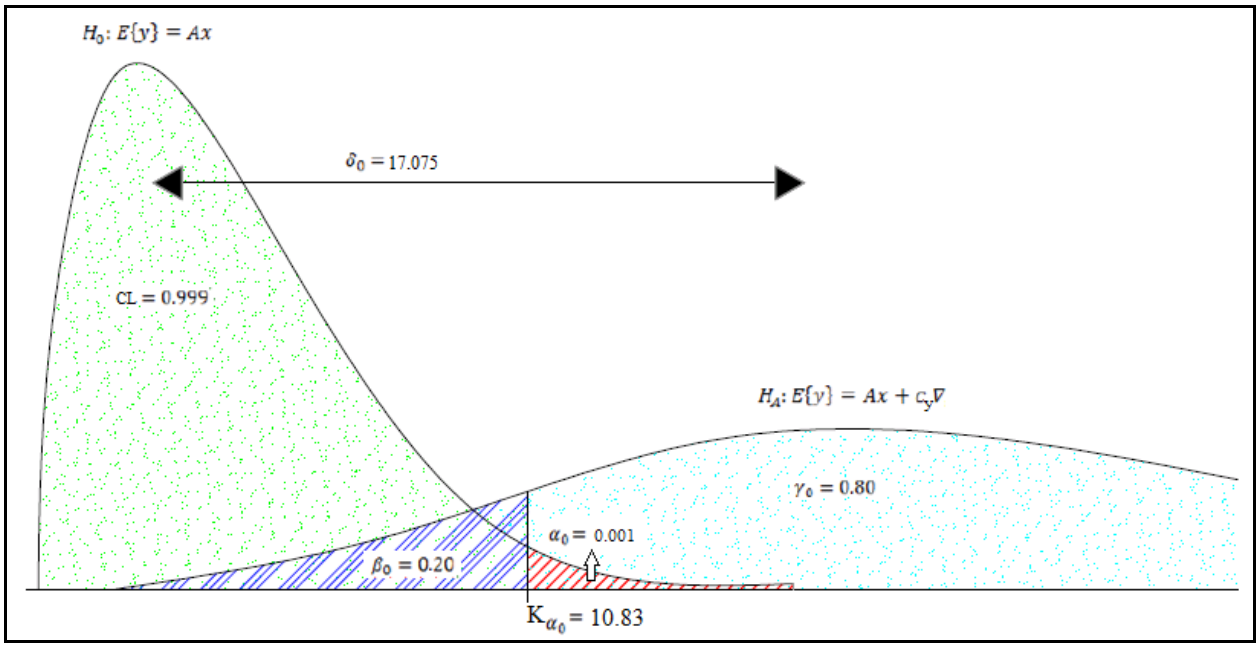

Figure 1. Probability levels related to testing hypotheses of data snooping.

Analyzing Figure 1, it can be concluded that the non-centrality parameter δ_0 as function of the significance level α_0, power of the test γ_0, and degree of freedom (or number of outliers considered). The γ_0 decreases with the significance level α_0 for a given value of δ_0. On the other hand, the γ_0 increases with the non-centrality parameter δ_0. Baarda (1968) provides the monograms for those interested in obtaining δ_0 values as a function of α_0 and γ_0 (for a given degree of freedom). Alternatively, Aydin and Demirel (2004) presented a procedure to obtain the same through approximations of the non-central chi-squared distribution. The necessity of obtaining the non-centrality parameter is widespread in Geodesy (Baarda, 1968; Kavouras, 1982; Teunissen, 2006; Knight et al. 2010).

In addition to these probabilities, the iterative data snooping procedure can identify more observations than real number of outliers (here we call "over-identification"). The "over-identification" may contain one or more observations correctly identified or, in the worst case scenario, all erroneously identified.

3. Monte Carlo approach to analyse the iterative data snooping procedure

The MCS is applied to compute the probabilities levels. To do so, a sequence of *m* random errors vector $\boldsymbol{e}_K$, $\boldsymbol{k = 1,\ldots,m}$ of a desired statistical distribution is generated. The "$\boldsymbol{m}$" is known as the number of Monte Carlo experiments. Usually, assume that the random errors of the good measurements are normally distributed with expectation zero. Thus, we generate the random errors using the well-known Box-Muller method (Box and Muller, 1958) based on multivariate normal

distribution, since the assumed stochastic model for random errors is based on matrix covariance of the observations, i.e. $\boldsymbol{e \sim N(0, \sigma_0^2 \Sigma_y)}$.

On the other hand, an outlier (q=1) is selected based on magnitude intervals of the outliers for each ***m*** Monte Carlo experiments. We use the standard uniform distribution to select the outlier magnitude. The uniform distribution is a rectangular distribution with constant probability and implies the fact that each range of values that has the same length on the distributions support has equal probability of occurrence (see e.g. Lehmann and Shuffler, 2011). For example, for 10,000 Monte Carlo experiments, if the one choices a magnitude interval of the outliers of |3σ to 9σ|, the probability of a 3σ error occurring is virtually the same as -3σ, and so on. At each iteration of the simulation, a specific observation is chosen to receive a gross error based on the discrete uniform distribution (i.e., all observations have the same probability of being selected).

Random and gross errors are assumed to be independent (by definition) and both are combined to the total error as follow (see e.g. Kavouras, 1982):

$$\varepsilon = e + c_y \nabla, \ \nabla \neq 0 \qquad [10]$$

Where ε is the n x 1 total error vector, e is n x 1 random errors vector and c_y is outlier model for q=1, and ∇ is a scalar value with the outlier at i^{th} observation being tested. Here, we assume that ∇>e. In order to avoid the compensation and potentiation problems, i.e. $\nabla - e$ and $\nabla + e$, respectively, the observation selected to contain a gross error has its random error removed in the Equation 10. Before computing statistical test T_q=1 it is necessary to relate the random error vector e and total error vector ε, since this statistical test depends on the estimated random error vector $\hat{e}_0$. In the sense of LSE, this relationship is given by:

$$e_\nabla = R\varepsilon \qquad [11]$$

In the Equation 11 the reader should note that the multiplication of the redundancy matrix (R) and the total error ε provides a total error vector e_∇. The total error vector is not only composed by random errors, but also it has one of its elements contaminated by an outlier.

Now it becomes possible to compute the test statistic T_q=1 considering e_∇ instead of $\hat{e}_0$ in the Equation 8 for all observations and perform the iterative data snooping procedure and qualify the efficiency in identifying simulated outlier The whole procedure described so far is performed again until the ***m*** Monte Carlo experiments are completed.

The redundancy matrix **R** in Equation 11 is based on the network geometry and covariance matrix. The redundancy matrix is given by:

$$R = I - A(A^T W A)^{-1} A^T W \qquad [12]$$

Where R is the $n \times n$ redundancy matrix and I is the $n \times n$ identity matrix. The diagonal elements of R are the **local redundancy numbers (*r*)**. The local redundancy numbers indicate the fraction of a possible outlier on the observation, which is reflected in the respective residue of this observation. Reliability measures such as local redundancy numbers are intrinsically related to the network

geometry/configuration and observation weights. Such measure is widely used for quality analysis of geodetic networks, both in the design stage as well as for quality control (Kuang, 1991; Klein et al. 2012). It is desirable to have approximately a constant value for all redundancy numbers so that the ability of detecting outliers will be the same in every part of the network. In other words, one seeks to minimize the magnitude of undetectable outliers in the observations by increasing the redundancy numbers in order to have an optimal network configuration (see e.g. Amiri-Seemkooei, 2001a, 2001b). Furthermore, the redundancy numbers are correlated with the robustness parameters proposed by Vaníček et al. (1990) and Vaníček et al. (2001).

Note that the method described here for evaluating the iterative data snooping depends only on the matrix A and W, and the magnitude of the desired outlier.

4. Experiments and results

In this study, the previously described method was applied considering a GNSS (*Global Navigation Satellite System*) network, with one control station (fixed) and five u stations with unknown 3D positions (X,Y,Z), totalling six minimally constrained stations (see Figure 2). For each pair of stations, there are four or five baseline vectors (ΔX, ΔY, ΔZ components). Thus, there are $n = 13 \times 3 = 39$ observations (baseline vector components), $u = 5 \times 3 = 15$ unknowns and $n - u = 39 - 15 = 24$ redundant observations. The stations are taken from the Brazilian Network for Continuous Monitoring **of** GNSS. Baseline vectors, free of outliers, consist in differences between the stations official coordinates in the SIRGAS2000 reference frame. The observation covariance matrix is obtained through data processing of 6-hour sessions for each baseline vector, resulting in a 3×3 full matrix, combined in a $13 \times (3 \times 3) = 39 \times 39$ block diagonal matrix. More details about network can be found in Klein (2014).

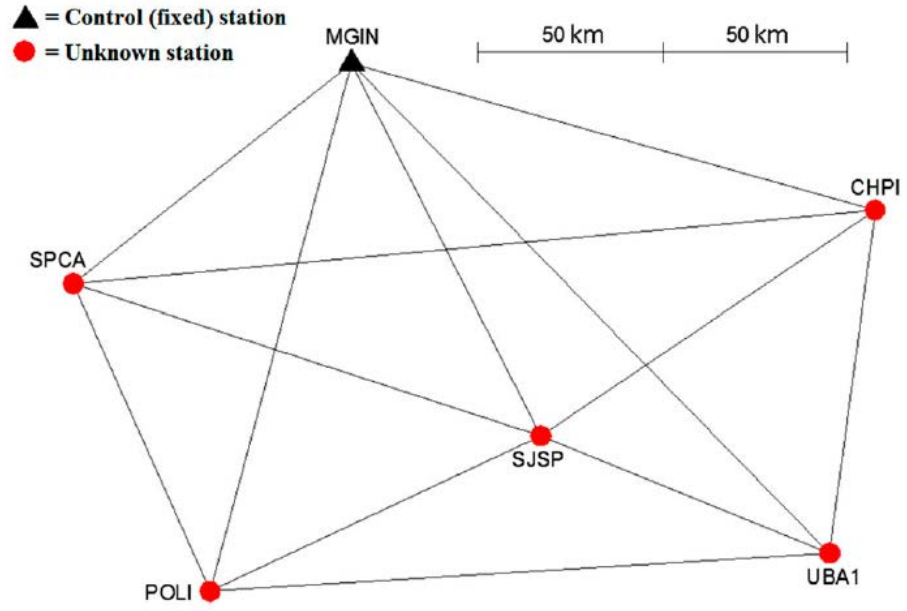

Figure 2. GNSS network analysed by means Monte Carlo approach for data snooping procedure

It is important to mention that the design matrix defined initially must have a minimum configuration to avoid rank deficiency as well as being able to identify at least one outlier as mentioned by Xu (2005) that 'in order to identify outliers, one also has to further assume that for each model parameter, there must, at least, exist two good data that contain the information on such a parameter'. For example, consider the one unknown height into a leveling network (one-dimensional - 1D). Two observations would lead to different solutions and allow the detection of an inconsistency between them. Three observations would lead to different solutions and the identification of one outlying observation, and so on. Thus, in a general case, the value for 'q' equal to the minimal number of redundant observations across each and every point, minus one. For more details on the choosing the number of outliers to be considered, see Klein et al. (2017).

We highlight that the redundancy numbers (diagonal elements of Equation 12) were between 0.43 (minimum) and 0.78 (maximum) for the network, being classified as a good controllability; further considerations about GNSS networks are outside the scope of this study.

Here, the significance level for iterative data snooping is varied, taken as $\alpha = 0.001$ (0.1%), $\alpha = 0.005$ (0.5%), $\alpha = 0.01$ (1%), $\alpha = 0.025$ (2.5%) and $\alpha = 0.05$ (5%). Each Monte Carlo simulation has a unique combination of significance level and magnitude of outliers. We ran 10,000 experiments for each simulation and compute the rate of type II error, type III error, the power of the test and the over-identification in iterative data snooping, totaling 12 x 4 x 10,000 = 480,000 Monte Carlo simulations. It is important to emphasize that the proposed method does not depend on the unknown parameters vector or the vector of observations as can seen in the section 3.

Random errors and outliers are synthetically generated and added to the observations. Each unknown station is involved in at least four baseline vectors; thus, the local-scale redundancy equals three. Positive and negative outliers are clipped between 3σ and 3.5σ, 3.5σ and 4σ; 4σ and 4.5σ; 4.5σ and 5σ; 5 and 5.5σ; 5.5σ and 6σ; 6σ and 6.5σ; 6.5σ and 7σ; 7σ and 7.5σ; 7.5σ and 8σ; 8σ and 8.5σ; 8.5σ and 9σ in each experiment.

Figure 3 shows the success rate (number of experiments that only outlying observation was identified), i.e. the power of the iterative data snooping testing procedure for one simulated outlier in the GNSS network (γ). The misidentifications rates are showed in the Figures 4-5. The misidentifications are divided in two types of classes are counted in the simulations: number of experiments where the procedure yielded none observation identification (type II error - β); number of experiments in which the procedure identified a single observation but wrong localization (type III error - κ). In addition to these classes, we consider "over-identification" case (more identified observation than one) and divided it into two categories: number of experiments where the procedure identified the outlying observation and others ("over-identification +") – see Figure 6; number of experiments where the procedure identified only non-outlier observations ("over-identification –") – see Figure 7.

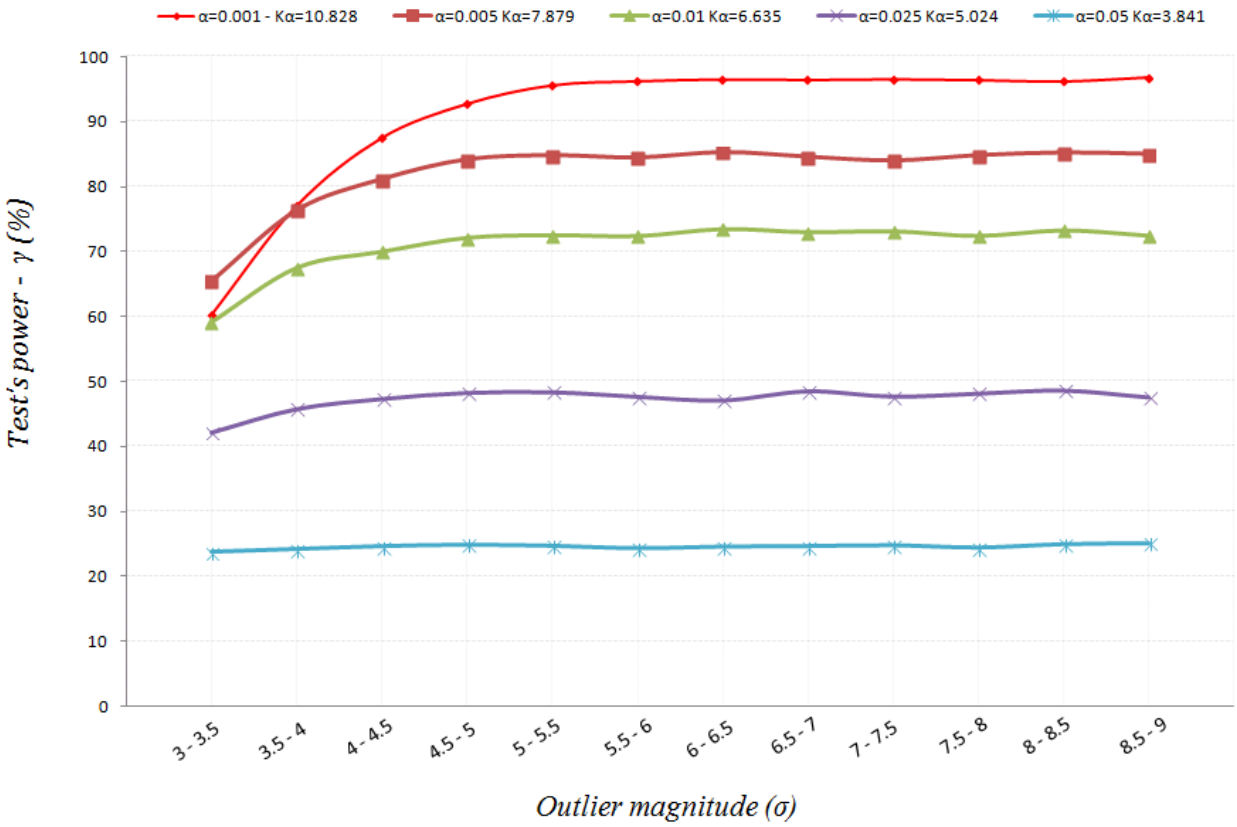

Figure 3. Power of the data snooping testing procedure for GNSS network vs. magnitude intervals of the outliers for each probability level α

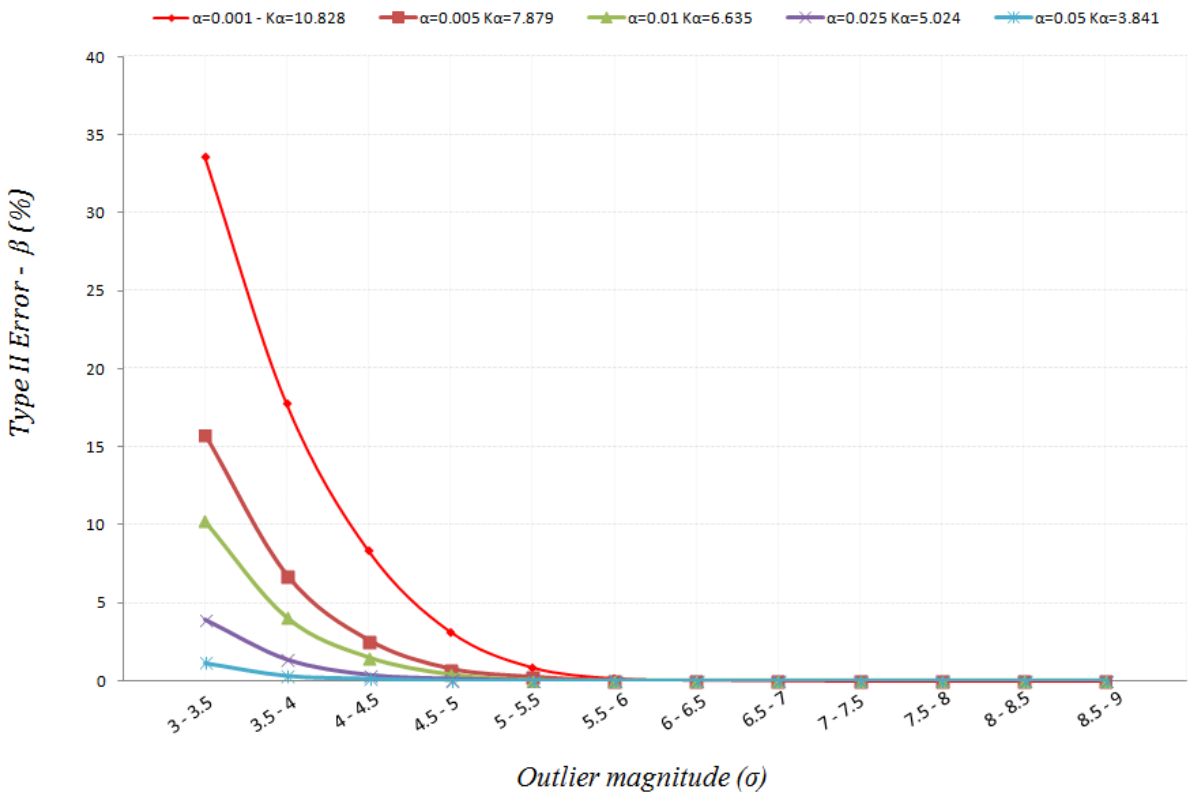

Figure 4. Type II error for GNSS network vs. magnitude intervals of the outliers for each probability level α

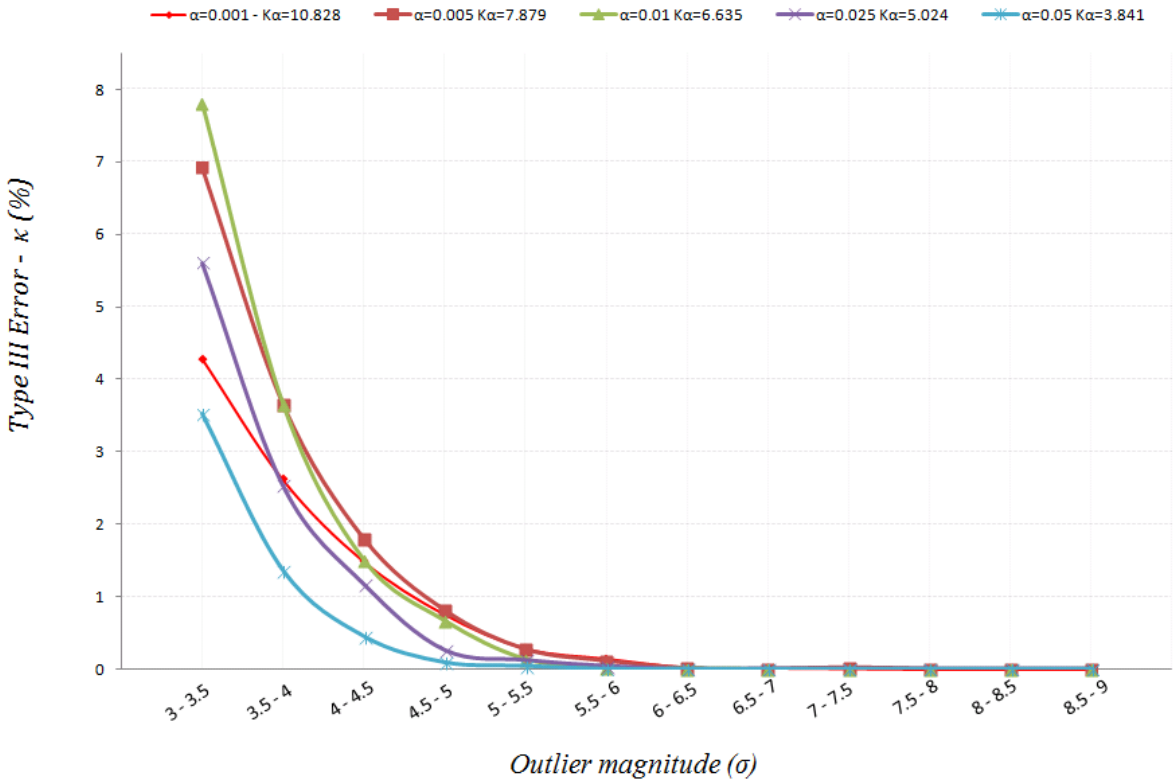

Figure 5. Type III error for GNSS network vs. magnitude intervals of the outliers for each probability level α

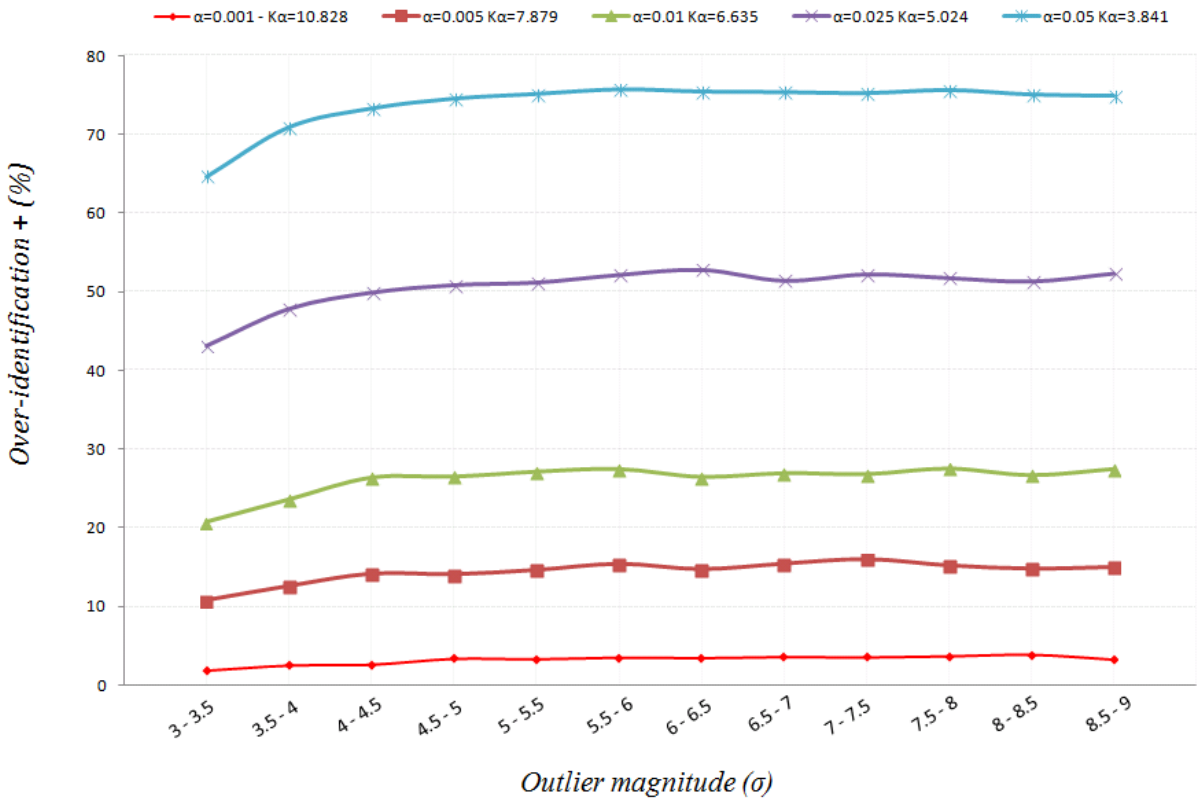

Figure 6. "Over-identification+" for GNSS network vs. magnitude intervals of the outliers for each probability level α

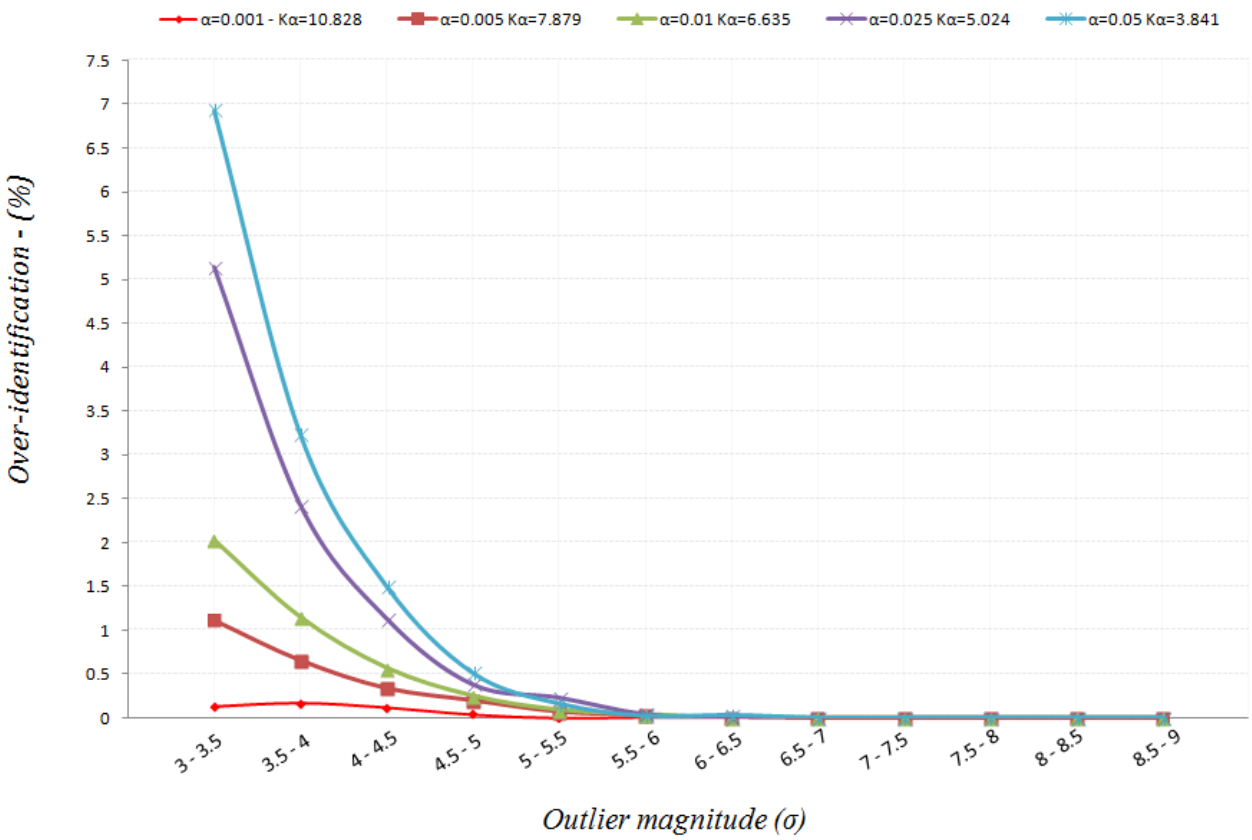

Figure 7. "Over-identification–" for GNSS network vs. magnitude intervals of the outliers for each probability level α

The "over-identification –" was also added to type III error case (k) (see Figure 8). Tables 1-5 show the success (γ), misidentifications (β and κ) and over-identifications rates for the various significance levels and intervals of the outlier size considered in this work.

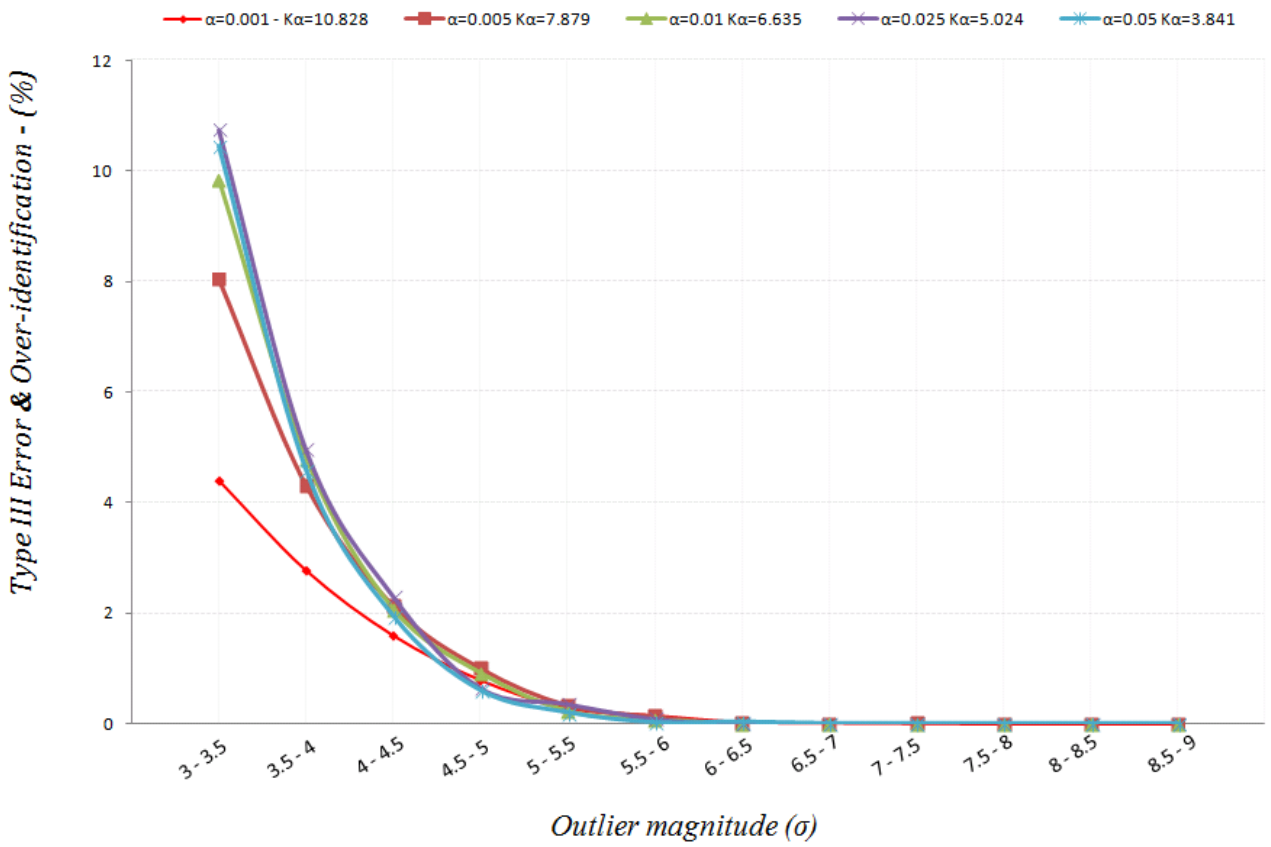

Figure 8. "Over-identification–" added to type III error for GNSS network vs. magnitude intervals of the outliers for each probability level α

Table 1. Success (γ), misidentifications (β and κ) and over-identifications for α = 0.001

Outlier Magnitude	γ %	β %	κ %	Over-identification(+) %	Over-identification (–) %
3 - 3.5	60.06	33.67	4.29	1.85	0.13
3.5 - 4	76.89	17.81	2.62	2.51	0.17
4 - 4.5	87.39	8.38	1.49	2.62	0.12
4.5 - 5	92.64	3.18	0.76	3.38	0.04
5 - 5.5	95.5	0.89	0.29	3.32	0
5.5 - 6	96.12	0.19	0.15	3.53	0.01
6 - 6.5	96.41	0.08	0.03	3.47	0.01
6.5 - 7	96.33	0.04	0.02	3.61	0
7 - 7.5	96.43	0	0.01	3.56	0
7.5 - 8	96.3	0	0	3.7	0
8 - 8.5	96.1	0	0	3.9	0
8.5 - 9	96.73	0	0	3.27	0

Table 2. Success (γ), misidentifications (β and κ) and over-identifications for α = 0.005

Outlier Magnitude	γ %	β %	κ %	Over-identification(+) %	Over-identification (–) %
3 - 3.5	65.39	15.79	6.93	10.77	1.12
3.5 - 4	76.3	6.78	3.66	12.6	0.66
4 - 4.5	81.09	2.6	1.8	14.17	0.34
4.5 - 5	84.13	0.77	0.81	14.09	0.2
5 - 5.5	84.76	0.25	0.27	14.65	0.07
5.5 - 6	84.41	0.02	0.11	15.43	0.03
6 - 6.5	85.24	0	0.01	14.75	0
6.5 - 7	84.54	0	0	15.46	0
7 - 7.5	83.95	0	0.02	16.03	0
7.5 - 8	84.77	0	0	15.23	0
8 - 8.5	85.18	0	0	14.82	0
8.5 - 9	84.97	0	0	15.03	0

Table 3. Success (γ), misidentifications (β and κ) and over-identifications for α = 0.01

Outlier Magnitude	γ %	β %	κ %	Over-identification(+) %	Over-identification (–) %
3 - 3.5	59.2	10.26	7.81	20.7	2.03
3.5 - 4	67.55	4.06	3.66	23.58	1.15
4 - 4.5	70.04	1.5	1.51	26.38	0.57
4.5 - 5	72.18	0.4	0.67	26.5	0.25
5 - 5.5	72.59	0.04	0.14	27.14	0.09
5.5 - 6	72.44	0.01	0.03	27.48	0.04
6 - 6.5	73.52	0	0.01	26.47	0
6.5 - 7	73.04	0.01	0	26.95	0
7 - 7.5	73.18	0	0	26.82	0
7.5 - 8	72.46	0	0	27.54	0
8 - 8.5	73.31	0	0	26.69	0
8.5 - 9	72.51	0	0	27.49	0

Table 4. Success (γ), misidentifications (β and κ) and over-identifications for α = 0.025

Outlier Magnitude	γ %	β %	κ %	Over-identification(+) %	Over-identification (–) %
3 - 3.5	42.21	3.91	5.61	43.13	5.14
3.5 - 4	45.8	1.37	2.53	47.87	2.43
4 - 4.5	47.35	0.39	1.17	49.95	1.14
4.5 - 5	48.29	0.14	0.26	50.92	0.39
5 - 5.5	48.38	0.02	0.12	51.24	0.24
5.5 - 6	47.69	0	0.04	52.22	0.05
6 - 6.5	47.13	0	0	52.87	0
6.5 - 7	48.53	0	0	51.47	0
7 - 7.5	47.72	0	0	52.28	0
7.5 - 8	48.18	0	0	51.82	0
8 - 8.5	48.63	0	0	51.37	0
8.5 - 9	47.61	0	0	52.39	0

Table 5. Success (γ), misidentifications (β and κ) and over-identifications for α = 0.05

Outlier Magnitude	γ %	β %	κ %	Over-identification(+) %	Over-identification (–) %
3 - 3.5	23.7	1.17	3.52	64.68	6.93
3.5 - 4	24.17	0.32	1.36	70.92	3.23
4 - 4.5	24.64	0.13	0.45	73.28	1.5
4.5 - 5	24.88	0	0.09	74.52	0.51
5 - 5.5	24.7	0	0.04	75.1	0.16
5.5 - 6	24.3	0	0	75.68	0.02
6 - 6.5	24.57	0	0	75.4	0.03
6.5 - 7	24.66	0	0	75.34	0
7 - 7.5	24.79	0	0	75.21	0
7.5 - 8	24.41	0	0	75.59	0
8 - 8.5	24.97	0	0	75.03	0
8.5 - 9	25.13	0	0	74.87	0

In general, the greater the magnitude of outliers, the greater is the efficiency of iterative data snooping procedure. It is important to note that from 4.5-5σ the type II error is practically absent. The results show also that the probability of committing different types of error depends more on the critical value than outlier magnitude for one-dimensional identification and the GNSS network analysed.

Iterative data snooping procedure was more efficient for outliers larger than 4.5σ. Considering all results for the GNSS network, the mean success rate was 90.6% for α=0.001(0.1%) and 82% for

α=0.005(0.5%). Thus, we consider α=0.001 and α=0.005 satisfactory significance level for the GNSS network analysed. Therefore, these results show the importance of a correct choice of α, as pointed out by Lehmann (2012); it also highlights the challenges in controlling the error rate in multiple hypotheses tests and iterative tests.

Regarding the two classes of over-identification rates, in general, the influence of committing the "over-identification+" and "over-identification–" is directly related to probability level α, i.e. the greater type I error, the greater is the over-identifications case. From the 5.5-6σ the "over-identification–" is practically null. Furthermore, if we disregard the "over-identification–" case, one could erroneously consider the significance level 0.05 the best scenario for type III error (see Figure 5). However, considering the type III error plus "over-identification–" the lowest probability of making a wrong exclusion is actually for a significance level of 0.001; as can be seen in the Figure 8. Therefore, the "over-identification–" rates should be considered for a more accurate and thorough analysis of the type III error as well as to avoid false interpretation of the results.

To conclude, it is always important to emphasize that the level of significance that is used to determine the theoretical critical value of the test is not the probability of the type I error of the iterative data snooping procedure. It is the only type I error of the local test (w-test) for a single alternative hypothesis. The type I error of iterative data snooping procedure can also be estimated based on Monte Carlo simulations. This topic will be covered in future work.

5. Conclusions

Monte Carlo methods are tools for solving problems using random numbers. Although this might sound somewhat specific and not very promising, Monte Carlo methods are fundamental tools in many areas of modern science. Here, the goal was to analyse the iterative data snooping testing procedure to locate an outlier by means of the Monte Carlo Simulations (MCS).

The MCS discards the use of the observation vector of Gauss-Markov model. In fact, the proposed method here depends only on design matrix A; the uncertainties of the observations Σ_y; and the magnitude intervals of simulated outlier. The random errors (or residues) are generated artificially from the normal statistical distribution, while the size of outliers is selected using standard uniform distribution.

Iterative data snooping shows high success rates in the experiments of a GNSS network for a single outlier randomly generated between four and five standard deviations. The efficiency of the data snooping also depends on the significance level α. Here, the optimal value for the significance level was 0.001 (0.1%) for the GNSS network analysed. This value depends more or less on the functional and stochastic model.

Furthermore, we note that the outlier identifiability in the iterative data snooping procedure is much more complex than that proposed by Prószyńsk (2015), because in our case a removal operation of observations is performed and we are not restricted only in the first adjustment run. Here the reliable identification of outlier not only depends on type III errors, but also the "over-

identifications case". The available probability of "over-identification–" allows enhancing the probability of type III error and avoids any misinterpretation. In the case of "over-identification+" requires further clarification in terms of theoretical basis and experiments (e.g. issues related to Minimal Detectable Bias and Minimal Identifiable Bias). The latter case will be investigated more closely in the next study.

Finally, we show that MCS is a feasible method to compute the probabilities level associated to a statistical testing procedure regardless of the statistical tables. Future studies should consider the case of multiple outliers which is a much more complicated problem and will be a topic of the next research.

6. Acknowledgments

The authors thank CNPq for financial support provided to the second author (proc. n. 305599/2015-1) and IBGE for the GNSS data. We also appreciate the support of all members of the research group: "Quality Control in Geodesy" (http://dgp.cnpq.br/dgp/espelhogrupo/3674873915161650). Last but not least we are also thankful to the two anonymous reviewers who provided contribution to the revision of the paper.

REFERENCES

Amiri-Seemkooei, AR 2001a, 'Comparison of reliability and geometrical strength criteria in geodetic networks', Journal of geodesy, vol. 75, issue 4, pp.227-233.

Amiri-Seemkooei, AR 2001b, 'Strategy for designing geodetic network with high reliability and geometrical strength', Journal of Surveying Engineering, vol. 127, issue 3, pp.104-117.

Aydin, C & Demirel, H 2004, 'Computation of baarda's lower bound of the non-centrality parameter', Journal of Geodesy, vol. 78, issue 4, pp. 437-411.

Baarda, W 1968, 'A testing procedure for use in geodetic networks', Publications on Geodesy 9, vol. 2, no. 5, Delft, Netherlands Geodetic Commission, viewed 10 August 2017, http://www.ncgeo.nl/phocadownload/09Baarda.pdf

Berber, M. & Hekimoglu, S 2003, 'What is the reliability of conventional outlier detection and robust estimation in trilateration networks?', Survey Review, vol. 37, issue 290, pp.308-318.

Box, GE & Muller, ME, 1958, 'A note on the generation of random normal deviates'. The Analls of Mathematical Statistics, vol. 29, no. 2, pp.610-611.

Erdogan, B 2014, 'An outlier detection method in geodetic networks based on the original observations', Boletim de Ciências Geodésicas,vol. 20, no. 3, pp.578-589.

Förstner, W 1983, 'Reliability and discernability of extended gauss-markov models' Seminar on Mathematical Models of Geodetic/Photogrammetric Point Determination with Regard to Outliers and Systematic Errors, Deutsche Geodätische. Kommision, Series A, no. 98, Munich, 1983, pp. 79-103.

Ghilani, C D, 2010, Adjustment computations: spatial data analysis, 5th edition, John Wiley & Sons, New Jersey.

Hawkins, DM 1980, 'Identification of outliers. Chapman and Hall', Chapman and Hall, London/New York.

Hekimoglu, S & Koch, KR, 1999, 'How can reliability of the robust methods be measured?'. In: Altan MO, Gründig L (eds) Third Turkish-German joint geodetic days, vol. 1, pp. 179–196.

Huber, PJ 1964, 'Robust estimation of a location parameter'. Annals of Mathematical Statistics, vol. 35, no. 1, pp.73-101.

Kavouras, M, 1982, 'On the detection of outliers and the determination of reliability in geodetic networks'. PhD thesis, Fredericton: Department of Surveying Engineering, University of New Brunswick.

Klein, I. 2014. Proposal of a new method for geodetic networks design (in Portuguese). PhD thesis, Porto Alegre: Universidade Federal do Rio Grande do Sul.

Klein, I; Matsuoka, MT; Souza, SF & Collischonn, C 2012, 'Design of geodetic networks reliable against multiple outliers' (in Portuguese), Boletim de Ciências Geodésicas, vol. 18, no. 3, pp.480-507.

Klein, I; Matsuoka, MT; Guzatto, MP; Souza, SF & Veronez, MR 2015a. 'On evaluation of different methods for quality control of correlated observations', Survey Review, vol. 47, issue 340, pp.28-35.

Klein, I; Matsuoka, MT & Guzatto, MP 2015b, 'How to estimate the minimum power of the test and bound values for the confidence interval of data snooping procedure' (in Portuguese), Boletim de Ciências Geodésicas, vol. 21, no. 1, pp.26-42.

Klein, I; Matsuoka, MT; Guzatto, MP; Nievinski, FG 2017. 'An approach to identify multiple outliers based on sequential likelihood ratio tests', Survey Review, vol. 49, Issue 357, pp. 449-457. DOI: 10.1080/00396265.2016.1212970

Knight, N L; Wang, J & Rizos, C 2010, 'Generalised measures of reliability for multiple outliers', Journal of Geodesy,vol. 84, issue 10, pp 625-635.

Koch, KR, 1999, Parameter Estimation and Hypothesis Testing in Linear Models, 2nd edition, Springer Verlag,Berlin Heidelberg, New York.

Kuang, S1991, 'Optimization and design of deformation monitoring schemes', PhD thesis, Fredericton: Department of Surveying Engineering, University of New Brunswick.

Larson, HJ 1974, Introduction to probability theory and statistical inference, 2nd edition, John Wiley & Sons, New York.

Lehmann, R & Scheffler, T 2011, 'Monte Carlo-based data snooping with application to a geodetic network'. Journal of Applied Geodesy, vol. 5, issue (3-4), pp.123-134. DOI: 10.1515/JAG.2011.014

Lehmann, R 2012, 'Improved critical values for extreme normalized and studentized residuals in gauss-markov models', Journal of Geodesy, vol. 86, issue 12, pp.1137-1146.

Lehmann, R 2013, 'On the formulation of the alternative hypothesis for geodetic outlier detection',Journal of Geodesy, vol. 87, issue 4, pp.373-386.

Lehmann, R & Lösler, M 2016, 'Multiple outlier detection: hypothesis tests versus model selection by information criteria', Journal of Surveying Engineering, vol. 142, issue 4. DOI: 10.1061/(ASCE)SU.1943-5428.0000189

Lehmann, R & Voß-Böhme, A 2017, 'On the statistical power of Baarda's outiler test and some alternative', Journal of Geodetic Science, vol. 7, issue 1, pp.68-78.

Niemeier, W & Tengen, D 2017, 'Uncertainty assessment in geodetic network adjustment by combining GUM and Monte-Carlo-simulations', Journal of Applied Geodesy, vol. 11, issue 2, pp. 67-76. DOI: 10.1515/jag-2016-0017.

Pope, AJ 1976, 'The statistics of residuals and the detection of outliers'. NOAA Technical Report NOS65 NGS1, US Department of Commerce, National Geodetic Survey Rockville, Maryland.

Prószyńsk, W 2015, 'Revisiting Baarda's concept of minimal detectable bias with regard to outlier identificability', Journal of Geodesy, vol. 89, issue 10, pp. 993-1003.

Rousseeuw, PJ and Leroy, AM 1987, 'Robust regression and outlier detection'. John Wiley & Sons, New Jersey.

Teunissen, PJG 2003, Adjustment Theory: An Introduction. Delft University Press, Delft University of Technology, The Netherlands.

Teunissen, PJG 2006, Testing theory: An Introduction. Series on Mathematical Geodesy and Positioning, 2nd edition, VSSD, Delft University of Technology, The Netherlands.

Teunissen, PJG 2017, 'Distributional theory for the DIA method', Journal of Geodesy, pp.1-22. DOI: 10.1007/s00190-017-1045-7.

Vaníček, P; Krakiwsky, EJ; Craymer, MR; Gao, Y & Ong, PS 1990, 'Robustness analysis'. Tech. Rep. no. 156, Dept. of Surveying Engineering, University of New Brunswick, Fredericton, Canada.

Vaníček, P; Craymer, MR & Krakiwsky, EJ 2001, 'Robustness analysis of geodetic horizontal networks', Journal of Geodesy, vol. 75, issue 4, pp. 199-209.

Wilcox, RR 2012, Introduction to robust estimation and hypothesis testing, Academic Press, Waltham, MA.

Xu, P 2005, 'Sign-constrained robust least squares, subjective breakdown point and the effect of weights of observations on robustness', Journal of geodesy, vol. 79, issue 1, pp. 146–159.

Yang, L; Wang, J; Knight, NL & Shen, Y 2013, 'Outlier separability with a multiple alternative hypotheses test', Journal of Geodesy, vol. 87, issue 6, pp.591-604. DOI: 10.1007/s00190-013-0629-0.